Revista Alicantina de Estudios Ingleses

nº 17, November 2004

Departamento de Filología Inglesa
Universidad de Alicante

ISSN 0214-4808 ● CODEN RAEIEX

The *Revista Alicantina de Estudios Ingleses* is published yearly by the Department of English at the University of Alicante in volumes of approximately 250 pages. The journal aims to provide a forum for debate and an outlet for research involving all aspects of English Studies.

NATURE AND FORMAT OF THE ARTICLES: The *Revista* would welcome articles of the following kinds: (1) Articles on linguistics and ELT, literature, literary theory and criticism, history and other aspects of the culture of the English-speaking nations. Articles should not exceed nine thousand words in length. (2) Bibliographies of studies on *very* specific topics, providing a brief introduction and a list of basic publications. A concise index of contents may optionally be included. (3) Reviews and review articles on recently published books in the field of English Studies. (4) Poetry translations (English-Spanish and Spanish-English). All articles submitted should follow the guidelines which can be obtained from the following Internet address:

http://www.ua.es/dfing/publicaciones/raei/general/instrucciones.htm

Manuscripts should include an abstract in English of about one hundred words in length. In normal circumstances, the editors will only consider for publication those contributions written in English and recorded on disk. Two print-outs of the contribution should also be included. Articles will only be returned at the authors' express wish, if so requested at the time of submission. All correspondence should be addressed to:

Revista Alicantina de Estudios Ingleses, Departamento de Filología Inglesa, Universidad de Alicante, P. O. Box 99, E-03080 ALICANTE (Spain)

● ADVERTISING: The journal will be pleased to carry advertisements in either full-page (17 x 24 cms. approx.) or half-page (17 x 12 cms. approx.) format. Prices and information are available on request at the above address

● EXCHANGES: The *Revista Alicantina de Estudios Ingleses* will be happy to make exchange arrangements with similar journals in the same field. All such proposals should be made in writing to the above address

● SUBSCRIPTIONS: The price of subscriptions for FOUR issues of the *Revista Alicantina de Estudios Ingleses* is as follows: (1) in Spain, 25€ for libraries and institutions, and 18€ for private subscribers; (2) in countries other than Spain, US $30 for libraries and institutions, and US $25 for private subscribers.

Correspondence on subscriptions should be adressed to:

Marcial Pons Librero
San Sotero 6
28037 MADRID (Spain)
revistas@marcialpons.es
Tel.: +34 913043303
Fax: +34 913272367

Portada: Enrique Pérez
Gabinete de Diseño de la Universidad de Alicante

ISSN: 0214-4808

Depósito Legal: A-22-1989

Edición de: COMPOBELL, S.L. Murcia

Estos créditos pertenecen a la edición impresa de la obra

Edición electrónica:

Espagrafic

Voice Onset Time and Foreign Accent Detection: Are L2 Learners Better Than Monolinguals?

Josep Alba-Salas

Contents

Voice Onset Time and Foreign Accent Detection: Are L2 Learners Better Than Monolinguals?

Josep Alba-Salas
College of the Holy Cross (Worcester, USA)
jalba@holycross.edu

Abstract

Using an audiovisual perception task, two groups of native English listeners (monolinguals, and L2 learners of Spanish) were asked to identify which /p,t,k/ tokens had been produced by native English speakers and which ones by foreigners. The experiment found that self-reported criteria for foreign accent detection tended to be consistent with actual perceptual behavior, and that the L2 learners performed better, both at the group and at the individual level. These results suggest that foreign accent detection is cued by Voice Onset Time differences and taps into both our tacit knowledge of the native segmental norm *and* our implicit awareness of what constitutes a *particular deviation from that norm. Moreover, the results suggest that, contrary to what some have*

assumed, monolinguals are not necessarily more sensitive to foreign accents than second language learners.

1. Introduction

The term *Voice Onset Time* (VOT) refers to the beginning of vocal fold vibration relative to the release of a consonantal closure, and it is relevant to the distinction between voiced and voiceless stops in the world's languages. English realizes the voicing distinction as a contrast between short lag (/b,d,g/) and long lag (/p,t,k/). Thus, English voiceless stops have a long VOT and are aspirated ([p^h, t^h, k^h]) in pre-stressed, syllable-initial position. By contrast, Spanish and other Romance languages make a distinction between voicing lead (/b,d,g/) and short lag (/p,t,k/). Hence, Romance /p,t,k/ have a short VOT and are always unaspirated [p,t,k] (e.g. Abramson and Lisker, 1972; Lisker and Abramson, 1964; Williams, 1977).

