Age of Learning Affects Rate-Dependent Processing of Stops in a Second Language

Abstract

The aim of this study was to assess the effect of speaking rate changes on the perception of English stop consonants by four groups of subjects: English and Spanish monolinguals, ‘early’ Spanish/English bilinguals who learned English in childhood, and ‘late’ bilinguals who learned English in adulthood. Subjects identified, and then later rated for goodness as exemplars of the English /p/ category, the members of two voice onset time (VOT) continua. The English monolinguals identified a well-defined range of VOT stimuli as English /p/, and stimuli with longer VOT values as ‘exaggerated’ instances of English /p/. Their goodness ratings increased as VOT increased, then showed a systematic decrease as VOT began to exceed values typical for English /p/. The English monolinguals’ goodness ratings also varied systematically as a function of speaking rate, which was simulated in the two continua by varying syllable duration. The Spanish monolinguals, on the other hand, failed to consistently identify any of the stimuli as English /p/. Although speaking rate influenced their goodness ratings, the Spanish monolinguals’ rate effects differed significantly from the English monolinguals’. The early bilinguals resembled the English monolinguals, and differed from the Spanish monolinguals to a greater extent than did the late Spanish/English bilinguals. This was taken as support for the hypothesis that early bilinguals are more likely than are late bilinguals to establish new phonetic categories for stop consonants in a second language.
Introduction

The stops /p t k/ are realized as voiceless unaspirated stops having short-lag voice onset time (VoT) values in Romance languages such as Spanish, but as voiceless aspirated stops with long-lag VoT values in Germanic languages such as English [Lisker and Abramson, 1964; Williams, 1977; Flege and Eefting, 1986]. The questions addressed by this study were these: Is the phonetic difference between Spanish /p/ and English /p/ sufficient to permit Spanish/English bilinguals to establish separate phonetic representations for Spanish /p/ and English /p/? If so, does the age at which English is learned affect the likelihood of phonetic category formation for English /p/?

Previous research has documented attempts by native speakers of Romance languages such as Spanish to negotiate the differences between /p t k/ in their L1 and English. Nonnatives who learn English as an L2 often produce English /p t k/ with shorter VoT values than do English monolinguals [e.g., Carmazza et al., 1973; Williams, 1979b; Flege and Hillenbrand, 1984; Flege and Eefting, 1988; Major, 1987, 1992; Nathan, 1987; Schmidt, 1988; Flege, 1987, 1991; Mack, 1990; Flege et al. 1995a]. At least two variables influence the extent to which nonnatives’ productions of English /p t k/ diverge from English phonetic norms. Greater divergences are likely for individuals who begin to learn English in adulthood (‘late bilinguals’) than in childhood (‘early bilinguals’). Also, among late bilinguals, accuracy may vary according to amount of English-language experience, although experience with English does not in itself guarantee accurate production. Indeed, if one considers the data now available for late bilinguals who are highly experienced in English, one can observe the production of English /p t k/ with VoT values ranging from Spanish-like short-lag VoT values to English-like long-lag VoT values [e.g., Flege, 1987, 1991; Schmidt, 1988; Schmidt and Flege, 1995; see also Flege et al., 1995a, for Italian/English bilinguals].

The cross-language differences in /p t k/ which give rise to differences in production also give rise to less immediately obvious differences in perception. Spanish/English bilinguals, especially late bilinguals, tend to label stimuli with relatively short VoT values as /p t k/ more often than do English monolinguals [e.g., Abramson and Lisker, 1973; Elman et al., 1977; Williams, 1979a,b, 1980; Flege and Eefting, 1988]. When taken together with the production results mentioned earlier, this raises the issue of whether Spanish/English bilinguals who differ in chronological age at the time they begin to learn English as an L2 also differ in ability to establish long-term memory representations for English /p t k/.

The present study made use of the speaking rate paradigm developed by Miller and her colleagues to assess the perception of English /p/ [e.g., Miller and Volaitis, 1989; Volaitis and Miller, 1992]. This paradigm is based on the observation that English monolinguals’ production of /p t k/ varies as a function of speaking rate. As English speakers increase their rate of speech, the VoT in word-initial tokens of /p t k/ shortens. Conversely, as speaking rate slows, VoT in word-initial voiceless stops lengthens [e.g., Miller et al., 1986; see also Pind, 1995, for Icelandic]. Rate effects on the production of word-initial stop consonants are mirrored in perception, as shown in experiments in which subjects are asked to rate for goodness consonant-vowel (CV) stimuli differing in VoT. In these experiments, a speaking rate difference is simulated by varying the duration of the CV stimuli. For English monolinguals, goodness ratings first increase systematically as VoT increases, then show an orderly decrease as stimulus
VOT values begin to exceed those typical for English /p/. Importantly, stimuli with longer VOT values are judged to be acceptable tokens of English /p/ in long-duration ('slow-rate') stimuli than in shorter-duration ('fast-rate') CV stimuli. (See Summerfield [1981] and Repp and Lin [1991] for evidence of other types of contextual effects on listeners’ stop voicing judgments.) Evidence for rate-dependent processing has been taken as evidence for internal phonetic category structure, and as support for the view that speech perception is 'finely tuned to the consequences of articulation' [Miller and Wayland, 1993].

Changes in speaking rate seem to influence the production of /p t k/ differently in Spanish than English. Schmidt and Flege [1995] found that Spanish monolinguals produced /p/ with only slightly longer VOT values in slow-rate compared to normal-rate utterances. And when speaking more rapidly than normally, the Spanish monolinguals were observed to produce /p/ with slightly longer VOT values, rather than with the shorter VOT values seen in English. These cross-language differences in production set the stage for the present speech perception study.

Spanish speakers – even those with little experience in English – might become aware that /p t k/ are produced with longer VOT values in English than in Spanish [Bohn and Flege, 1993]. Given only superficial knowledge of the phonetic difference between corresponding Spanish and English stops, Spanish speakers might alter their judgments of /p t k/ postperceptually in a forced-choice test. In a goodness rating experiment, they might rate as 'good' exemplars of the English /p/ category any stimuli having longer VOT values than are typical for Spanish /p/. However, it is unlikely that Spanish/English bilinguals could show the same pattern of rate-dependent processing as English monolinguals in a goodness rating experiment unless they had established a central phonetic representation for English /p/ in long-term memory.

