Native Speakers of Spanish Show Rate-Dependent Processing of English Stop Consonants

Abstract

English monolinguals and native Spanish speakers of English used a 9-point scale to rate syllable-initial stops for goodness as realizations of the English /p/ category. Voice onset time (VOT) was varied in a set of short-duration ('fast-rate') consonant-vowel (CV) stimuli, and in a set of longer-duration ('slow-rate') CV stimuli. The VOT values ranged from values typical for English /b/ to values which exceeded those typical for English /p/. Results for the native English (NE) subjects replicated those obtained previously using the same two continua. Goodness ratings systematically increased, then decreased as VOT values extended beyond the range typical for English /p/. The NE subjects gave their highest ratings to stimuli with VOT values of about 50 ms. For stops with longer VOT values, their ratings were higher for stimuli in the slow-rate than fast-rate continuum. The native Spanish (NS) subjects were assigned to two subgroups based on degree of foreign accent in their pronunciation of English sentences. Both proficient and nonproficient NS subjects gave their highest ratings to stimuli with much the same VOT values as the NE subjects, even though /p/ is realized with short-lag VOT in Spanish. The nonproficient, but not the proficient NS subjects showed significantly smaller rate effects on goodness ratings than did the NE subjects. However, the subjects in both NS groups gave significantly higher ratings to stimuli with short-lag VOT values than did the NE subjects. The results are discussed in terms of the hypothesis that adult learners of a second language may establish new perceptual phonetic categories for phonetic segments not found in the native language.
Introduction

Phonetic segments said to be ‘shared’ by two languages sometimes differ acoustically and articulatorily. For example, /p t k/ are realized as unaspirated stops with short-lag voice onset time (VOT) values in Romance languages such as Spanish, Italian, and French, but as aspirated stops with long-lag VOT values in English [Lisker and Abramson, 1964]. As a result, English speakers require longer VOT values to perceive stops as phonologically voiceless than do speakers of Romance languages [Abramson and Lisker, 1973; Williams, 1977; Flege and Eefting, 1986]. The purpose of this research was to examine the ability of native Spanish speakers of English to perceive English /p/. Specifically, the research addressed the issue of whether native Spanish (NS) speakers who learned English in adulthood had, or had not, established a phonetic category representation for the long-lag /p/ of English, which differs in terms of VOT and other dimensions from the /p/ of Spanish.

Many previous second-language (L2) speech production studies have focused on VOT in /p t k/ tokens produced in the initial position of English words by native speakers of Romance languages [e.g., Carmazza et al., 1973; Williams, 1979a, b; Williams, 1980; Flege and Port, 1981; Flege and Hillenbrand, 1984; Flege and Eefting, 1987; Major, 1987; Nathan, 1987; Schmidt, 1988; Flege, 1987, 1991; Mack, 1990; Flege et al., in press]. Subjects have usually read words from a list, either in isolation or in a carrier phrase. The style of speech examined in most previous L2 studies might therefore be characterized as relatively slow and careful. Only one previous study has examined the production of L2 stops in casual speech [Major, 1992]. No VOT study, to our knowledge, has examined stops produced by nonnative subjects at a fast speaking rate.

Research has shown that the earlier in life English is learned as an L2, the longer (more accurate) are the VOT values produced in English /p t k/ [Flege et al., in press]. Most adult learners of English produce English /p t k/ with ‘compromise’ VOT values that are intermediate to the values typical for English and the native language (L1). However, there are sometimes exceptions to this general rule (e.g., Schmidt, 1988; Flege, 1991; see also Suomi, 1976). A few adult learners have been observed to produce English /p t k/ with the short-lag VOT values characteristic of their L1. This suggests the replacement of an English stop by one from the L1. Also, a few late learners have been observed to produce English /p t k/ with VOT values that equaled, or even exceeded, the values observed for English monolinguals.

One possible explanation for the intersubject variability just described is that although adult L2 learners may perceive the acoustic-phonetic properties of English /p t k/ accurately, only some possess the motoric ability to produce stops in the L2 with a laryngeal timing pattern differing from that used in the L1 [Abramson, 1977]. Alternatively, adult L2 learners’ accuracy in producing English /p t k/ may vary as a function of their accuracy in perceiving the acoustic-phonetic properties of English stops.

Elman et al. [1977] observed a difference in the identifications of naturally produced short-lag bilabial stops by ‘moderate/weak’ and ‘strong’ Spanish/English bilinguals. The weak/moderate bilinguals identified roughly half of the stops as /p/ in both English and Spanish perceptual ‘sets’. The strong bilinguals, on the other hand, identified fewer short-lag stops as /p/ in the English than Spanish set. They therefore resembled Spanish monolinguals (who heard most stops as /p/) in the Spanish set, and they resembled English monolinguals (who heard most of the short-
lag stops as /b/) in the English set. The strong bilinguals may have had more accurate perceptual representations for English /p/ than the moderate/weak bilinguals. One might even hypothesize that they, but not the moderate/weak bilinguals, had established a phonetic category representation for English /p/ [but cf. Bohn and Flege, 1993].

The strength of bilingualism of subjects in the Elman et al. [1977] study was based on ratings of overall ‘fluency’ by two experimenters following an oral interview. The extent to which the fluency ratings depended on the subjects’ age of learning English, and the amount and kind of English-language input these subjects may have received, is unknown. Most native Spanish-speaking subjects examined by Elman et al. [1977] were early learners. The few late learners who were examined showed little or no effect of the language set manipulation [Diehl, 1986, personal commun.].

Flege [1992a,b, in press] hypothesized that the likelihood of phonetic category formation for L2 phonetic segments is influenced importantly by the age at which L2 learning commences. More specifically, he hypothesized that the range of L2 segments for which additional phonetic categories are established decreases through childhood, but that even adult learners of an L2 may establish phonetic categories for L2 segments that differ substantially from the nearest L1 segment.

A study by Flege and Eefting [1988] provided indirect evidence that native speakers of Spanish who learn English in early childhood may establish phonetic categories for English /t/. The early L2 learners examined in that study identified the initial stop in consonant-vowel (CV) stimuli differing in VOT as ‘d’ or ‘t’. They then attempted to imitate stops as accurately as possible when the CV stimuli were randomly presented following completion of the forced-choice task. VOT values produced by the subjects increased as stimulus VOT increased, but the relation between VOT values in the stimuli and the VOT values produced by the subjects in their imitation responses was nonlinear. Large discontinuities in the stimulus-response functions occurred at VOT values which coincided closely with the phoneme boundaries obtained in the earlier identification experiment. This result suggested that although the subjects had not been asked to do so, they may nevertheless have phonetically labeled initial stops in the CV stimuli before (or during) their imitations of the stimuli.