Because of these cross-linguistic differences, Spanish-English bilinguals must acquire different timing patterns for each language in order to realize stop consonants authentically in both languages. Research shows that native speakers of Romance acquiring English as a second language (L2) gradually stretch their VOT values to approach the native English norm (Flege, 1991; Nathan, 1987; Nathan, Anderson, and

Budsayamongkon, 1987). Yet research also shows that native Romance speakers who begin learning English in late childhood or after puberty, unlike those who start at an earlier age, are typically unable to produce the L2 stops in a native-like manner. Specifically, these late L2 learners tend to produce English /p,t,k/ with substantially shorter (i.e. more Romance-like) VOT values than those typical for English monolinguals (Caramazza, Yeni-Komshian, Zurif, and Carbone, 1973; Flege, 1984, 1991; Flege and Eefting, 1987; Flege and Hillenbrand, 1984; Flege, Munro, and MacKay, 1995a; Major, 1987; Williams, 1979). This observation is consistent with the finding that age of learning is the single most important predictor of success in the acquisition of overall native-like pronunciation (e.g. Flege, Munro, and MacKay, 1995b; Long, 1990; Oyama, 1976; Piske, MacKay, and Flege, 2001; Tahta, Wood, and Loewenthal, 1981; Thompson, 1991).

Researchers have proposed a variety of explanations for the L2 learner's typical inability to achieve native-like pronunciation. Some have attributed foreign accents to the existence of a critical or sensitive period for speech learning (Lenneberg, 1967; Patkowski, 1990; Penfield and Roberts, 1959; Scovel, 1988; Walsh and Diller, 1981). Others have emphasized motivational, cognitive, psychosocial and affective factors, as well as differences in the type and nature of the L2 experience (Guiora, Beit-Hallami, Brannon, Dull, and Scovel, 1972; Krashen, 1975, 1981; Schumann, 1975). Yet others, particularly phoneticians, have tended to see foreign accents as arising from fossilized articulatory "shortcuts", from an inability to fully isolate the phonetic systems of the L2 and the native language (L1), or from a lack of accurate perceptual targets caused by an age-related decline in the L2 learner's recognition that certain auditorily detectable differences between L1 and L2 sounds are phonetically relevant (e.g. Flege, 1991, 1995; Flege, Frieda, and Nozawa, 1997; Flege et al., 1995b).

These hypotheses have been formulated in the context of a growing body of research on L2 phonetic development and the perception of foreign accents. The traditional paradigm in this line of research involves so-called foreign accent-ratings, where native listeners are asked to rate relatively long stretches of speech –usually sentences— on different scales of foreign accent (e.g. Asher, 1969; Bongaerts, Planken, and Schils, 1995; Flege and Fletcher, 1992; Scovel, 1969; Tahta et al., 1981; Thompson, 1991). In general, these studies have shown that native speakers are quite accurate in distinguishing between native- and foreign- produced utterances.

Besides accent ratings, other approaches have been used in the phonetics literature. For example, Flege and Hammond (1982) asked a group of native English speakers to mimic Spanish-accented English. They found that speakers imitating a Spanish accent produced English /t/ with VOT values that were considerably shorter than those typical for English. These results suggest that native English speakers are tacitly aware of subtle phonetic differences distinguishing native- from foreign-produced stop consonants –differences which apparently constitute a perceptually salient cue to detect a foreign accent. This possibility is supported by Major's (1987) finding that Romance speakers of L2 English produce English voiceless stops with VOT values that correlate directly with their global pronunciation accuracy in the second language (cf. Riney and Takagi, 1999, for a similar claim involving Japanese speakers of English as a foreign language).

The idea that stop consonants alone can cue the perception of a foreign accent has also been supported by Flege (1984) and Flege and Munro (1994). Flege's (1984) study included five experiments that examined the ability of native English listeners to detect foreign accents in English words, sentences and segments spoken by native speakers of French and American English. The first three experiments found that both naive and phonetically trained listeners familiar with French-accented English could detect foreign accents well above chance level regardless of the length of the speech segment. Experiments 4 and 5, both involving only phonetically trained listeners, are particularly relevant here. Experiment 4 explored whether differences between native and nonnative speakers in just the aperiodic portion of /tu/ and /ti/ syllables provided perceptually salient cues for the detection of a foreign accent. To avoid any interference from vowel context, the experiment used hybrid syllables consisting of a constant vowel (/i/ and /u/) and either an English /t/ produced by a native speaker of English or an English /t/ produced by a native French speaker of English. The results showed that these phonetically trained listeners were able to detect those tokens produced by French speakers in 67% of the cases. Finally, experiment 5 neutralized the temporal differences between the stop consonants produced by English and French speakers by presenting only 30 milliseconds of the initial portion of /t/ (including the burst and a portion of aspiration, without the following vowel). Again, these listeners were able to discriminate native- from foreign-produced tokens above chance level, although with substantial individual variation. Flege concluded that listeners apparently develop very detailed phonetic category prototypes against which to evaluate speech sounds occurring in their native language. Moreover, he also concluded that single phonetic segments contain sufficient acoustic information to permit foreign accent detection, although he could not determine whether VOT differences alone sufficed to cue a foreign accent.