Flege and Schmidt [1995] examined rate-dependent processing of English /p/ by two groups of late Spanish/English bilinguals whose overall pronunciation of English was relatively good ('proficient') or poor ('non-proficient'). Both groups resembled English monolinguals in showing an increase, then decrease in goodness ratings as VOT increased. However, only the proficient late bilinguals showed rate-dependent processing similar to that of English monolinguals.

The present study compared the performance of 'early' bilinguals, who began to learn English as young children, to that of 'late' bilinguals who began to learn English in adulthood. (The late bilinguals in this study were not differentiated according to proficiency in English, as in our previous study.) We hypothesized that early bilinguals are more likely than are late bilinguals to establish a phonetic category for English /p/. If so, then the early bilinguals should show a pattern of rate-dependent processing that is similar or identical to that of English monolinguals, whereas the late bilinguals should differ from English monolinguals.

Spanish monolinguals have apparently never been tested using Miller’s speaking rate paradigm, but we supposed their performance would resemble that of the nonproficient late bilinguals examined by Flege and Schmidt [1995]. If so, then the Spanish monolinguals examined here might be expected to differ from English monolinguals in at least two ways. They should show significantly smaller effects of the speaking rate manipulation when rating CV stimuli for goodness than would English monolinguals. They should also show a lower limit for acceptable VOT values in fast-rate compared to slow-rate stimuli. That is, based on the recent VOT production results obtained by Schmidt and Flege [1995], we ex-
Table 1. Characteristics of the 15 subjects (roughly half male, half female) in each of four groups

<table>
<thead>
<tr>
<th>Group</th>
<th>CA  (SD)</th>
<th>AOA (SD)</th>
<th>LOR (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English monolinguals</td>
<td>22.9 (5.7)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Early bilinguals</td>
<td>24.5 (5.6)</td>
<td>1.5 (2.3)</td>
<td>23.1 (6.7)</td>
</tr>
<tr>
<td>Late bilinguals</td>
<td>32.2 (5.6)</td>
<td>20.9 (4.4)</td>
<td>11.2 (4.6)</td>
</tr>
<tr>
<td>Spanish monolinguals</td>
<td>26.0 (5.8)</td>
<td>25.4 (5.6)</td>
<td>0.6 (0.3)</td>
</tr>
</tbody>
</table>

Standard deviations are in parentheses.

CA = Chronological age, in years; AOA = age of arrival in the US, in years; LOR = length of residence in the US, in years.

Expected the Spanish monolinguals to show a rate effect that was opposite to that observed previously for English monolinguals.

Method

Subjects (table 1)

Forty of the 60 subjects tested here had participated previously in the production study of Schmidt and Flege [1995]. All 60 were tested individually in a quiet room on the University of Texas campus after passing a pure-tone hearing screening at octave frequencies between 250 and 4,000 Hz. Roughly one half of the 15 subjects in each of four groups were male.

The subjects in two groups will be referred to here as 'monolinguals', although we readily admit that this designation is not accurate in the strictest sense. This term will be used, however, because the subjects in these groups could not carry on even a simple conversation in any language other than their L1. This was verified for the Spanish monolinguals by a native English-speaking experimenter, and for the English monolinguals by a Spanish-speaking assistant.

The English monolinguals were University of Texas students who had never lived outside the United States (US). Eight were from Texas; the remaining 7 were born and reared elsewhere (New York, Connecticut, Massachusetts, New Jersey). Some of them had studied Spanish as an academic subject for several years in high school or college. The Spanish monolinguals came from 9 different countries, including Colombia (n = 4), Mexico (n = 3) and Venezuela (n = 2). The subjects in this group had been living in the US for an average of 7 months at the time they were tested, the maximum being 14 months. They lived in a predominantly Spanish-speaking community in Austin. Some of them had studied English in school before coming to the US.

Subjects in the remaining two groups consisted of Spanish/English bilinguals who differed according to the age at which they were first massively exposed to native-produced English. The early bilinguals had either been born in the US (Texas or California) or had arrived in the US prior to the age of 7 years from Mexico, Honduras or Puerto Rico. The late bilinguals, on the other hand, arrived in the US at an average age of 21 years (range = 15-28 years). Seven of them were from Mexico; the remaining 8 were from El Salvador, Puerto Rico, Ecuador, Argentina, Venezuela, Guatemala, or Colombia.

In recruiting subjects for the late bilingual group, we attempted to locate individuals who had lived in the US for as long as subjects in the early bilingual group. However, we were not successful in matching the two bilingual groups for length of residence (LOR) so that, as in many previously published studies of L2 acquisition [e.g., Bahrick et al., 1994], the age variable was confounded with overall amount of L2 experience as indexed by LOR in a predominantly L2-speaking country. In addition to having been exposed to English earlier in life, the early bilinguals had also lived for a longer time in the US (mean = 23 years, range: 12-39 years) than the late bilinguals (mean = 11 years, range: 5-20 years).

We had reason to suppose, however, that the confound just mentioned would not undermine conclusions that might be drawn concerning the role of age of L2 acquisition on the perception of L2 consonants. First, previous studies have shown [e.g., Flege et al.,
that in nonnative populations made up of individuals who have lived in a predominantly L2-speaking environment for 5 years or more, variations in age of L2 learning generally account for substantially more variance in measures of L2 phonological proficiency than do variations in LOR. Second, a recent study by Flege and Schmidt [1995] showed that LOR exerted little influence on the perceptual measures of interest in the present study. For example, a significant difference between groups of Spanish/English late bilinguals who had lived in the US for averages of 2.1 vs. 9.4 years was not observed for the effect of speaking rate on goodness ratings in stop consonants. Thus, we will refer only to the differing ages of our two groups of bilinguals when presenting and discussing the results of this study.