Spanish monolinguals examined by Flege and Eefting [1988] produced mostly stops with lead and short-lag VOT values. English monolinguals produced mostly stops with short-lag and long-lag VOT values. However, three groups of Spanish/English bilinguals produced stops falling into all three modal VOT ranges [Lisker and Abramson, 1964]. Given the brevity of the interval between presentations of the CV stimuli to be imitated, the bilinguals’ composite pattern of VOT production in the imitation task suggested that they may have had three phonetic categories: one for /d/, one for voiceless unaspirated /t/ (such as Spanish /t/), and one for voiceless aspirated /t/ (such as English /t/). However, the study did not shed light on the hypothesis that adult L2 learners may establish phonetic categories for English /p t k/. This is because the subjects in all three bilingual groups had begun learning English by the age of 7 years. (Moreover, one might argue that since the dependent variable in the Flege and Eefting [1988] study was a measure of speech production, it provided little insight into perceptual representations in long-term memory.)

Two assumptions underlying the research reported here were that perceptual phonetic category representations specify how phonetic elements ‘ought’ to sound when produced [Linell, 1979], and that perceptual representations for phonetic segments are related to their
production. Perceptual phonetic category representations specify values for a wide range of acoustic-phonetic dimensions (or ‘cues’). The absolute normative values of these dimensions, as well as their relative importance, may vary as a function of phonetic context, lexical context, degree of stress or emphasis, and speaking rate [e.g., Massaro, 1987]. One might expect small differences in cue weighting to exist among individuals who speak a particular language [e.g., Samuel, 1977; Flege and Hillenbrand, 1985]. However, one would expect even larger differences in normative parameter values for certain phonetic segments to exist between speakers of different languages, as well as between the relative importance of various acoustic phonetic dimensions.

Spanish and English monolinguals’ representations for /p/ are apt to differ in numerous ways. For example, English monolinguals’ representation for /p/ will surely specify a longer VOT value than Spanish monolinguals’. The laryngeal timing difference that distinguishes English /p/ from Spanish /p/ gives rise to many acoustic-phonetic differences [Abramson, 1977], however. Also, VOT is not always an overriding cue to the voicing feature in stops either for English or Spanish monolinguals [Forrest and Rockman, 1988; Bohn and Flege, 1993]. Spanish and English monolinguals’ representations might therefore be expected to differ along many dimensions in addition to VOT. These dimensions include the amplitude of the release burst in syllable-initial stops [Williams, 1977], release burst duration [Bohn and Flege, 1993], the amplitude of the aspiration noise following stop release [Repp, 1979], fundamental frequency (F₀) variations occurring in the tens of milliseconds following release [e.g. Ohde, 1985], the presence versus absence of energy immediately following the stop release [Serniclaes, 1976; Williams, 1977], F₁ onset frequency [Simon and Fourcin, 1978], and amplitude rise time [Munro, 1987]. It would be a daunting task indeed to test for the full range of acoustic-phonetic dimensions that might differentiate English and Spanish monolinguals’ phonetic category representations for /p/. The present research therefore focused just on VOT.

The paradigm and stimuli used by Miller and Volaitis [1989] and Volaitis and Miller [1992] were used here to test the hypothesis that some native Spanish late learners establish a phonetic category for English /p/. The paradigm is based on the observation that, in speech production, the duration of VOT intervals, especially those of voiceless stops, is shortened as speaking rate increases [e.g., Miller and Baer, 1983]. Listeners’ perceptual responses to phonetic distinctions cued in part by VOT change as a function of speaking rate, as if listeners somehow ‘adjusted’ (or normalized) their perceptual responses to accommodate perceived changes in speaking rate [see Miller and Wayland, 1993, for discussion]. In the Miller paradigm, speaking rate variations are simulated by varying overall syllable duration in two VOT continua. The VOT of syllable-initial bilabial stops ranges from short-lag values typical for English /b/ to values that exceed those typical for English /p/. English monolinguals have rated the goodness of the bilabial stops as realizations of the English /p/ category. Their ratings have been observed to increase systematically as VOT increases, then to decrease as VOT values extended beyond typical English values.

Of primary interest to the present study was the finding that native English (NE) subjects’ ratings varied as a function of stimulus duration in a way that corresponded to the differences in VOT that exist when /p/ is produced at relatively slow versus fast speaking rates. The NE and NS subjects who participated in the present study rated the members of
Miller’s VOT continua as realizations of the English /p/ category. The pattern of ratings expected for the NE subjects is represented schematically in figure 1a. The remaining three panels of figure 1 present data patterns that might be obtained from the NS subjects if they diverged from the pattern typical for NE subjects. NS subjects might give their highest ratings only to stops with short-lag VOT values, as illustrated in figure 1b. However, even if the NS subjects resembled the NE subjects in showing a systematic increase in goodness ratings as VOT increased from 10 to about 50 ms, it would not necessarily prove that they had a perceptual phonetic category representation for English /p/. Such a response pattern could, for example, result from the assignment of higher ratings to stops as their perceived distance from Spanish /p/ increased.

We therefore required that additional criteria be met before reaching the conclusion that the NS subjects had established a representation for English /p/. Another criterion was that the NS subjects give their highest ratings to stimuli with much the same VOT values as the NE subjects, and that their goodness ratings show a systematic decrease as VOT increased beyond the values typical for English /p/, as illustrated in figure 1c.

Like the first kind of evidence mentioned, this would not in itself be a convincing demonstration of category formation. A systematic decrease in ratings beyond the VOT values typical for English might be based on judgments of the extent to which long-lag VOT values (or the ‘voiced’ portion of the CV stimuli; see below) seemed abnormal or ‘exaggerated’ in some universal sense.

No previous research has examined rate-dependent processing of stops by Spanish monolinguals. Given that short-lag stops show far less variation as a function of speaking rate than do long-lag stops [Miller et al., 1986; Schmidt and Flege, 1995], and given that the /p/ of Spanish is realized with short-lag VOT values [Lisker and Abramson, 1964], one might expect Spanish monolinguals to show little or no rate effect for long-lag stops.
Table 1. Characteristics of the 20 NS subjects each who were assigned to relatively proficient and nonproficient subgroups based on their pronunciation of English sentences (top) and characteristics of the same NS subjects assigned to subgroups based on total English input (bottom)