More recently, Flege and Munro (1994) found that VOT differences do in fact trigger the perception of a foreign accent. Their experiment considered the relation between VOT in Spanish speakers' production of /t/ in the word *taco* and native English listeners' classifications (as Spanish or English) and foreign accent ratings. Among other findings, the study found that variation in VOT accounted for 54% of the variance in the classifications and 87% of the variance in the ratings, so that longer VOT values tended to be associated with more native-like pronunciation of English /t/.

Like Flege (1984) and Flege and Munro (1994), the present study focuses on the role of VOT differences in foreign accent detection. This study, however, presents three important innovations. First, it uses the entire voiceless stop series (/p,t,k/), instead of just one consonant. Second, the experiment includes a cross-modal audiovisual perception task in which participants heard each /p,t,k/ segment while looking at a visual display of the word from which the token had been excised (see details below). Third, and more importantly, the study explicitly explores any differences between monolingual listeners and listeners who have been exposed to a second language in terms of their ability to detect a foreign accent in individual speech sounds.

Although not tested empirically, the implicit consensus in both the phonetics and the Second Language Acquisition (SLA) literature is that monolinguals should be more sensitive to foreign accents than L2 learners. This view is made explicit in Long's (1990) influential review of the research on maturational constraints on language acquisition. In his critique of several foreign accent rating studies using French-English bilinguals from Canada, Long notes that in cosmopolitan areas like Montreal, Toronto and Ottawa "continuous immigration has resulted in heterogeneous English and French speech communities (...), as well as tolerance for and expectation of within-language variation" (267). These factors, he argues, "may be expected to cause [foreign accent] raters to think twice before rejecting accented English or French as definitely non-native" (267). According to this line of reasoning, individuals who have been exposed to fewer varieties of their native language, including foreign-accented varieties, should be more sensitive to foreign accents than speakers who have been exposed to more varieties.

A similar view is expressed in Flege et al. (1995b) within the phonetics literature. Flege et al.'s study unexpectedly found some important individual differences among native English listeners with respect to their ability to detect foreign accents. The researchers hypothesized that this individual variation could be due to dialectal and even idiolectal differences in the listeners' "representations for segmental and prosodic characteristics of English, or to differences in 'tolerance' ranges for English phonetics structures" (3132). Citing Long's (1990) speculation, Flege et al. suggest that listeners' increased tolerance and thus decreased sensitivity toward foreign accents could be a function of their previous exposure to foreign-accented varieties of their native language. The implication is that our ability to detect a foreign accent is somehow compromised by increased exposure to different varieties of our native language and, by extension, also by exposure to a second language.

The putative superiority of monolingual listeners is in principle consistent with Flege's Speech Learning Model (e.g. Flege, 1995; Flege et al., 1995a, 1995b). One of the hypotheses proposed within this model is that late L2 learners typically fail to develop appropriate perceptual targets for L2 sounds that are phonetically similar to those in the native language (as is the case with /p,t,k/ for Spanish learners of English), so they are forced to use a single phonetic category to process both the L2 sounds *and* the L1 segments. Since, by hypothesis, late L2 learners lack separate phonetic categories for the L2 segments, they must expand their L1 categories to encompass also the sounds of their second language. Following this line of reasoning, it seems reasonable to conclude that by expanding our L1 phonetic categories, late exposure to a second language, like exposure to foreign-accented varieties of our native tongue, 'contaminates' our L1 perceptual targets, potentially decreasing our ability to recognize any departures from the native norm. This would entail that monolingual listeners should be better at detecting foreign accents than (late) L2 learners.

The present experiment tried to explore this possibility by including two groups of native English participants: four monolingual listeners, and four listeners who had learned Spanish as an L2. Since native English speakers have been shown to be sensitive to VOT differences in their language, it was expected that, overall, both groups of listeners would be able to discriminate between foreign- and native-produced consonants. Specifically, the prediction was that the VOT differences found in the /p,t,k/ tokens produced by the English and

Spanish natives in the input would provide all listeners with a perceptually salient cue to detect a foreign accent. Given Long's (1990) and Flege et al.'s (1995b) speculation, however, it was hypothesized that, overall, the monolingual group would be more accurate in discriminating between the native- and foreign-produced tokens than the L2 group.