Stimuli
Cross-language differences in VOT arise primarily from differences in the timing of supraglottal gestures (e.g., the labial constriction for /p/) vis-à-vis a rapid abduction-adduction of the vocal folds [e.g., Abramson, 1977]. Cross-language differences in laryngeal timing are apt to give rise to acoustic phonetic dimensions in addition to VOT. These include the amplitude and duration of the release burst [Williams, 1977; Bohn and Flege, 1993], the amplitude of aspiration following stop release [Repp, 1979], fundamental frequency variations following the release [e.g. Ohde, 1983], the presence vs. absence of low-frequency energy immediately following release [Williams, 1977], F1 onset frequency [Simon and Fourcin, 1978], and amplitude rise time [Munro, 1987]. Some of these acoustic dimensions might influence stop perception by speakers of both English and Spanish [see, e.g., Bohn and Flege, 1993]. However, VOT is known to differentiate Spanish from English stops, and has been used in previous studies of rate-dependent processing, so it was the acoustic phonetic variable manipulated in the synthetic stimuli used here.

The stimuli were drawn from a study by Miller and Voulaitis [1989]. They consisted of a labial stop followed by an /l/-quality vowel. In the Miller and Voulaitis [1989] study, VOT in 17 fast-rate CVs increased in 5-ms steps from 10 to 60 ms, then in 10-ms steps from 60 to 120 ms. VOT in 37 slow-rate CVs increased in 5-ms steps from 10 to 60 ms, then in 10-ms steps from 60 to 320 ms. Given that only a single experimental session was practical, we eliminated 9 of the original slow-rate stimuli (i.e., those with VOT values of 150, 170, 190, 210, 230, 250, 270, 290 and 310 ms).

The CV stimuli were synthesized at 10.0 kHz using the cascade branch of a software synthesizer. A difference in speaking rate was simulated by varying syllable duration: 125 ms for CVs in a ‘fast-rate’ continuum versus 325 ms for CVs in a ‘slow rate’ continuum. The initial stop in each CV contained a 5-ms release burst followed by 5 ms of silence and formant transitions lasting from 20 to 45 ms. VOT was increased by widening the F1 bandwidth and switching from a periodic to an aperiodic noise source at different times with respect to the initial release burst. The F1 transition into the ‘vowel’ portion lasted 20 ms, which meant that the effective 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 CVs having the longest VOT values.

Procedure
Materials written and tape-recorded in English were used to provide instructions to the English monolinguals and the two groups of bilinguals. The Spanish monolinguals were tested through the use of instructions that had been written and tape-recorded in Spanish. (A native Spanish-speaking assistant was also available to answer questions posed in Spanish by the Spanish monolinguals.) After responding to a language background questionnaire, the subjects participated in a speech production task that is reported in a companion article [Schmidt and Flege, 1995]. They then participated in the identification and goodness rating tasks to be reported here.

The subjects were told that the purpose of the study was to investigate how bilinguals produce and perceive /p/ in Spanish and English. Prior to participating, the subjects were given practice using a mouse, which was used to obtain perceptual responses. The CV stimuli described earlier were low-pass-filtered (4.8 kHz) and presented over headphones at a comfortable level using a Toshiba notebook computer equipped with a 16-bit sound card. The interval between each response and the presentation of the next stimulus was 1.5 s.

The 28 slow-rate CVs and the 17 fast-rate CVs were randomly presented 4 times each for identification. The order in which the two continua were presented was counterbalanced across the subjects in each group. A short familiarization phase preceded each continuum. Prior to each, the CVs (either fast-rate or slow-rate) were played twice in order of ascending VOT. The subjects were told that they might hear a /b/ or a Spanish /p/ at the beginning of the sequence, and that they might hear ‘exaggerated,
breathy /p/ sounds' near the end of the sequence. During the identification test, subjects used three labels to identify the CV stimuli: 'b/ or Spanish /p/', 'English /p/', and 'exaggerated/breathy /p*/'. The subjects were required to identify each stimulus with one of the labels provided, and were told to guess if uncertain.

After a short break, members of the two continua were randomly presented 8 times each for goodness ratings. The two continua were presented in the same counterbalanced order as in the identification experiment. Given the small number of random presentations in the rating experiment (n = 8), a different random order was used for each subject in the four groups. The subjects were told that although the computer-synthesized stimuli would not necessarily sound 'just like' sounds from either English or Spanish, they should try to determine which stimuli contained the 'best example' of English /p/, as found in words such as peach and Pete. During the familiarization phase, the stimuli to be rated were presented a single time in order of increasing VOT. During the rating task, subjects were told to rate each stimulus. They used a scale rating from 1 (for a very poor exemplar of English /p/) to 9 (for a very good exemplar of English /p/). The subjects were told explicitly that a CV judged to contain a 'good' example of English /p/ should receive a higher rating than a CV judged to contain a /b/, a Spanish /p/, or an exaggerated English /p/. The subjects were required to rate each stimulus, and were urged to use the entire 9-point scale.

Results

Identification

The number of times each subject used the three labels to identify the final 3 (of 4) random presentations of each stimulus was tabulated. The total number of times the three labels were used by all 15 subjects in each group was also tabulated. The results obtained for the English monolinguals for the fast-rate and slow-rate continua are shown in figure 1. Comparable data for the early bilinguals, late bilinguals, and Spanish monolinguals are shown in figures 2–4.

The first question of interest was whether the subjects who spoke Spanish as a native language would consistently identify any of the VOT stimuli as English /p/. 'Consistent' was defined operationally here as at least 40 of 45 possible identifications (15 subjects x 3 replicate judgments) as English /p/. The English monolinguals and early bilinguals met this criterion for both the slow-rate and fast-rate CVs. The late bilinguals met the criterion for the fast-rate but not slow-rate CVs, although their identification function did show a clear peak at the slow rate. (As for the English monolinguals, the late bilinguals' peak was centered at a VOT value of about 50 ms.) The Spanish monolinguals, on the other hand, used the 'English /p/' label only sporadically. They did not apply this label to any stimulus in more
than 50% of instances. This suggests that, although they were living in Austin at the time of the study, few if any of the Spanish monolinguals had received enough exposure to English to reliably identify stimuli having the properties of English /p/.

Another question of interest was whether the native Spanish-speaking subjects would differentiate stimuli having VOT values acceptable for English /p/ from stimuli with VOT values that were too long for English /p/. The English monolinguals and early bilinguals (fig. 1, 2) stopped using the 'English /p/' label once VOT exceeded a certain value, both for fast-rate and normal-rate stimuli. The late bilinguals and Spanish monolinguals (fig. 3, 4) continued to use the 'English /p/' label for even those stimuli with the longest VOT values in the slow-rate continuum. The Spanish monolinguals, and to a lesser extent the late bilinguals, showed less differentiation of the 'English /p/' and '/p*/' responses for the slow-rate CVs than did the English monolinguals. The Spanish monolinguals' identification data are consistent with the notion that they did not have a clear definition of the upper limit of acceptable VOT values for English /p/.