<table>
<thead>
<tr>
<th></th>
<th>Nonproficient subjects</th>
<th>Proficient subjects</th>
<th>Difference</th>
<th>t(38)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age</td>
<td>30.8(7.2)</td>
<td>32.3(7.7)</td>
<td>1.5(2.3)</td>
<td>0.63</td>
<td>0.532</td>
</tr>
<tr>
<td>Gender</td>
<td>4 m, 16 f</td>
<td>9 m, 11 f</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age of arrival in US</td>
<td>27.1(5.4)</td>
<td>24.6(6.7)</td>
<td>2.5(1.9)</td>
<td>1.28</td>
<td>0.209</td>
</tr>
<tr>
<td>Residence in the US</td>
<td>3.7(3.3)</td>
<td>7.7(7.7)</td>
<td>4.0(1.9)</td>
<td>2.10</td>
<td>0.043*</td>
</tr>
<tr>
<td>% use of English</td>
<td>36%(18)</td>
<td>61%(22)</td>
<td>26%(6)</td>
<td>3.98</td>
<td>0.000**</td>
</tr>
<tr>
<td>Total English input</td>
<td>1.4(1.4)</td>
<td>5.5(6.5)</td>
<td>4.1(1.5)</td>
<td>2.78</td>
<td>0.008**</td>
</tr>
<tr>
<td>Pronunciation score</td>
<td>65(18)</td>
<td>138(35)</td>
<td>73(9)</td>
<td>8.24</td>
<td>0.000**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Inexperienced subjects</th>
<th>Experienced subjects</th>
<th>Difference</th>
<th>t(38)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age</td>
<td>29.2(6.5)</td>
<td>33.8(7.6)</td>
<td>4.7(2.2)</td>
<td>2.10</td>
<td>0.042*</td>
</tr>
<tr>
<td>Gender</td>
<td>6 m, 14 f</td>
<td>7 m, 14 f</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age of arrival in US</td>
<td>27.1(6.2)</td>
<td>24.5(5.9)</td>
<td>2.6(1.9)</td>
<td>1.35</td>
<td>0.184</td>
</tr>
<tr>
<td>Residence in the US</td>
<td>2.1(1.1)</td>
<td>9.4(7.1)</td>
<td>7.3(1.6)</td>
<td>4.56</td>
<td>0.000**</td>
</tr>
<tr>
<td>% use of English</td>
<td>33%(18)</td>
<td>64%(18)</td>
<td>31%(6)</td>
<td>5.41</td>
<td>0.000**</td>
</tr>
<tr>
<td>Total English input</td>
<td>0.7(0.5)</td>
<td>6.2(6.1)</td>
<td>5.6(1.3)</td>
<td>4.10</td>
<td>0.000**</td>
</tr>
<tr>
<td>Pronunciation score</td>
<td>79(33)</td>
<td>124(47)</td>
<td>45(13)</td>
<td>3.50</td>
<td>0.001**</td>
</tr>
</tbody>
</table>

Mean values and standard deviations (in parantheses) are given.

A third criterion for category formation was that the NS subjects show evidence of rate-dependent processing, as illustrated in figure 1d.

Method

Subjects
Sixty residents of Birmingham, Ala., all of whom passed a pure-tone hearing screening, participated as subjects. Twenty subjects (5 males, 15 females) were monolingual native speakers of American English between the ages of 21 and 55 years (mean = 32 years). Seven NE subjects were from the Southeastern part of the US (mostly Ala.), 8 were from the Midwest, 3 were from the Northeast, and 2 were from the Western US. The remaining 40 subjects were native speakers of Spanish who had learned English as an L2. All of the NS subjects began learning English in the US after the age of 13 years. The NS subjects were from nine different Latin American and Caribbean countries, including Columbia (n = 10), Mexico (n = 9), Peru (n = 6), and Puerto Rico (n = 5). As summarized in Table 1, the mean age of the 13 male and 27 female NS subjects was 32 years (range = 19–54 years). The NS subjects had arrived in the US at an average age of 26 years (range = 14–39 years) and had lived in the US for an average of 6 years (range = 1 month to 35 years). They estimated using English 48% of the time on a daily basis (range = 3–93%).

A number of subjects had to be replaced. One NE subject failed to pass the hearing screening, and another failed to use at least half of the available range in the rating task. Also, unlike any of the subjects examined by Miller and Volaitis [1989] and the other 21 NE subjects tested, 3 NE subjects apparently did not hear any /b/s in the identification task. Two NS subjects failed to pass the hearing screening. Eight NS subjects failed to use at least half of the range in the rating task, and 1 gave steadily increasing goodness ratings as VOT increased. This NS subject apparently interpreted the task as a kind of psychoacoustic evalu-
ation of VOT (or vowel duration), not as an evaluation of the phonetic typicality of /pl/ tokens.

Twenty NS subjects each were assigned to relatively 'proficient' and 'nonproficient' subgroups according to their pronunciation of the following English sentences: I can read this for you, The red book was good, The good shoe fit Sue. These sentences were rated for degree of global foreign accent using procedures described previously [e.g., Flege and Fletcher, 1992]. Briefly, sentences spoken by the 40 NS subjects and 5 randomly selected NE subjects were band-pass-filtered (80-9,000 Hz), digitized at 20.0 kHz, normalized for peak intensity, and stored on disk. The productions of each sentence were randomly presented in a separate block to 10 monolingual NE listeners, who were informed that most of the sentences had been spoken by nonnative speakers of English. The listeners were instructed to position the lever on a response box at some place along a scale ranging from 'no foreign accent' at the top of the lever’s range, to 'medium foreign accent' at the middle, to 'strongest foreign accent' at the very bottom of the scale. The listeners were further instructed to reserve the top of the scale for sentences they believed to have been spoken by a native speaker of English. They were to reserve the bottom of the scale for the one talker who had the strongest foreign accent of all talkers being considered. A value of 0 (strongest accent) to 255 (no accent) was stored on the PC depending on where the lever had been positioned. Each sentence was presented 1.0 s after a rating was received for the last sentence. A mean rating was calculated for each sentence based on the final three (of four) replicate judgments.

Intraclass correlation analyses of the ratings obtained for the 10 listeners were carried out for each of the three sentences. The R values ranged from 0.912 to 0.953 (p < 0.001), indicating that the listeners performed the foreign accent rating task in much the same way. The scores obtained for the Spanish subjects' productions of the three sentences were correlated significantly (r = 0.70-0.77, p < 0.001). Therefore, a single mean foreign accent score was calculated for each subject, each of which was based on 90 ratings (10 listeners x 3 sentences x 3 replicate judgments). The NS subjects' ratings averaged 102, and ranged from 29 to 238. Only 1 NS subject obtained a score falling within the range obtained for the NE subjects [i.e., 236-251]. Many NS subjects' scores fell below the scale midpoint (i.e. 128), so they might be said to have spoken English with a strong foreign accent.

The NS subjects were rank-ordered according to their foreign accent scores. The 20 subjects with the highest (i.e. least foreign-accented) scores were assigned to a 'proficient' group, and those with the 20 lowest scores were assigned to a 'nonproficient' group.

Stimuli

The synthetic CV stimuli used in the two experiments reported here were those used by Miller and Volaitis [1989]. The CV stimuli were synthesized at 10.0 kHz using the cascade branch of a software synthesizer. CV stimuli in the 'fast-rate' continuum had a duration of 125 ms; those in the 'slow-rate' continuum had a duration of 325 ms. Each stimulus contained a 5-ms release burst followed by 5 ms of silence. Formant transitions lasting from 20 to 45 ms cued a syllable-initial bilabial stop.