2. Experiment

2.1. Methods

2.1.1. Participants

The experiment included eight native English listeners divided in two groups: monolinguals, and L2 learners of Spanish (n = 4 each). All listeners were paid for their participation, and none reported a history of hearing disorders. Relevant background data were collected from each participant through a detailed questionnaire administered orally after the experiment.

All listeners were born and raised in the US and had English as the main language of instruction and social interactions from elementary through high-school. None had received any training in phonetics or taken a course that explicitly focused on the pronunciation of English or any other language, nor had they been told explicitly how English /p,t,k/ differ from their Spanish counterparts.

The monolingual group consisted of one male and three females representing two varieties of US English (Northeast and Midwest). It included participants L#1, L#2, L#3 and L#4, with a mean age of 37.25 years (range: 21-45). None had studied a foreign language or had the ability to speak or understand a language other than English.

The group of English speakers of L2 Spanish consisted of three males and one female, with a mean age of 27.5 years (range: 21-32). It included participants L#5, L#6, L#7 and L#8, who represented two varieties of US English (Northeast and Northwest). All four listeners started learning Spanish as a foreign language after age 11. On average, they had received approximately five years of formal instruction in Spanish with native speaker instructors, and they had lived in an environment in which Spanish was the dominant language for an average of 3.81 years (range: 1-5).

The L2 learners rated their current skills in Spanish from ‘good’ to ‘excellent’. Their oral Spanish proficiency was informally assessed with a 5-minute interview with the experimenter, who is a native speaker of Spanish and a trained, though not certified, ACTFL Oral Proficiency interviewer. According to this informal assessment, the oral skills of the L2 learners ranged between intermediate-high on the ACTFL OPI scale for L#6 and L#8 to advanced for L#5 and advanced-high for L#7. Besides Spanish, three L2 learners reported knowledge of at least another L2, but rated their current skills in the corresponding languages as ‘poor’ or ‘very poor’.

With respect to their familiarity with foreign-accented English, all eight listeners reported being primarily exposed to Spanish-accented English during the past six months, though to different degrees (see below). Other foreign accents to which participants were regularly exposed included Chinese (for L#4, L#5 and L#6), French (for L#2 and L#3), Hindi (for L#5), Japanese (for L#5 and L#6), and a variety of African accents (for L#2 and L#5).

To assess their familiarity with Spanish-accented English, participants were asked to indicate how often they were exposed to Spanish accents during a typical day in the six months prior to the experiment, using a scale ranging from ‘never’ to ‘very frequently’. Individual responses were subsequently coded into a numerical value ranging from 0 (no exposure to Spanish accents) to 5 (maximal exposure). Numerical values for the monolingual listeners ranged between 2 and 3, with an average of 2.5. In the L2 group there was far more variability, with a range of 1 to 4 and a mean of 2.25. A two-tailed Mann-Whitney test (chosen because of the small sample size) revealed that the difference between both groups is not statistically significant ($U = 6$, $z = -.619$, $p = .536$). This result suggests that our two listener groups did not differ substantially with respect to their daily exposure to Spanish-accented English during the six months prior to the experiment.

2.1.2. Materials

The materials included a set of English /p,t,k/ tokens produced by six native speakers of American English and six native speakers of Spanish who had learned English as a second language (see below). The consonants were produced in pre-stressed word-initial position in the context of nine English words (*peas*, *par*, *pool*, *tease*, *tar*, *tool*, *keys*, *car* and *cool*) embedded in a carrier sentence ('I say ___ again') and interspersed with a variety of distractor items. The words were chosen so that the stop consonants appeared before three vowels that are relatively similar, though not identical, in English and Spanish: /i,a,u/. To further minimize any effect for vowel context, the consonants were edited out from each word in a manner to be described below. The experiment thus included a total of 108 /p,t,k/ tokens (3 consonants x 3 vowel contexts x 12 speakers).

The six native speakers of Spanish who produced the stimuli represented three varieties of Latin American Spanish (Venezuelan, Puerto Rican and Ecuadorian). All of them were born outside the US to Spanish-speaking parents. Although they began learning English in a formal L2 setting between the ages of 5 and 8, their first systematic exposure to native English input occurred after age 15, when they arrived in the US.

The linguistic profile of the native speakers of Spanish was complemented with an informal assessment of their global accent. This measure was obtained by asking two native English speakers to listen to the carrier sentences produced by each of the six participants and then provide an overall rating of their pronunciation accuracy on a scale ranging from a 'very strong' foreign accent to 'none'. Accents ranged from 'slight' (in one case) to 'strong' (in three other participants).

The speakers were recorded in a sound-proof booth using a cardioid microphone and a high-quality cassette recorder. The stimuli were subsequently digitized on a SUN SPARCstation2 at 11,025 Hz with a low-pass filter setting of 5 KHz. The tokens were subsequently stored as files to be processed by the commercial software package WAVES+/ESPS.