Visual inspection of figures 1–4 suggests that the Spanish monolinguals applied the 'b/' or Spanish /p/ label more often than did subjects in the other three groups. Indeed, statistical analyses (a significant Group x Speaking Rate interaction in an ANOVA, followed by a Tukey's HSD test) confirmed that the Spanish monolinguals used this label significantly more often (mean = 29.6) than the English monolinguals, early bilinguals, or late bilinguals (mean = 17.4, 18.0, and 19.8, respectively; p < 0.05).

Interpreting this finding is complicated by the fact that the label in question embraced two phonetic categories, /b/ and Spanish /p/. This
unusual procedure was motivated by the possibility that subjects might hear stimuli with short-lag VOT values either as /b/ or as /p/, coupled with the desire to have the subjects in all four groups participate in a comparable, three-alternative forced-choice task. If one assumes that the Spanish monolinguals did not hear the short-lag stimuli as /b/ [Abramson and Lisker, 1973; Williams, 1977; Flege and Eefting, 1986], then the question arises as to why they heard stimuli with longer VOT values as /p/ than did Spanish speakers examined previously [e.g., Williams, 1977]. The range of VOT values present in the stimulus set used here is a likely explanation. VOT values extended from the short-lag range to well beyond values typical for English /p/, but no stimuli with lead VOT values were included. We speculate that the Spanish monolinguals heard Spanish /p/ more often than did the Spanish/English bilinguals (especially the early bilinguals) because they did not have a representation for a long-lag English /p/ to constrain their judgments of stops as Spanish /p/.

Boundaries between the three response alternatives (i.e., '/b/ or Spanish /p/', 'English /p/', '/p*/') were calculated by determining, through linear interpolation, the VOT value of the 50% crossovers. For the English monolinguals and the two groups of bilinguals, 4.4% of the 180 boundary values (3 groups x 15 subjects x 2 boundaries x 2 rates) were missing. These 8 values were replaced by the appropriate cell mean for the group. For the Spanish monolinguals, on the other hand, 23% of the values, all for the boundary between English /p/ and /p*/, were missing. Accordingly, we examined the boundary values for the Spanish monolinguals separately. The mean boundary values for all four groups are presented in table 2.

The crossover from '/b/ or Spanish /p/' to 'English /p/' occurred at higher VOT values for slow-rate than fast-rate CVs for the English monolinguals and the two groups of bilinguals. The crossover from 'English /p/' to '/p*/' also occurred at higher VOT values for slow-rate than fast-rate CVs. The boundary values were submitted to a (3) Group x (2) Speaking Rate x (2) Boundary ANOVA. Averaged over groups, the differences in /b/-/p/ boundaries at the fast and slow rates (mean = 33.1 vs. 39.3 ms) were smaller than those for /p/-/p*/ boundaries (mean = 74.0 vs. 118.5 ms), yielding a significant Rate x Boundary interaction [F(1, 42) = 39.3, p < 0.01]. Of primary interest, however, was the fact that the Group factor did not enter into significant two-way interactions, and the three-way interaction was also nonsignificant [F(2, 42) = 0.82, p > 0.10]. This suggested that the rate manipulation affected labeling of the stimuli by subjects in all three groups in a similar way.

Boundaries between '/b/ or Spanish /p/' and 'English /p/' were available for all 15 Spanish monolinguals. For these boundaries, the dif-

---

**Fig. 4.** Identification responses obtained from Spanish monolinguals (see fig. 1 legend).
Table 2. Mean boundaries between the three-response alternatives in the forced-choice identification experiment

<table>
<thead>
<tr>
<th>Group</th>
<th>/p/ (Spanish /p/)-/p/*</th>
<th>English /p/ -/p/*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fast</td>
<td>slow</td>
</tr>
<tr>
<td>English monolinguals</td>
<td>31 (11)</td>
<td>39 (9)</td>
</tr>
<tr>
<td>(n=15)</td>
<td>(n=15)</td>
<td>(n=15)</td>
</tr>
<tr>
<td>Early bilinguals</td>
<td>35 (6)</td>
<td>38 (6)</td>
</tr>
<tr>
<td>(n=15)</td>
<td>(n=15)</td>
<td>(n=15)</td>
</tr>
<tr>
<td>Late bilinguals</td>
<td>34 (7)</td>
<td>42 (20)</td>
</tr>
<tr>
<td>(n=15)</td>
<td>(n=15)</td>
<td>(n=15)</td>
</tr>
<tr>
<td>Spanish monolinguals</td>
<td>57 (18)</td>
<td>73 (22)</td>
</tr>
<tr>
<td>(n=15)</td>
<td>(n=15)</td>
<td>(n=6)</td>
</tr>
</tbody>
</table>

Standard deviations are in parentheses.

The difference between fast-rate and slow-rate CVs (mean = 56.8 vs. 73.0 ms) was significant [F(1, 14) = 9.67, p < 0.01]. Boundaries between English /p/ and /p/* were available for only 6 Spanish monolinguals. However, the fast-rate vs. slow-rate difference for these subjects (mean = 89.1 vs. 172.5 ms) was also significant [F(1, 5) = 9.92, p < 0.05].

Mean Goodness Ratings

The average ratings given by each subject to the final 7 (of 8) random presentations of CVs in the fast-rate and slow-rate continua were calculated. Figure 5 shows the mean goodness ratings obtained from the English monolinguals and the early bilinguals for the two continua, and figure 6 shows data for late bilinguals and Spanish monolinguals. Each mean in these figures is based on 105 ratings (15 subjects x 7).