As described by Miller and Volaitis [1989], VOT was increased by widening the F1 bandwidth and switching from a periodic to an aperiodic noise source. VOT in the 17 fast-rate stimuli increased in 5-ms steps from 10 to 60 ms, then in 10-ms steps from 60 to 120 ms. VOT in the 37 slow-rate stimuli increased in 5-ms steps from 10 to 60 ms, then in 10-ms steps from 60 to 320 ms. The F1 transition lasted for 20 ms. This meant that F1 onset frequency and VOT covaried in the first 4 stimuli. F0 rose from 100 to 125 Hz over 45 ms, so F0 and VOT covaried in the first 9 stimuli. The voiced portion of the 'vowel' decreased in the constant-duration stimuli as VOT increased. Thus, only 5 ms of voiced vowel remained in the fast-rate and slow-rate stimuli having the longest VOT values. Given the difficulty in locating bilingual subjects, only a single session was practical. To accommodate this need, 9 of the original 37 slow-rate stimuli (i.e., those with VOT values of 150, 170, 190, 210, 230, 250, 270, 290 and 310 ms) were eliminated from the set of stimuli used originally by Miller and Volaitis [1989].

Procedure

The 60 subjects were tested one at a time in a sound booth. They first responded to a language background questionnaire. Next, they participated in a speech production task that is reported in a companion article [Schmidt and Flege, 1995]. Finally, they rated word-initial stops for goodness.

The CV stimuli were low-pass-filtered (4.8 kHz) and presented over headphones at a comfortable level. The interval between each response and the presentation of the next stimulus was fixed at 1.5 s. The 17 fast-rate and 28 slow-rate stimuli were randomly presented 11 times each in separate, counterbalanced blocks. A brief period of familiarization preceded the two blocks. The familiarization consisted of playing
out the stimuli to be rated a single time in order of increasing VOT. The subjects were told that they would probably first hear /b/, then eventually a 'good' /p/ as in the words pea or Pete, and finally 'exaggerated' tokens of /p/. They were instructed to type a number from 1 to 9 on a keyboard depending on how good an example of English /p/ they judged the initial consonant in each CV to be. The subjects were advised explicitly that 'good' examples of /p/ should get a higher rating than either /b/ or 'exaggerated' /p/ tokens. The subjects were required to rate each initial consonant and were told to use the entire scale. The dependent variable was the average of the final ten (of 11) ratings of each CV.

The NE subjects returned for a second session, during which they identified members of the two VOT continua they had rated previously for goodness. (Had the identification task preceded the rating task, as it did in the 1989 study of Miller and Volaitis, the NE subjects would have had more familiarity with the stimuli in the rating task than did the NS subjects.) Once again, the 17 fast-rate and 28 slow-rate stimuli were randomly presented 11 times each in separate, counterbalanced blocks preceded by a brief period of familiarization. The NE subjects were instructed to push a button marked 'b' if they heard /b/, a button marked 'p' if they heard /p/, and a button marked 'breathy/exaggerated' if they heard /#p/. The subjects were required to respond to each stimulus and were told to guess if unsure. The dependent variable was the percentage of times (out of the final 10 judgments) that each stimulus was identified as /b/, /p/, or /#p/.

Results

Identification

Figure 2 shows the results of the identification task, in which only the NE subjects participated. Figure 2a plots the percentage of times stimuli in the fast-rate continuum (i.e., the continuum made up of the relatively short-duration CVs) were identified as /b/, /p/, and exaggerated (breathy) /#p/. Figure 2b plots the results for the slow-rate (i.e., long-duration) stimuli. In the fast-rate continuum, stimuli with VOT values ranging from 10 to 25 ms were identified predominantly (>80%) as /b/. Those with VOT values ranging from 40 to 60 ms were identified predominantly as /p/, and those with VOT values of 110–120 ms were identified predominantly as /#p/. In the slow-rate continuum, stimuli with VOT values of 10–30 ms were identified predominantly as /b/; those with VOT values ranging from 45 to 100 ms as /p/, and those with VOT values of 200–320 ms as /#p/.

Boundaries between /b/-/p/ and between /p/-/#p/ appeared to occur at longer VOT values in the slow-rate than fast-rate continuum. To test this, phoneme boundaries equivalent to 50%
crossovers were computed by fitting regression lines to the data in the boundary regions. The /b/-/p/ boundaries averaged 37.7 ms in the slow-rate continuum, and 34.4 ms in the fast-rate continuum. The /p/-/*p/ boundaries averaged 150.1 ms in the slow-rate continuum, and 89.4 ms in the fast-rate continuum. The phoneme boundary scores were examined in a (2) Speaking Rate x (2) Phoneme Boundary repeated-measures ANOVA. The significant two-way interaction it yielded \( F(1, 19) = 70.8, p < 0.05 \) was explored by tests of simple main effects. As expected, the difference between the phoneme boundaries (/b/-/p/ versus /p/-/*p/) was significant in both continua \( p < 0.05 \). The effect of speaking rate (i.e., CV stimulus duration) was significant for the /p/-/*p/ boundary \( F(1, 37) = 137.9, p < 0.05 \), but not for the /b/-/p/ boundary \( F(1, 37) = 0.415, p > 0.10 \). This is because 5 of the 20 NE subjects failed to show a crossover from /b/ to /p/ at shorter VOT values for stimuli in the fast-rate than slow-rate continuum.

The lack of a significant effect of speaking rate on the location of /b/-/p/ boundaries diverges from the results obtained by Miller and Volaitis [1989] and Volaitis and Miller [1992] using the same stimuli. The manipulation of speaking rate shifted the /b/-/p/ phoneme boundary a nonsignificant 3.3 ms in the present experiment. The rate effects obtained in five experiments by Miller and Volaitis [1989] and Volaitis and Miller [1992], which ranged from 2.5 to 8.3 ms (mean = 4.8 ms), were all significant. The difference between experiments may have been due to the dialect of the subjects tested, but were more likely to have been due to methodological differences.

For example, Flege and Hillenbrand [1985] reported a significant difference between northern and southern American listeners in the identification of final fricatives. Only 3 of the 20 subjects examined in the present experiment were from the northeastern part of the US, whereas the majority of subjects examined in previous experiments appear to have been from the Northeast.

The previous research [Miller and Volaitis, 1989; Volaitis and Miller, 1992] included a practice phase, whereas the present experiment did not. Subjects in the present experiment identified each stimulus 11 times each, whereas subjects in the earlier experiments identified each stimulus 15 or 20 times each. Also, subjects in the present experiment identified stimuli after having rated them in an earlier session whereas in the experiments of Miller and Volaitis [1989] and Volaitis and Miller [1992] subjects identified stimuli before rating them for goodness.

Most importantly, the set of VOT stimuli used here differed somewhat from the stimulus sets used previously. Nine slow-rate stimuli used by Miller and Volaitis [1989] and Volaitis and Miller [1992] were not included in the present experiment (i.e., the stimuli with VOT values of 150, 170, 190, 210, 230, 250, 270, 290 and 310 ms). The range and frequency of VOT values presented to listeners are known to influence their identification of stops as voiced or voiceless [e.g., Brady and Darwin, 1978]. Eliminating the 9 stimuli did not change the overall range of VOT values in the stimulus set, but it did affect the overall frequency of VOT values that exceeded the English phonetic norm in the slow-rate continuum. All of the stimuli eliminated were from the ‘exaggerated’ (‘breathy’) portion of the continuum.