VOT values for each of the 108 /p,t,k/ tokens were measured (in milliseconds, ms) so as to determine if there were any systematic differences between native- and foreign-produced stimuli. VOT was defined as the interval between the release of the burst and the onset of periodicity of the following vowel up to the right of a cursor positioned at approximately the first upward going zero crossing in the vocalic segment. All measurements were first made on the corresponding waveform and then checked on a wide-band spectrogram for further accuracy.

Mean VOT values produced by each individual speaker for each place of articulation were calculated by averaging out the three measurements obtained for each consonant (one measurement for each of the vowel contexts in which the consonant appeared before being edited out). Table 1 shows the VOT values obtained for each group of speakers. As we can see, VOT values produced by native English speakers are longer than those produced by the native Spanish group for each /p/, /t/ and /k/. Importantly, there is no overlap in the VOT values produced by the two groups in each place of articulation. The results of three separate ANOVAs of the VOT values by place of articulation showed that the English natives produced /p,t,k/ with significantly higher VOT values than the Spanish natives in each place of articulation ($F(1, 94) = 310.54$, $p < .001$, for /p/; $F(1, 94) = 150.17$, $p < .001$, for /t/; and $F(1, 94) = 271.1$, $p < .001$, for /k/).

Consonant	SPANISH NATIVES (n = 6)			ENGLISH NATIVES (n = 6)		
	Mean	SD	Range	Mean	SD	Range
/p/	46	21	41 - 50.5	96.5	17.5	93 - 100
/t/	71.5	22.5	66.5 - 76.5	110.5	17	106.5 - 114. 5
/k/	66	21	60.5 - 71	111	12.5	109 - 113
Total /p,t,k/	61	24	41 - 76.5	106	17	93 - 114.5

Table 1. Mean VOT values (in ms) for English /p,t,k/ by speakers' L1.

N.B. All values rounded up to the nearest 0.5 value. The difference between the native Spanish and English groups is significant at $p<.01$ in each place of articulation.

The VOT values produced by the native English speakers here are substantially higher than the average values reported in other studies. For example, Flege, Munro and MacKay (1995a) report an average VOT of 57 ms for /p/, 78 ms for /t/, and 77 ms for /k/ for their native English participants. In Lisker and Abramson (1964), the average values were 58 ms for /p/, 70 ms for /t/, and 80 ms for /k/. However, it is important to note that the VOT values obtained for the native participants here fall within the native speaker range reported in Lisker and Abramson (1964) for /p/ (20-120 ms) and /k/ (50-135 ms), and they are only a few milliseconds higher in the case of /t/ (30-

105 ms). It is thus reasonable to conclude that the higher VOT values reported in this study were mostly due to individual differences. At any rate, what is crucial here is that our English natives produced all three stops with significantly longer VOT values than their Spanish counterparts, so that such differences could trigger the perception of a foreign accent.

In preparation for the perception experiment, the recorded stimuli were computer-edited so as to excise each /p,t,k/ token according to the segmentation criteria mentioned above (i.e. the excised consonants included the interval between the release of the burst and the onset of periodicity of the following vowel up to the right of a cursor positioned at approximately the first upward going zero crossing in the vocalic segment). To control for any differences in intensity, the peak intensities of the tokens were equalized in decibels (dB).

2.1.3. Task and Procedure

The experiment consisted of a cross-modal audiovisual perception task. The participants heard each audio stimulus (the excised /p/, /t/ or /k/) one second after the word from which it had been excised was presented visually on a computer screen. Thus, for example, listeners saw the word *keys* on the screen, and a second later they heard a /k/ token that had been excised from the word /kiz/. There were two repetitions of each token, for a total of 216 trials. The stimuli were presented in randomized order, with no pause between trials. Each listener was tested individually in a sound-treated room, and the audio stimuli were presented binaurally using stereo headphones.

This audio-visual task was chosen for two main reasons. First, by seeing the word in which each /p,t,k/ token had been originally produced, listeners could determine whether the token they were judging was a /p/, a /t/ or a /k/ --a basic, yet important, precaution given the very short duration of the auditory stimuli. Second, and more importantly, the visual cues provided participants with some key information about the phonetic context in which each token originally appeared. Specifically, it showed listeners that the consonants had been produced in word-initial, pre-stressed position before a given vowel context (/i/, /a/ or /u/). It is reasonable to assume that this information allowed listeners to activate the perceptual standards needed to judge context-specific realizations of English stops. As such, the audiovisual task served to minimize the departure from ecological validity involved in judging /p,t,k/ tokens in isolation. Granted, this goal could have also been accomplished by adding synthesized vowels to the /p,t,k/ tokens, so that the listeners would have had more phonetic context to anchor their perceptual judgments. However, the visual feedback had the added advantage of minimizing the possibility that the information provided by the additional vowel might have interfered with foreign accent detection. This possibility is consistent with the observation that even vowels that have been traditionally considered equivalent in English and Spanish, such as /i,a,u/, probably differ phonetically across the two languages (e.g. Flege and Munro, 1994), thus raising the possibility that even adding synthesized (as opposed to naturally-produced) vowels could introduce extraneous information into our foreign accent detection task.