As expected, the English monolinguals’ goodness ratings first increased, then decreased systematically in both continua as stimulus VOT values began to exceed values typical for English /p/. The speaking rate manipulation influenced the English monolinguals’ ratings in three ways. The range of VOT stimuli that received relatively high ratings was somewhat wider for the slow-rate than fast-rate CVs. The rating peak occurred at somewhat higher VOT values for slow-rate than fast-rate CVs. Finally, the goodness ratings decreased less precipitously beyond the peak ratings for slow-rate than fast-rate CVs. As a consequence, stimuli having VOT values of 60–120 ms received higher ratings in the slow-rate than fast-rate continuum. The two groups of bilinguals, and even the Spanish monolinguals, showed the same general pattern. However, on average, the Spanish monolinguals used less of the available 9-point scale than did the English monolinguals. Thus, their rating functions were less sharply peaked in the middle than the English monolinguals’. Somewhat surprisingly, the subjects in all four groups gave their highest ratings to stimuli having VOT values of roughly 50 ms.

A close inspection of figures 5 and 6 nevertheless reveals differences between the four groups. The English monolinguals and early bilinguals gave somewhat lower ratings to the CVs with short-lag VOT values of 0–30 ms than did the late bilinguals and the Spanish monolinguals.

Also, the effect of the speaking rate manipulation, which is evident in the differing rat-
Fig. 5. Mean goodness ratings of stimuli differing in VOT by English monolinguals (a) and early Spanish/English bilinguals (b). The fast-rate and slow-rate stimuli differed in duration (125 vs. 325 ms).

Fig. 6. Mean goodness ratings of stimuli differing in VOT by late Spanish/English bilinguals (a) and Spanish monolinguals (b). The fast-rate and slow-rate stimuli differed in duration (125 vs. 325 ms).

ings for CVs with VOT values of 60–120 ms in the two continua, was somewhat smaller for the late bilinguals and Spanish monolinguals than for the early bilinguals and the English monolinguals.

To assess these differences, the average ratings given by subjects to the 5 stimuli in each continuum that had VOT values within what is traditionally defined as the 'short-lag' VOT range (i.e., 0–30 ms) were computed, as well as an average for 7 stimuli having 'long-lag' VOT values of 60–120 ms. (This latter value represented the upper limit of VOT in the fast-rate continuum). The 4 average ratings obtained for each of the subjects were submitted to a Group (4 levels) x Speaking Rate (fast vs. slow) x VOT (short-lag vs. long-lag) ANOVA, with repeated measures on the last two factors. The significant three-way interaction [F(3, 56) = 3.46, p < 0.05] yielded by this analysis seems to have arisen because the speaking rate manipulation had the same small effect on the short-lag stimuli for the subjects in all four groups, whereas it exerted differing effects across groups for long-lag stimuli.

The difference in ratings for short-lag stimuli in the two continua averaged 0.5 or less for each group. A (4) Group x (2) Speaking Rate ANOVA examining the short-lag stimuli yielded a nonsignificant two-way interaction [F(3, 56) = 0.50, p > 0.10]. (The main effect of Group on ratings of short-lag stimuli also proved to be nonsignificant [F(3, 56) = 1.40, p > 0.10].) The difference in ratings given to long-lag CVs in the two continua was larger for the English monolinguals and early bilinguals (mean = 2.4 and 2.3, respectively) than for the late bilinguals and Spanish monolin-
goals (mean = 1.5 and 1.0, respectively). As a result, a (4) Group x (2) Speaking Rate ANOVA examining the long-lag stimuli yielded a significant two-way interaction \( [F(3, 56) = 4.38, p < 0.01] \).

The simple main effect of Speaking Rate on the average ratings of long-lag stimuli proved to be significant for all four groups (F values ranging from 7.6 to 154.3, d.f. = 1, 14; p < 0.02). This finding suggests that the two-way interaction was due to between-group differences in the size of speaking rate effects. Indeed, a one-way ANOVA examining 'rate effect' scores (i.e., the mean difference between the two continua for stimuli with VOT values of 60–120 ms) yielded a significant effect of Group \( [F(3, 56) = 4.56, p < 0.01] \). According to a Tukey's test, the English monolinguals' and early bilinguals' rate effects were significantly larger than the Spanish monolinguals' \( (p < 0.05) \), but did not differ significantly from the late bilingual's. (No other between-group differences reached significance.)

**Derived Variables**

Additional analyses were carried out to examine three variables derived from the goodness rating functions obtained from individual subjects. These variables, which will be designated the 'preferred VOT', 'VOT lower limit', and 'VOT upper limit' scores, defined the center and range of VOT values that might be said to define a 'good' exemplar of English /p/. Preferred VOT was operationally defined as 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 the average VOT value of those stimuli. The VOT lower limit scores were calculated by multiplying the highest rating given by each subject to the stimuli in a continuum by 0.9, then calculating (through interpolation) the stimulus VOT that coincided with this value. The VOT lower limit scores calculated in this way were always smaller than the preferred VOT scores. A constant of 1.1 was used in calculating the upper limit scores, which always specified VOT values that were higher than the preferred VOT.

Unlike the average ratings obtained for CV stimuli defined by ranges of VOT values (see above), it was not possible to obtain preferred VOT, VOT lower limit, and VOT upper limit scores for each subject. This is because the rating functions obtained from certain subjects were not as orderly as the grouped mean values shown in figures 5 and 6. The two authors independently inspected plots of the rating functions prepared for each subject. All three scores that might be derived from a function were excluded if, according to the judgment of both authors, the observed pattern diverged from the expected up-down pattern in one of the following ways: (a) a subject gave uniformly high ratings to all stimuli, (b) a bimodal rather than unimodal pattern of ratings was evident, (c) ratings increased or decreased steadily, or (d) there was either no increase in ratings at the beginning, or no decrease in ratings at the end of the function. Patterns a and b made it impossible or unreasonable to calculate preferred VOT, and patterns c and d prevented calculation of upper and/or lower limit scores.