**Rating Data**

The NE subjects showed a variety of rating patterns. The mean goodness ratings of slow-rate and fast-rate stimuli obtained from 3 NE subjects are shown in figure 3. Subject NE-3 showed a speaking rate effect that was similar to the grouped data presented below for the NE subjects. In both continua, the goodness
ratings given by this subject first increased, then decreased systematically as VOT values exceeded those typical for English /p/.

Subject NE-3 showed a clear effect of speaking rate (i.e., CV duration) on goodness ratings. In the portion of the two continua which followed the highest goodness ratings, stimuli with the same VOT values were given higher ratings in the slow-rate than fast-rate continuum. NE-3 gave low ratings to the stimuli with short-lag VOT values, which are typical for English /b/ and for Spanish /p/. Subject NE-22 also gave low ratings to the stimuli with short-lag VOT values. However, this subject’s rate effect was far smaller than that of NE-3 and, in fact, was confined to stimuli with VOT values ranging from 80 to 120 ms. Subject NE-21 showed a fairly large rate effect, but gave ratings to short-lag VOT stimuli that were higher than those given by the other 2 NE subjects.

There was even greater variation in the rating patterns obtained from the NS subjects. The data obtained from a few NS subjects resembled the hypothetical pattern for Spanish monolinguals shown in figure 1b. However, none of them matched that pattern perfectly. Data for 5 NS subjects are shown in figure 4. Subjects NS-10 and NS-17 gave high ratings to stimuli with short-lag VOT values. Subjects NS-4 and NS-31 gave low ratings to short-lag stimuli, on the other hand, and NS-25 gave intermediate ratings. Three subjects (NS-4, NS-17, NS-25) showed rate effects, whereas 2 others (NS-10, NS-31) showed little or no speaking rate effect when rating stimuli from the fast-rate and slow-rate continua.

The mean goodness ratings obtained from the three groups of subjects are shown in figure 5. As shown in figure 5a, the NE subjects’ mean ratings increased as VOT increased from 10 to about 50 ms, then decreased systematically as VOT increased beyond values typical for English /p/. Stimuli with VOT values ranging from about 60 to 120 ms received higher mean ratings in the slow-rate than fast-rate continuum. Figure 5b shows the mean ratings obtained for the proficient NS subjects. These subjects differed from the NE subjects primarily in their ratings of short-lag stimuli, which were higher than the NE subjects’ ratings. Figure 5c shows the results for the nonproficient NS subjects. These subjects also gave higher ratings to short-lag stimuli than did the NE subjects. Unlike the NE subjects and the proficient NS subjects, however, the nonproficient NS subjects showed a very small rate effect.
To quantify the effect of speaking rate on goodness ratings, six variables were derived from the goodness rating functions obtained from each subject. Three were for the slow-rate continuum, and three were for the fast-rate continuum. The preferred VOT value was the VOT value of the highest-rated stimulus in each continuum. In instances where two or more stimuli received the highest rating, the preferred VOT value was calculated by averaging the VOT values of those stimuli. The upper limit and the lower limit of acceptable VOT values were determined by multiplying the highest rating by 0.90, then calculating (through interpolation) the VOT values that coincided with that rating both above and below the preferred VOT value.

Averaged across the two continua, the lower limit, preferred VOT, and upper limit scores obtained for the NE subjects were 43, 53, and 83 ms, respectively. The corresponding scores obtained for the proficient NS subjects were all somewhat smaller (lower limit: 31 ms; preferred VOT: 51 ms; upper limit: 75 ms), as were the scores for the nonproficient NS subjects (lower limit: 28 ms; preferred VOT: 49 ms; upper limit: 78 ms). This undoubtedly reflects the NS subjects’ experience with short-lag realizations of /p/ in their L1.

Figure 6a shows the magnitude of rate effects obtained for the three groups. (Data in figure 6b will be discussed below.) The rate effects shown here were obtained by subtracting values obtained for stimuli in the fast-rate con-
The mean effect of speaking rate on goodness judgments by NE subjects and NS subjects differing in English pronunciation proficiency. 'Preferred' refers to the VOT values of stimuli receiving the highest ratings. The 'lower' and 'upper limit' scores were meant to define the range of acceptable VOT values in realizations of English /p/. Data for the NE subjects juxtaposed to NS subjects, who have been reassigned to subgroups based on their English-language experience. One and two asterisks indicate significance at the 0.05 and 0.01 levels.

Fig. 5. The mean goodness rating obtained from NE subjects (a), relatively proficient NS speakers of English (b) and relatively nonproficient NS subjects in response to relatively short-duration (fast-rate) and long-duration (slow-rate) CV stimuli differing in VOT (c).

Fig. 6. a The mean effect of speaking rate on goodness judgments by NE subjects and NS subjects differing in English pronunciation proficiency. 'Preferred' refers to the VOT values of stimuli receiving the highest ratings. The 'lower' and 'upper limit' scores were meant to define the range of acceptable VOT values in realizations of English /p/. b Data for the NE subjects juxtaposed to NS subjects, who have been reassigned to subgroups based on their English-language experience. One and two asterisks indicate significance at the 0.05 and 0.01 levels.
limit than preferred VOT scores which, in turn, were greater than the lower limit scores. The same held true for the two NS groups.

The proficient NS subjects’ rate effects appear to have resembled the NE subjects’ to a greater extent than the nonproficient NS subjects’. To test this, the three sets of scores were submitted to separate (3) Group x (2) Speaking Rate ANOVAs, with repeated measures on Rate. A significant main effect of Group was obtained only for the lower limit scores \(F(2, 57) = 8.37, p < 0.05\). Significant Rate main effects were obtained for the preferred VOT scores \(F(1, 57) = 6.5 p < 0.05\) and the upper limit scores \(F(1, 57) = 15.9 p < 0.05\), but not for the lower limit scores. The two-way interaction was significant only for the lower limit scores \(F(2, 57) = 14.5 p < 0.05\).

The lack of a significant Group main effect or two-way interaction in the analysis of the preferred VOT scores indicated that the stimuli receiving the highest ratings were much the same for the NS and NE subjects. This came as something of a surprise. Given that \(/p/\) is produced with shorter VOT values in Spanish than English, one might have expected the NS subjects to reserve their highest ratings for stimuli with shorter VOT values than the NE subjects. Perhaps the NS subjects, even those who pronounced English sentences with strong foreign accents, had native-like perceptual representations for English \(/p/\), and thus were able to perform the goodness rating task just like the NE subjects.

Other explanations might be offered for the similarity of the NE and NS subjects’ preferred VOT scores. For example, the primary difference between the ‘fast-rate’ and ‘slow-rate’ stimuli was overall duration. However, the range of VOT values in the two continua was confounded with stimulus duration. Whereas VOT ranged from 10 to 120 ms in the fast-rate stimuli, it ranged from 10 to 320 ms in the slow-rate stimuli. The so-called ‘rate’ effect observed for some or all of the NS subjects may have been a kind of range effect [Brady and Darwin, 1978]. In the identification experiment, NE subjects labeled slow-rate stimuli having VOT values of 160–320 ms as ‘exaggerated’, and they labeled fast-rate stimuli with VOT values of 100–120 ms as ‘exaggerated’. To the authors, most of these exaggerated stimuli did not sound like possible speech sounds.