Participants were asked to determine whether the tokens they heard had been produced by native English speakers or foreigners. Instead of a forced-choice paradigm (cf. Flege, 1984) or other rating scales used in previous studies (e.g. Flege et al., 1995b), the task involved four response categories: ‘definitely native’, ‘possibly native’, ‘definitely foreign’, and ‘possibly foreign’. These categories were chosen for two reasons. First, the distinction between ‘definitely’ and ‘possibly’ responses could capture a qualitative distinction in terms of degree of certainty in perceptual judgments. Second, the four categories allowed for a straightforward comparison of total ‘native’ vs. ‘foreign’ responses by collapsing ‘definitely’ and ‘possibly’ responses.

Information about the task was provided via written directions immediately before the experiment. Listeners were instructed to identify the L1 background of the speaker who produced each token by pushing one of four response buttons (four keys on a computer keyboard) after hearing the corresponding sound. The four response choices included, from right to left, ‘definitely native’, ‘possibly native’, ‘possibly foreign’, and ‘definitely foreign’. Thus, for example, if listeners thought that a particular /p/ token was likely to have been produced by a native speaker of English but they were not sure, they would press the key labeled ‘possibly native’. If, on the other hand, they were sure that the token had been produced by a native English speaker, they would push the button labeled ‘definitely native’. Listeners were not told about the relative proportion or the exact number of native and nonnative speakers in the input, nor were they told about the L1 background of the foreigners. To familiarize the participants with the task, there was one practice round immediately preceding the actual experiment. This practice round included 15 randomly presented audio-visual stimuli, with no feedback provided.

The perception task lasted approximately 15 minutes, and it was followed by the oral background questionnaire. For the L2 Spanish listeners, the interview also included a 5-minute conversation in Spanish to informally assess their oral skills in the language. The questionnaire focused primarily on the linguistic background of the listeners, but the last section also explored the criteria they had used to discriminate between native- and foreign-produced tokens. Each participant was asked to identify the foreign accent(s) represented in the input, first with an open-ended question and then by choosing one or more of seven options (Hindi, Italian, Spanish, Thai, German, Russian and Chinese). In addition, participants were asked to imitate the pronunciation of a sample token (*pool*) as it had been pronounced by the natives in the input, and then as it had been produced by the foreigners. Finally, participants listened to two different realizations of /p/ (aspirated [pʰ] and unaspirated [p]) and determined which one sounded more similar to the pronunciation of the native speakers in the input and which one sounded like the /p/ produced by the foreigners by choosing the number associated with each production (1 for [pʰ] and 2 for [p]).

3. Results

3.1. Self-reported Criteria for Foreign Accent Detection

Immediately after the experiment, listeners were asked about the criteria they had used to discriminate between native- and foreign-produced tokens. Table 2 summarizes the responses for the L2 Spanish group.

		L2 SPANISH LISTENERS			
		L#5	**L#6**	**L#7**	**L#8**
Foreign accent in input was similar to...		French	Spanish	Spanish	Spanish
Prompted choice of the speakers' L1*		Italian	Spanish	Spanish	Spanish
Imitation:	native /p/ sounded like...	[p^h]	[p^h]	[p^h]	[p^h]
	foreign /p/ sounded like...	[p]	[p]	[p]	[p]
Choice of	native /p/ sounded like...	[p^h]	[p^h]	[p^h]	[p^h]
	foreign /p/ sounded like...	[p]	[p]	[p]	[p]

Table 2. L2 Spanish listeners' self-reported criteria for foreign accent detection.

*Choices included Hindi, Thai, Spanish, Italian, German, Russian and Chinese.

As Table 2 shows, all four L2 listeners correctly identified Romance as the foreign accent represented in the input. Three of them (L#6, L#7 and L#8) chose Spanish when asked if the accent they had heard reminded them of a familiar foreign accent and also when asked to choose among the seven options given. The fourth L2 learners (L#5) chose French in her response to the first question and then Italian when given the seven-language choice.