Far more data sets obtained from the Spanish monolinguals (15, or 50%) than from the English monolinguals (3, or 10%) met one of these criteria and so were declared missing. The large amount of missing data for the Spanish monolinguals is not surprising given that they were being asked to make judgments of an unfamiliar phonetic category. (The missing data, in fact, lend support to our decision to refer to these subjects as Spanish monolinguals.) For the early and late bilinguals, 18% of the data sets (late bilinguals \( n = 5 \), early learners \( n = 6 \)) were declared missing. We attribute the
Table 3. Mean (and standard deviations) for three variables derived from goodness rating functions

<table>
<thead>
<tr>
<th>Group</th>
<th>Lower limit</th>
<th>Preferred VOT</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fast</td>
<td>slow</td>
<td>fast</td>
</tr>
<tr>
<td>English monolinguals</td>
<td>40 (11)</td>
<td>44 (12)</td>
<td>61 (14)</td>
</tr>
<tr>
<td>Early bilinguals</td>
<td>38 (10)</td>
<td>41 (5)</td>
<td>54 (6)</td>
</tr>
<tr>
<td>Late bilinguals</td>
<td>38 (15)</td>
<td>33 (15)</td>
<td>63 (16)</td>
</tr>
<tr>
<td>Spanish monolinguals</td>
<td>40 (4)</td>
<td>30 (4)</td>
<td>55 (12)</td>
</tr>
</tbody>
</table>

The means were based on a differing number of observations for the four groups due to missing data (English monolinguals n=13, early bilinguals n=10, late bilinguals n=11, Spanish monolinguals n=5). Data for the Spanish monolinguals were not included in ANOVAs testing for between-group differences (see text).

Lower limit = Lower limit of acceptable VOT values; preferred VOT = VOT value of the highest-rated stimulus in a continuum; upper limit = upper limit of acceptable VOT values (see text).

missing data for subjects who spoke English primarily to the fact that the protocol administered to English monolinguals by Miller and Volaitis [1989] over three sessions was administered to subjects in the present study in a single experimental session. Importantly, we obtained fewer judgments of each stimulus than did Miller and Volaitis (8 vs. 20).

We planned to submit the derived variables to mixed-design ANOVAs in which Speaking Rate served as a within-subjects factor. Thus, if the data for one continuum were missing for a subject, data for the other continuum also had to be included. Complete data sets were available for 13 English monolinguals, 10 early bilinguals, 11 late bilinguals, and just 5 Spanish monolinguals. An $F_{\text{max}}$ test revealed that variances obtained in preliminary (4) Group x (2) Speaking Rate ANOVAs examining the derived variables were nonhomogeneous. This is a serious violation of the assumptions of ANOVA when group size differs substantially, as was the case here. Accordingly, the ANOVAs to be reported examined just the scores obtained for the English monolinguals and the two groups of bilinguals. (For the sake of comparison, the mean values obtained for the 5 Spanish monolinguals are presented along with those of the other three groups in table 3.) To the extent that the missing data affected the outcome of analyses presented below, they were likely to have reduced the size of differences between the English monolinguals and one or both groups of bilinguals. As will be shown, however, the bilinguals did differ significantly from the English monolinguals in a number of important respects.

The VOT lower limit scores obtained from the English monolinguals were somewhat smaller for fast-rate than slow-rate CVs (mean = 40 and 44 ms, respectively). This meant that somewhat shorter VOT values were ‘acceptable’ for English /p/ in the relatively short-duration CVs, which simulated a fast rate, than in the longer-duration CV stimuli which simulated a slower speaking rate. Scores for early bilinguals were also smaller for the fast-rate than slow-rate CVs (mean = 38 and 41 ms), whereas the opposite held true for the late...
bilinguals (mean = 38 and 33 ms) and Spanish monolinguals (mean = 40 and 30 ms).

The VOT lower limit scores obtained for the first three groups were submitted to a Group x Speaking Rate ANOVA, which yielded a significant two-way interaction [F(2, 31) = 4.4, p < 0.05]. The simple effect of Rate was marginally significant for the English monolinguals [F(1, 12) = 3.15, p = 0.10], but clearly nonsignificant for the early bilinguals [F(1, 9) = 0.71, p > 0.10]. The effect observed for the late bilinguals was significant [F(1, 10) = 7.06, p = 0.02] but, as mentioned earlier, it was opposite in direction from the English monolinguals' rate effect.

The preferred VOT scores were smaller for fast-rate than slow-rate CVs for the English monolinguals (mean = 48 vs. 61 ms) and early bilinguals (mean = 52 vs. 54 ms), whereas no difference was obtained for the late bilinguals (51 ms in both instances). As a result, the analysis of the preferred VOT values yielded a significant Group x Speaking Rate interaction [F(2, 31) = 3.78, p < 0.05]. The simple effect of Rate was significant for the English monolinguals [F(1, 12) = 23.7, p < 0.01], but it was nonsignificant for the early bilinguals [F(1, 9) = 0.55] and the late bilinguals [F(1, 10) = 0.01, p > 0.10].

Finally, the VOT upper limit scores were smaller for fast-rate than slow-rate CVs for the English monolinguals (mean = 63 and 81 ms, respectively), early bilinguals (mean = 63 and 81 ms), and late bilinguals (mean = 63 and 69 ms). The ANOVA examining these scores yielded a significant main effect of Speaking Rate [F(1, 31) = 22.6, p < 0.01], but a nonsignificant Group x Speaking Rate interaction [F(2, 31) = 1.56, p > 0.10]. One possible interpretation of this finding is that rate exerted the same effect on the upper limit of acceptable VOT values for English /p/ for all three groups. A consideration of the mean values, however, led us to consider the possibility that the nonsignificant interaction may have been due to sampling error.

As mentioned earlier, we had complete data sets for relatively few subjects (13 English monolinguals, 10 early bilinguals, 11 late bilinguals). The mean VOT upper limit scores for fast-rate and slow-rate CVs observed for the 11 late bilinguals (mean = 63 and 69 ms) were similar to those for the 5 Spanish monolinguals for whom we had data (mean = 62 and 66 ms). We therefore pooled the data for these 16 subjects, and the data for the 23 English monolinguals and early bilinguals. These scores were submitted to a (2) Group x (2) Speaking Rate ANOVA. This analysis of VOT upper limit scores yielded a significant two-way interaction [F(1, 37) = 5.0, p < 0.05]. For the English monolinguals and early bilinguals, the upper limit scores obtained for fast-rate and slow-rate CVs (mean = 63 and 81 ms, respectively) differed significantly [F(1, 22) = 30.5, p < 0.01] whereas those of the late bilinguals and Spanish monolinguals (mean = 62 and 69 ms) did not [F(1, 15) = 2.08, p > 0.10]. The English monolinguals' and early bilinguals' upper limit scores did not differ significantly from those of the late bilinguals and Spanish monolinguals for fast-rate CVs [F(1, 37) = 0.01], but they were significantly higher for slow-rate CVs [F(1, 37) = 5.21, p < 0.05].