If NS subjects had the same auditory impression as the authors concerning which stimuli contained possible speech sounds, they might have chosen the midpoint of the ‘possible speech’ range as the best exemplar of English \(/p/\). The midpoint of the VOT range of 10–90 ms in the fast-rate continuum is 50 ms. The preferred VOT values of the proficient and nonproficient NS subjects averaged 45 and 48 ms, respectively. A stimulus range explanation is less satisfactory for the slow-rate continuum, however. The midpoint of the ‘possible’ VOT range of 10–140 ms in the slow-rate continuum is 75 ms, whereas the preferred VOT values of the proficient and nonproficient NS subjects averaged 57 and 50 ms, respectively. It is possible, of course, that the NS subjects did not define the range of ‘possible’ \(/p/\) sounds in the same way as did the NE subjects (or the authors). Additional work with stimuli differing in VOT range, as well as with monolingual NS subjects, would be helpful in determining the role of stimulus range on goodness ratings. If NS subjects who had not been exposed to English show the same ‘preferred’ VOT values for English \(/p/\) as do NE speakers, it would argue strongly in favor of a stimulus range explanation.

The significant two-way interaction obtained in the analysis of the lower limit scores suggests that the NS subjects’ processing of short-lag stops was influenced by their phonetic representation for Spanish \(/p/\). The two-
way interaction was explored by simple effects tests. The effect of speaking rate on lower limit scores was significant for the NE subjects ($p < 0.05$). It was nonsignificant for proficient NS subjects. The effect of speaking rate was significant, but in a direction opposite to that of the NE subjects, for the nonproficient NS subjects ($p < 0.05$). That is, the nonproficient NS subjects judged stops with longer VOT values to be more acceptable in the fast-rate than slow-rate continuum. Thus, although the proficient NS subjects showed a smaller rate effect on the lower limit scores, their lower limit scores were clearly more English-like than were the nonproficient NS subjects.

Significant Group x Rate interactions were not obtained in the analyses of preferred VOT and upper limit scores. Despite this, the simple effect of speaking rate was tested for these variables as well. As shown in figure 6a, the effect of speaking rate on the preferred VOT scores was significant for the NE and proficient NS subjects but not for the nonproficient NS subjects ($p < 0.05$). This also supports the conclusion that the proficient NS subjects had a more native-like perception of English stops than did the nonproficient NS subjects. For upper limit scores, the rate effect was significant only for the NE subjects ($p < 0.05$).

**English-Language Experience**

In analyses presented so far, the NS subjects were rank-ordered according to their pronunciation of English sentences, then assigned to relatively proficient and nonproficient subgroups. This procedure was predicated on the assumption that the accuracy with which the NS subjects pronounced English was related to how accurately they perceived English phonetic segments [see Flege, 1992a,b]. That is, we assumed that the so-called ‘proficient’ subjects would be more likely to have native-like perceptual representations for English /p/ than would the ‘nonproficient’ NS subjects.

We also considered an alternative hypothesis, i.e. that accuracy in perceiving L2 phonetic segments increases simply as a function of amount of exposure to the L2. To test this, the NS subjects were rank-ordered according to English-language experience. L2 experience was estimated by multiplying the NS subjects’ self-estimates of percentage daily use of English by their length of residence in the US. The 40 NS subjects were then reassigned to relatively ‘experienced’ and ‘inexperienced’ subgroups of 20 subjects each based on L2 experience. Characteristics of the relatively experienced and inexperienced NS subjects are summarized in the lower half of table 1.

The relatively experienced NS subjects were estimated to have had over 5 years more of English-language experience than the relatively inexperienced NS subjects. As expected, the relatively experienced NS were more likely to have been drawn from the proficient than nonproficient NS subgroups. However, although the proficient NS subjects had undoubtedly received more English-language input than the nonproficient ones (table 1), the confound between amount of English-language experience and pronunciation was not complete. The correlation between total English input and the foreign accent scores was nonsignificant ($r = 0.274$, d.f. = 38, $p = 0.08$). Furthermore, 12 of the 40 subjects ‘switched’ groups (i.e., 6 experienced subjects were drawn from the nonproficient group, and 6 relatively inexperienced subjects were drawn from the proficient NS group).

As shown in figure 7, the rate effects obtained for both the experienced and inexperienced NS subjects were somewhat smaller than the NE subjects’. Both NS subgroups gave higher ratings to stimuli with short-lag VOT values than did the NE subjects. As
Fig. 7. The mean goodness rating obtained from NE subjects (a), relatively experienced NS speakers of English (b) and relatively inexperienced NS subjects (c).

shown in figure 6b, the experienced NS subjects' rate effects were somewhat less like the NE subjects' rate effects than those of the inexperienced NS subjects.

ANOVA's were carried out to compare the performance of the experienced and inexperi-enced NS subjects to that of the NE subjects. These tests yielded results that were similar to the results reported earlier. Once again, a significant main effect of Group was obtained only for the lower limit scores \(F(2, 57) = 8.07, p < 0.01\). Significant Rate main effects were obtained for the preferred VOT scores \(F(1, 57) = 7.06, p < 0.05\) and for the upper limit scores \(F(1, 57) = 16.4, p < 0.01\), but not for the lower limit scores. A significant Group x Rate interaction was obtained for the lower limit \(F(2, 57) = 5.74, p < 0.01\) and preferred VOT scores \(F(2, 57) = 3.64, p < 0.05\); the two-way interaction in the analysis of the upper limit scores was marginally significant \(F(2, 57) = 2.72, p = 0.07\). Simple effects tests revealed that the NE subjects, but neither NS group, showed significant rate effects on the lower and upper limit scores \((p < 0.05)\). The NE and inexperienced NS subjects, but not the experienced NS subjects, showed significant rate effects on preferred VOT scores. (The experienced NS subjects actually showed a small average difference in a direction opposite to that of the NE subjects.)

The inexperienced NS subjects' rate effects on preferred VOT and upper limit scores resembled the NE subjects' scores to a greater extent than the experienced NS subjects' scores. This is paradoxical. One would expect additional L2 experience to result in more rather than in less native-like performance in the L2. Recall that a comparable finding was not obtained in analyses comparing the NE subjects to NS subgroups differing in English pronunciation proficiency. Taken together, then, the analysis just presented suggests that late learners' accuracy in perceiving an L2 phonetic segment is tied more closely to their overall ability to pronounce the L2 than to amount of L2 experience.
Stimuli with Short-Lag VOT

Regardless of how the NS subjects were grouped (i.e., according to L2 proficiency or experience), they tended to differ from the NE subjects in terms of how stimuli with short-lag VOT values were rated. Averaged across the two VOT continua, the NE subjects’ lower limit scores were 13 ms higher than those of 40 NS subjects’ (43 vs. 30 ms). This means that, on the average, the NS subjects gave a higher rating to the stimuli with a VOT of 30 ms (which is often taken to define the high end of the short-lag VOT range) than did the NE subjects.