When asked to illustrate what native-produced tokens sounded like, all four L2 learners produced clearly aspirated [p^h]. Conversely, they produced unaspirated [p] to imitate the pronunciation of those pool tokens that, in their view, had been produced by the foreigners. Moreover, when given a choice between [p^h] and [p], all four L2 learners chose [p^h] as the closest pronunciation of the tokens they believed had been produced by the native speakers, and [p] for the realization that they associated with the foreigners. In sum, these participants seemed to associate aspirated tokens with the natives and unaspirated items with the foreigners.

The responses for the monolingual group are summarized in Table 3. As we can see, responses in the monolingual group were more heterogeneous. Three listeners (L#1, L#3 and L#4) correctly identified Spanish as the foreign accent both in the open question and when asked to choose among the seven options. Of these three listeners, L#3 reported that he recognized a second Romance accent (French), and L#4 also chose Italian from the seven language choices. The only listener who did not recognize a Romance accent was L#2. This participant was unable to determine whether what she had heard was similar to a foreign accent with which she was familiar, and she incorrectly chose Thai when given the seven options.

		MONOLINGUAL LISTENERS			
		L#1	**L#2**	**L#3**	**L#4**
Foreign accent in input was similar to...		Spanish	?	Spanish, French	Spanish
Prompted choice of the speakers' L1*		Spanish	Thai	Spanish	Spanish, Italian
Imitation:	native /p/ sounded like...	[p^{h}]	[p^{h}]	[p^{h}]	[p^{h}]
	foreign /p/ sounded like...	[p^{h}]	[p^{h}]	[p]	[p^{h}]
Choice of [p^{h}] vs. [p]:	native /p/ sounded like...	[p^{h}]	[p]	[p^{h}]	[p^{h}]
	foreign /p/ sounded like...	[p]	[p^{h}]	[p]	[p]

Table 3. Monolingual listeners' self-reported criteria for foreign accent detection.

*Choices included Hindi, Thai, Spanish, Italian, German, Russian and Chinese.

As Table 3 also shows, monolingual listeners also differed with respect to their self-reported use of aspiration to discriminate between foreign- and native-produced tokens. L#3 apparently used similar criteria as the bilingual participants. In fact, this listener produced [p^h] when asked to imitate what native-produced tokens such as pool sounded like, and [p] for the realization he associated with the foreigners. Moreover, when given a choice between [p^h] and [p], L#3 also chose [p^h] for the pronunciation of the native speakers, and [p] for the foreigners.

This situation contrasts with that of L#1 and L#4, who patterned together in their mixed responses. When asked to imitate the pronunciation of native- and foreign-produced *pool*, they used aspirated [p^h] for both. However, when given a choice between [p^h] and [p], they chose [p^h] for the natives and [p] for the foreigners.

Finally, the fourth monolingual listener (L#2) apparently used criteria that were almost a mirror image of those used by L#3 and the L2 listeners. In fact, L#2 produced aspirated [p^h] when asked to imitate the natives, and highly aspirated [p^h] for the foreigners. Moreover, when given the binary choice, she chose unaspirated [p] for the natives, and aspirated [p^h] for the foreigners.

3.2. Total Percentages Obtained for Each Response Category

As a preliminary step in the analysis, individual percentages were tabulated for each listener by averaging out the responses given to both repetitions of the same token across all three vowel contexts. A second step involved computing the percentage of individual responses falling in each of the four response categories ('definitely foreign', 'possibly foreign', 'possibly native' and 'definitely native'). Table 4 summarizes the results.

LISTENER GROUP		RESPONSE CATEGORY			
		'Definitely foreign'	'Possibly foreign'	'Possibly native'	'Definitely native'
MONOLINGUALS	Mean	6.6**	39.1	49.2**	5.3**
	SD	12.9	21.9	23.9	9.8
L2 SPANISH	Mean	18.7**	36	28.6**	16.7**
	SD	17.2	21.6	19.8	16.9

Table 4. Total percentages obtained by both groups of listeners for each response category.

Double asterisk (**) indicates that the difference between monolingual and L2 learners is significant at $p<.01$.

As Table 4 shows, the monolingual group gave very few 'definitely foreign' and 'definitely native' responses (6.6% and 5.3%, respectively). In fact, over 88% of the monolingual group's responses corresponds to 'possibly' responses. The L2 group also gave a majority of 'possibly' responses, but the total percentage (64.6%) was substantially lower than that for the monolingual group. More importantly, the L2 group gave three times as many 'definitely foreign' (18.7%) and 'definitely native' responses (16.7%) as its monolingual counterpart.

Four separate ANOVAs of response category by listener group revealed that the L2 group gave a significantly higher percentage of 'definitely foreign' responses ($F (1, 286) = 45.996$, $p < .001$) and 'definitely native' responses ($F (1, 286) = 48.338$, $p < .001$), but a significantly lower percentage of 'possibly native' responses ($F (1, 286) = 63.437$, $p < .001$) than the monolingual group. There was no significant difference with respect to 'possibly foreign' responses ($F (1, 286) = 1.436$, $p = .232$).