Discussion

This speech perception study had two primary aims. One was to test the hypothesis that changes in speaking rate exert less influence on Spanish monolinguals' judgments of CV stimuli differing in VOT than on English monolinguals' judgments. The other aim was to test the hypothesis that 'early' Spanish/English bilinguals who learned English as an L2 in childhood would resemble English monolinguals to a greater extent than 'late' Span-
lish/English bilinguals who were first massively exposed to English in adulthood. The two groups of monolinguals and the two groups of bilinguals identified, and then later rated for goodness as exemplars of the English /p/ category, the members of 2 synthetic continua. CVs in the 2 continua had VOT values that ranged from values appropriate for English /p/ to values exceeding those typical for English /p/. The CV stimuli in the two continua differed in overall duration (125 vs. 325 ms), thereby simulating a change in speaking rate.

**Spanish vs. English Monolinguals**

Work by Miller and Volaitis [1989] and Volaitis and Miller [1992] has shown that perceptual judgments of stop consonants differing in VOT are closely linked to variations in VOT that accompany changes in speaking rate. One possible interpretation of this finding is that the speech perception mechanism is attuned to the consequences of articulation [e.g., Liberman and Mattingly, 1985]. If so, then one question of interest is whether such attunement reflects a general and direct (possibly unlearned) linkage between human speech perception and production, or whether it reflects a learned attunement to language-specific phonetic characteristics of segmental articulation. This question was addressed by examining the effect of simulated speaking rate changes on Spanish and English monolinguals' perception of stops differing in VOT.

Schmidt and Flege [1995] observed that changes in speaking rate had a substantial effect on VOT in English monolinguals' productions of /p/ [see also Miller et al., 1986; Volaitis and Miller, 1992]. However, comparable speaking rate changes had relatively little effect on VOT in Spanish monolinguals' productions of /p/, apparently because /p/ is realized with short-lag VOT values in Spanish. (In English, rate exerts relatively little effect on VOT in short-lag tokens of /h/.) Moreover, English monolinguals produced /p/ with shorter VOT values at a fast compared to a normal rate whereas the Spanish monolinguals lengthened VOT at a fast rate. From this, we hypothesized that the Spanish monolinguals examined here would show less perceptual effect of the speaking rate manipulation than the English monolinguals and might, in certain instances, show effects that were opposite to those observed for the English monolinguals.

These hypotheses were supported. First, the English and Spanish monolinguals differed in identifying stops in the two perceptual continua. The English monolinguals consistently labeled certain stimuli as English /p/, and, as VOT extended beyond values typical for English, they stopped using the 'English /p/' label and began using the '/p*/' label (which signified a 'breathy' or 'exaggerated' token of English /p/). The Spanish monolinguals, on the other hand, did not consistently identify any of the VOT stimuli as English /p/ nor did they differentiate stimuli having VOT values appropriate for English /p/ from stimuli having VOT values that were too long for English. These findings implied that the Spanish monolinguals were not cognizant of the range of acceptable values for English /p/.

This is understandable in light of the fact that they had received little previous exposure to English.

The goodness ratings of the English and Spanish monolinguals also differed. The English monolinguals gave significantly higher ratings to long-lag CV stimuli (VOT = 60–120 ms) in the slow-rate than fast-rate continuum. This finding, which agrees with results obtained previously, corresponds to the differences in VOT one sees in English monolinguals' production of utterances at a relatively slow vs. fast speaking rate [Miller et al., 1986; Volaitis and Miller, 1992]. The Spanish monolinguals also showed a signifi-
cant rate effect for long-lag stimuli. However, their effect was significantly smaller than the English monolinguals', which may reflect cross-language differences in the effect of speaking rate on the production of /p t k/. Schmidt and Flege [1995] found that as speaking rate shifted from normal to slow, English monolinguals showed a much larger average increase in VOT for /p/ than did Spanish monolinguals (mean = 11.0 vs. 1.5 ms).

Analyses were also undertaken to assess subjects' average ratings of short-lag CV stimuli (VOT = 0 – 30 ms). Although the subjects were instructed to rate the stimuli as exemplars of the English /p/ category, we nonetheless expected the Spanish monolinguals to give higher goodness ratings to the short-lag stimuli than the English monolinguals because Spanish /p/ is realized with short-lag VOT. However, neither the main effect of Group nor the Group x Rate interaction were significant. Perhaps the Spanish monolinguals adjusted their goodness judgments based on their limited exposure to English. In support of this, the Spanish monolinguals gave their highest ratings to stimuli with much the same VOT values as did the English monolinguals (although they used a much smaller part of the available scale than did the English monolinguals). Another possibility is that the short-lag stimuli simply did not sound like 'good' realizations of Spanish /p/ to the Spanish monolinguals because synthesis was based on acoustic properties of English /p/ rather than Spanish /p/ [Munro, 1987].

Whatever the explanation, the finding just reported bears on the criteria one might use in testing for phonetic category formation. Flege and Schmidt [1995] suggested that, to be credited with category formation, Spanish/English bilinguals might be required to give low goodness ratings to short-lag stops as exemplars of the English /p/ category. The present findings suggest that such a criterion may be unrealisti-
linguals, on the other hand, the VOT lower limit scores were somewhat higher for the fast-rate than slow-rate continuum (mean = 40 vs. 30 ms).

**Bilinguals vs. English Monolinguals**

The effect of the speaking rate manipulation on the categorization of stop consonants was much the same for the English monolinguals and the two groups of bilinguals. However, one or both bilingual groups differed from the English monolinguals in rating the VOT stimuli. The rating results, when taken together, supported the prediction that the early bilinguals’ perceptual judgments would more nearly resemble the English monolinguals’ than would the late bilinguals’ perceptual judgments. This prediction was derived from the hypothesis that category formation is more likely for early than late bilinguals.

As mentioned earlier, the English monolinguals gave higher ratings to long-lag stimuli (VOT = 60–120 ms) in slow-rate than fast-rate CVs. Their rate effect and that of the early bilinguals was significantly larger than that of the Spanish monolinguals. However, the late Spanish/English bilinguals did not differ significantly from the Spanish monolinguals. Between-group differences were also obtained for VOT lower limit, preferred VOT, and VOT upper limit scores derived from individual subjects’ goodness rating functions.