To further explore how stimuli with short-lag VOT values were rated, the average rating given to stimuli having VOT values of 10–30 ms were calculated. The two mean values obtained for the NE subjects and for NS subjects differing in L2 proficiency were submitted to a (3) Group x (2) Speaking Rate ANOVA. This analysis yielded a significant two-way interaction \[F(2, 57) = 5.75, p < 0.05\]. The simple main effect of Group was significant for both the fast-rate and slow-rate stimuli \[F(2, 70) = 7.33 and 11.8, respectively, p > 0.05\]. Tukey’s HSD tests revealed that the proficient and nonproficient NS subjects gave significantly higher ratings than did the NE subjects to stimuli with short-lag VOT in both the slow-rate and fast-rate continua (\(p < 0.05\)). The nonproficient NS subjects gave significantly higher rating to stimuli in the slow-rate than fast-rate continuum (6.0 vs. 5.0; \(p < 0.05\)), whereas the NE subjects’ ratings of slow-rate and fast-rate stimuli (2.6, 2.9) and those of the proficient NS subjects (5.3, 4.8) did not differ significantly (\(p > 0.10\)). A similar analysis comparing the average ratings of the NE subjects and the NS subgroups differing in English-language experience revealed that both experienced and inexperienced NS subjects gave significantly higher ratings to short-lag stops in both continua than did the NE subjects.

Average Speaking Rate Effect on Long-Lag Stimuli

The analyses of upper limit scores presented earlier failed to yield significant Group x Rate interactions despite the fact that the NE subjects produced rate effects of larger magnitude than did various subgroups of NS subjects. Recall that the upper limit scores were derived from the average response functions obtained for each subject. An additional analysis was carried out to examine the effect of speaking rate on the goodness judgments. The average difference in ratings given to stimuli with VOT values of 60–120 ms in the slow-rate and fast-rate continua was calculated for each subject. The longest VOT value in the fast-rate continuum was 120 ms; 60 ms was chosen as the lower limit of the VOT range in this analysis because it occurred beyond the peak (i.e., preferred) VOT rating in most subjects’ rating functions.

The difference scores obtained for the NE subjects averaged 1.7. Those obtained from the proficient and nonproficient Spanish subjects averaged 1.7 and 0.7, respectively. A one-way ANOVA examining the difference scores was significant \([F(2, 57) = 3.91, p < 0.05]\). A Tukey’s HSD test revealed that the NE subjects’ difference scores were significantly larger than those of the nonproficient but not the proficient NS subjects (\(p < 0.05\)). The two NS groups did not differ significantly from one another (\(p > 0.10\)).

Relation between Perception and Production

The NS examined here also participated in a production study by Schmidt and Flege [1995]. The question addressed in this section was whether the perceptual results obtained here for the NS subjects were related to their production of English /pl/. The NS subjects produced the phrase A green pea repetitively at self-selected slow, normal, and fast speak-
Table 2. Strength of correlation (Pearson $r$) between perceptual variables obtained here and production variables obtained from the same 40 NS subjects by Schmidt and Flege [1995]

| Perception variables | Production variables | VOT in normal-rate speech | VOT in fast-rate speech | normal vs. fast-rate difference | Correlation between VOT and speaking rate$^a$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate effect on lower limit scores</td>
<td>0.392</td>
<td>0.361</td>
<td>0.228</td>
<td>0.301</td>
<td></td>
</tr>
<tr>
<td>Rate effect on preferred VOT scores</td>
<td>0.201</td>
<td>0.248</td>
<td>-0.024</td>
<td>-0.192</td>
<td></td>
</tr>
<tr>
<td>Rate effect on upper limit scores</td>
<td>0.112</td>
<td>0.109</td>
<td>0.052</td>
<td>-0.085</td>
<td></td>
</tr>
<tr>
<td>Average rate effect on long-lag stimuli</td>
<td>0.303</td>
<td>0.295</td>
<td>0.141</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>Average rating of short-lag stimuli</td>
<td>-0.036</td>
<td>-0.106</td>
<td>0.141</td>
<td>0.129</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Mean values were obtained for each NS subject's production of /p/ in the English word pea spoken in a carrier phrase at normal and fast speaking rates. The strength of correlation between the NS subject's phrase duration and VOT of /p/ provided an index of the extent to which they adjusted their production of /p/ when changing speaking rate.

Mean normal and fast-rate VOT values for each subject were based on six measurements per rate. In addition, the extent to which each of the NS subjects modified VOT in /p/ when changing speaking rates was estimated by determining the strength of correlation between the subject's phrase duration and VOT for /p/. The correlations obtained for the NE and proficient NS subjects were significantly larger than the nonproficient NS subjects' correlations.

Table 2 presents the correlations (Pearson $r$) between perceptual variables from the present study and production variables obtained by Schmidt and Flege [1995]. Small but significant correlations ($p < 0.05$) were found to exist between the rate effect on perceptual lower limit scores (i.e., the difference in the shortest VOT values rated as 'good' at the simulated slow and fast speaking rates) and the VOT values produced in /p/ by the 40 NS subjects at a normal rate and at a fast speaking rate. The longer (more English-like) were the VOT values produced by the NS subjects, the greater was their difference in goodness ratings for stops found in the two perceptual continua. Somewhat surprisingly, the differences in VOT values that were produced in stops at a fast versus a slow speaking rate were not correlated significantly with the perceptual lower limit scores ($p > 0.10$).

The correlations between the average perceptual rate effects for long-lag stops and VOT production variables were marginally significant ($p < 0.10$). However, the correlation between the difference in VOT values in stops produced at a normal versus fast speaking rate, on the one hand, and the average perceptual rate effect on the ratings of long-lag stops, on the other hand, was again nonsignif-
The relation between production and perception was further explored by plotting perception scores from the present perception experiment and production data obtained by Schmidt and Flege [1995]. The two sets of data examined were those that showed the strongest correlation of any considered above. The variables examined were (1) the perceptual rate effect on lower limit scores, which represented a difference in ratings obtained in the normal versus fast-rate continua and (2) the VOT values produced in /p/ at a normal rate of speech. The correlation between these variables was \( r = 0.392 \) for the 40 NS subjects (\( p < 0.01 \)).

The perceptual scores obtained for all 60 subjects were converted to Z scores, as were the mean VOT values obtained from the 60 subjects. Figure 8a plots the perception Z scores (x axis) against the production Z scores (y axis). The upper right-hand quadrant of figure 8a is labeled ‘Good Production and Perception’. This is because the data points located in this quadrant had Z scores that were greater than average (i.e., a Z score of 0) for both production and perception. The data for most of the NE subjects (filled circles) cluster in this quadrant. Data for 9 native Spanish subjects (unfilled circles) are located in the lower right-hand quadrant, which indicates relatively more English-like perception than production. The data for 7 native Spanish subjects are located in the upper left-hand quadrant, indicating relatively more English-like production than perception.