The English monolinguals’ VOT lower limit scores were significantly shorter for CVs in the fast-rate than low-rate continuum. The early bilinguals showed a nonsignificant trend in the same direction. The late bilinguals, on the other hand, showed a significant difference in the opposite direction, thereby resembling the Spanish monolinguals (see above). These findings for the late bilinguals agree with those obtained by Flege and Schmidt [1995]. In that study, Spanish/English late bilinguals who pronounced English poorly showed a significant rate effect on VOT lower limit scores that was the opposite of English monolinguals’.

The English monolinguals’ preferred VOT scores were significantly shorter for fast-rate than slow-rate CVs. That is, the optimal VOT for /p/ was judged to be smaller in short-duration than in longer-duration CVs. (A similar finding was obtained for English monolinguals by Flege and Schmidt [1995].) The speaking rate effect on the early bilinguals’ preferred VOT scores did not reach significance. The late bilinguals showed no difference whatsoever between fast-rate and slow-rate CVs. Similarly, Flege and Schmidt [1995] found that nonproficient late bilinguals showed virtually no effect of speaking rate on preferred VOT values. (However, a group of more proficient late bilinguals did show a significant rate effect, one which went in the same direction as English monolinguals’.)

Finally, the English monolinguals’ and early bilinguals’ VOT upper limit scores were substantially higher on average (by 18 ms in both instances) for slow-rate than fast-rate CVs, whereas the late bilinguals showed only a small difference (mean = 6 ms) in the same direction. The Group x Rate interaction did not reach significance. On the other hand, Flege and Schmidt [1995] found that both proficient and nonproficient late Spanish/English bilinguals showed significantly smaller rate effects on VOT upper limit scores than did English monolinguals. The apparent difference between the two studies may have been due to sampling error. (Flege and Schmidt [1995] examined a total of 40 late bilinguals.) When we pooled the data for the late bilinguals and Spanish monolinguals, a significant Group x Rate interaction was obtained. For English monolinguals and early bilinguals, the VOT upper limit scores were significantly higher for slow-rate than fast-rate CVs. This perceptual finding is congruent with the effect
of speaking rate changes on VOT in the production of English /p/. However, for the late bilinguals and Spanish monolinguals, the speaking rate effect on VOT upper limit scores was nonsignificant. This may be due to the fact that a decrease in speaking rate leads to smaller VOT increases in Spanish than English [Schmidt and Flege, 1995].

**Theoretical Implications**

These results bear on the issue of whether bilinguals establish phonetic category representations when they encounter L2 stops differing from stops in the L1 inventory. According to the L2 Speech Learning Model (SLM) presented by Flege [1987, 1991, 1992], early bilinguals eventually establish phonetic categories for all L2 'sounds' (i.e., position-sensitive allophones) that are auditorily distinct, as a class, from the closest L1 sound. Late bilinguals, on the other hand, were hypothesized to establish phonetic categories for 'new' L2 sounds that differ substantially from the closest L1 sound, but not for less distant L2 sounds that might be described as 'similar' to sounds in the L1. On the assumption that late Spanish/English bilinguals treat English /p/ as similar to Spanish /p/ rather than as a new sound, the SLM would predict category formation for English /p/ by early but not late Spanish/English bilinguals. This, in turn, would lead one to expect substantial differences between early and late bilinguals in an experiment like the one presented here.

Although we did indeed note differences between the early and late bilinguals (see above), the differences were not as substantial as one might have expected from the SLM. The results obtained here agree better with predictions generated by a revised version of the SLM. The revised model [Flege, in press] treats both the chronological age at which L2 learning commences and perceived cross-language phonetic distances as continuous rather than dichotomous variables (as was the case in the original model). According to this proposal, the likelihood of category formation varies inversely with age of L2 learning, but directly as a function of perceived cross-language phonetic distance. From this, one might expect some but not all late Spanish/English bilinguals to establish a phonetic category for the long-lag /p/ of English. VOT production data consistent with hypotheses of the revised SLM were obtained in a study that examined the production of English /p t k/ by highly experienced Italian/English bilinguals differing in age of L2 learning [Flege et al., 1995a]. Additional support is provided by a consideration of perception and production data obtained from 8 of the late Spanish/English bilinguals examined in this study.

**Fig. 7.** Mean goodness ratings by 4 late Spanish/English bilinguals who produced English /p/ with Spanish-like short-lag VOT values (a) and 4 late Spanish/English bilinguals who produced English /p/ with English-like long-lag VOT (b).
Figure 7 shows the mean goodness ratings obtained from the 4 late learners who produced English /p/ with the most English-like VOT values (of the 10 subjects whose production was examined) and by the 4 subjects who produced English /p/ with Spanish-like VOT values of 13–18 ms [Schmidt and Flege, 1995]. The late learners who produced English /p/ accurately showed a substantially larger rate effect than did those who produced English /p/ with Spanish-like VOT values. If the presence of a perceptual rate effect on VOT goodness judgments is indicative of category formation, then one might take these data as support for the view that some but not all late Spanish/English bilinguals establish a phonetic category for English /p/. However, additional research will be needed before a firm conclusion can be reached. What is needed are studies examining large numbers of late bilinguals (to ensure the requisite range of data patterns) in which both production data and relevant speech perception are obtained.

In summary, the present study showed that, as in speech production, speaking rate exerts a smaller effect on Spanish monolinguals’ perceptual judgments of /p/ than on English monolinguals’ judgments. The difference between Spanish and English carries over to the perception of stop consonants in an L2. Spanish/English bilinguals who were first massively exposed to English as adults, and to a lesser extent Spanish/English bilinguals who began to learn English as children, showed some resemblance to the Spanish monolinguals, even though the subjects were asked to rate VOT stimuli for goodness as exemplars of the English /p/ category. The most notable evidence of cross-language phonetic interference involved the lower limit of acceptable VOT values for /p/. For English monolinguals, the range of acceptable VOT values extended to shorter VOT values in short-duration (fast-rate) than longer-duration (slow-rate) stimuli, whereas the opposite held true for the late bilinguals and Spanish monolinguals.

Acknowledgments

This research was supported by NIH grant DC00257. The authors are grateful to J. Miller for providing the VOT stimuli used here, to R. Diehl for providing space in which to test subjects in Austin, and to Ana Escomel for help locating subjects and preparing test materials.

References