The data in figure 8a do not support the view that perception of L2 phonetic segments is more native-like than the production of those phonetic segments. The possibility existed, however, that the inclusion of data for the NE subjects in the computation of the Z scores may have obscured differences among the 40 NS subjects. Accordingly, the analysis just described was performed a second time on
data obtained from just the 40 NS subjects. These results are presented in figure 8b. The data points for 10 NS subjects are located in the lower right-hand quadrant (more English-like perception than production), whereas the data for 5 NS subjects are located in the upper left-hand quadrant (more English-like production than perception). This provides modest support for the view that perception 'leads' production.

**Discussion**

The question addressed by this research was whether NS speakers who learn English as an L2 in adulthood establish a phonetic category representation for English /p/ in long-term memory. In the 'Introduction', three criteria were proposed as a test for the establishment of a new, English /p/ category. One criterion was that the NS subjects give their highest rating to stimuli having much the same VOT values as NE subjects, rather than to stimuli with the short-lag VOT values typical for Spanish /p/. This criterion was fulfilled by 'proficient' NS subjects who pronounced English sentences relatively well, and also by non-proficient NS subjects who pronounced English sentences more poorly. As discussed earlier, however, the 'preferred' VOT values of some or all NS subjects might have been determined, at least in part, by the range of available VOT values.

Another criterion for category formation was that the NS subjects show an effect of simulated speaking rate on their goodness ratings. Smaller speaking rate effects were anticipated for the NS than NE subjects because Spanish /p/ is realized with short-lag VOT values [Lisker and Abramson, 1964], and because speaking rate changes have a greater influence on long-lag than short-lag stops [Miller et al., 1986; Schmidt and Flege, 1995]. The effect of speaking rate was assessed by having the subjects rate the goodness of stops in relatively short-duration and long-duration CV syllables. The effects of speaking rate obtained for the NE and proficient NS subjects on the perceived goodness of stimuli with long-lag VOT values did not differ significantly. However, the non-proficient NS subjects' rate effects were significantly smaller than the NE subjects'. This finding fulfilled a second criterion for category formation, at least for the proficient Spanish subjects.

A third criterion was that the NS subjects give lower ratings to stops with short-lag VOT (which is typical for Spanish /p/) than to stimuli with VOT values typical for English /p/ (ostensibly, the category being rated for goodness). Both NS groups gave significantly higher goodness ratings to stimuli with short-lag VOT values than did the NE subjects. This suggests that even the proficient NS subjects' judgments of stops as realizations of the English /p/ category were influenced by their representations for Spanish /p/. The proficient NS subjects differed from the NE subjects in one other important respect. Unlike the NE subjects, they did not show a significant effect of speaking rate on 'upper limit' scores. These scores represented the upper limit of the range of acceptable VOT values for English /p/. They may have continued to identify stimuli with English-like VOT values (for /p/) as 'distorted' realizations of the Spanish /p/ category even if they were able to detect auditorily the acoustic difference between Spanish /p/ and English /p/ tokens [see, e.g., Flege and Munro, in press].

If so, this may have prevented the NS subjects from establishing a perceptual category representation for English /p/, and resulted in their having a single /p/ representation based on the acoustic characteristics of /p/ tokens encountered in both Spanish and English. On this view, many NS late learners of English...
would possess a single phonetic category representation for /p/. This single representation would be used to process bilabial stops in both the L1 and the L2. That is not to say that the NS subjects were unable to detect auditorily phonetic differences between English and Spanish /p/. Their original Spanish /p/ category may have evolved so that it embraced the acoustic properties of short-lag Spanish /p/s and long-lag English /p/s.

Another possible interpretation is that it was impossible for some or all NS subjects to ignore their representation for Spanish /p/. Grosjean [1989] has suggested that both the L1 and L2 systems are engaged at all times. On this view, a Spanish category for /p/ would be expected to influence a category for English /p/, assuming that one were established. If so, then the third criterion set up as a test for category formation might be unrealistic.

Given the uncertainty just mentioned, it would probably be imprudent to conclude from the data presented here that the proficient NS subjects had, or had not, established a phonetic category for the long-lag /p/ of English. Another caveat, of course, is the considerable intersubject variability that was observed. Still another caveat has to do with the perceptual stimuli used in the present experiment. Only VOT was manipulated systematically, and so the stimuli did not represent the full range of perceptually important acoustic dimensions that may distinguish English /p/ from Spanish /p/ (see the 'Introduction'). Bohn and Flege [1993] showed recently that certain short-lag stops are often identified as voiceless by English monolinguals; and, conversely, that certain short-lag stops are often identified as voiced by NS monolinguals. Detailed acoustic analyses failed to reveal the acoustic properties responsible for this pattern of perceptual results. Perhaps larger rate effects would have been obtained from the NS subjects in the present study, or lower ratings of short-lag stops would have been given, had natural stimuli been used, or had synthetic stimuli that represented a wide range of relevant acoustic properties been employed [but cf. Miller and Wayland, 1993].

It would be interesting to determine in future research if NS subjects who provide evidence of having established a perceptual representation for English /p/ are more likely to produce English /p/ accurately than those who have not. The perception data from the present study was compared to production data obtained from the same subjects [Schmidt and Flege, 1995]. Correlational analyses provided only weak support for the hypothesis that perception 'leads' production. This pattern of results is in keeping with the view (see 'Introduction') that motoric and perceptual representations are stored separately in long-term memory.

In summary, this experiment examined goodness ratings of word-initial stops as realizations of the English /p/ category by NE and NS speakers who learned English as adults. The NE subjects' results replicated results obtained by Miller and Volaitis [1989] in earlier experiments using the same VOT continua. Ratings increased as VOT values increased beyond the values typical for English /b/, then decreased systematically as VOT increased beyond the range typical for English /p/. Higher ratings were given to certain stimuli having the same VOT values when they occurred in long-duration CV stimuli (which simulated speech produced at a slow rate) than in shorter CVs (which simulated fast-rate speech). Both NE and NS subjects gave their highest ratings to stimuli having VOT values of about 50 ms. Additional research is needed to determine to what extent, if at all, the 'preferred' VOT values of NS subjects were determined by the range of VOT values present in the fast-rate and slow-rate continua. The effect of speaking rate on goodness rat-
ings obtained from nonproficient but not the proficient NS subjects was significantly smaller than effects obtained from the NE subjects. However, both NS groups gave significantly higher goodness ratings to stimuli with short-lag VOT values than did the NE subjects. Thus the proficient NS subjects, as a group, failed to satisfy one of three criteria set up to test for the formation of a phonetic category for English /p/. Additional work will be needed to determine if adult Spanish learners of English do establish perceptual phonetic categories for English /p t k/ and, if so, to what extent their English representations differ from those of NE speakers.

Acknowledgments

This research was supported by NIH grant DC00257. The authors thank J. Miller for supplying the VOT stimuli used here, and W. Strange and N. Takagi for comments on an earlier version of the article.

References


Massaro, D.: Speech perception by ear and eye (Lawrence Erlbaum, Hillsdale 1987).