3.3. Overall Correct Identification Scores

Correct identification scores were calculated both for overall 'native' vs. 'foreign' responses (thereby collapsing 'possibly' and 'definitely' responses within each category) and for 'definitely' only responses (taken to be a more nuanced measure of sensitivity towards foreign accents than ‘possibly’ responses).

Table 5 shows the correct identification scores for overall ‘native’ vs. ‘foreign’ responses for each individual listener and for both groups as a whole. As we can see, overall the L2 group correctly discriminated between native- and foreign-produced tokens in 58.4% of the cases. By contrast, the monolingual group as a whole obtained only 47.3% correct responses. An ANOVA of overall correct identification scores by listener group revealed that this difference is highly significant ($F(1, 286) = 18.858$, $p < .001$).

MONOLINGUAL LISTENERS			L2 SPANISH LISTENERS		
Listener	**Mean**	**SD**	**Listener**	**Mean**	**SD**
L#1	49.5	23	**L#5**	64.8	18.6
L#2	35.2	20.2	**L#6**	57.4	21.6
L#3	53.7	22.9	**L#7**	51.8	18.6
L#4	50.9	21.8	**L#8**	59.7	21.2
Total	47.3**	22.9	**Total**	58.4**	20.4

Table 5. Overall percent correct identification scores, including both ‘definitely’ and ‘possibly’ responses.

Double asterisk (**) indicates that the difference between monolingual and L2 learners is significant at $p<.01$.

As Table 5 also shows, in the L2 group all four listeners obtained above 50% correct responses (range: 51.8 - 64.8%), with three participants above 57%. By contrast, in the monolingual group three participants obtained around 50% correct responses (range: 49.5 - 53.7%), and one participant (L#2) obtained only 35.2% correct responses. It is important to note that although the most accurate monolingual listener (L#3) was almost two points above the least accurate L2 participant (L#7), there was no further overlap between both groups.

The L2 group also performed better as a whole with respect to ‘definitely native’ and ‘definitely foreign’ responses. Table 6 shows the correct identification scores for all ‘definitely’ responses (including both ‘definitely native’ and ‘definitely foreign’) obtained by each individual participant and by the two listener groups as a whole.

MONOLINGUAL LISTENERS			L2 SPANISH LISTENERS		
Listener	**Mean**	**SD**	**Listener**	**Mean**	**SD**
L#1	0.9	3.9	**L#5**	24.5	16.6
L#2	3.2	6.7	**L#6**	27.8	18.7
L#3	2.8	6.3	**L#7**	19	17.9
L#4	11.6	13.7	**L#8**	13.4	15.8
Total	4.6**	9.3	**Total**	21.2**	18

Table 6. Percent correct identification scores for 'definitely' responses.

Double asterisk (**) indicates that the difference between monolingual and L2 learners is significant at $p<.01$.

As Table 6 shows, overall, the L2 learners performed correctly almost five times as often as their monolingual counterparts (21.2% vs. 4.6%, respectively). An ANOVA analysis of percent correct 'definitely' responses by listener group revealed that this difference is highly significant ($F (1, 286) = 96.093$, $p < .001$).

The L2 group also performed better in terms of individual results. Whereas the percentage of correct 'definitely' responses among the monolinguals ranged from 0.9 to 11.6%, their L2 counterparts were correct from 13.4 to 27.8% of the cases. Importantly, there is no overlap in correct scores between both groups.

3.4. Correct Identification Scores by Speaker L1

As a further step in the analysis, overall correct identification scores (including both 'definitely' and 'possibly' responses) were tabulated according to the L1 of the speakers represented in the input (English vs. Spanish). Table 7 displays the percent of total correct identification scores obtained *for the native-produced tokens* by each individual listener and by both listener groups as a whole.

MONOLINGUAL LISTENERS			L2 SPANISH LISTENERS		
Listener	**Mean**	**SD**	**Listener**	**Mean**	**SD**
L#1	63	19.4	**L#5**	56.5	14.2
L#2	37	17.7	**L#6**	55.5	19.8
L#3	63.9	20.8	**L#7**	45.4	14.9
L#4	44.4	19.8	**L#8**	57.4	19.1
Total	52.1	22.4	**Total**	53.7	17.5

Table 7. Overall percent correct identification scores for the native-produced tokens (including both 'definitely' and 'possibly' responses).

N.B. The difference between monolingual and L2 learners is not statistically significant.

As Table 7 shows, overall both listener groups correctly identified the native English speakers in a slight majority of cases. The L2 group performed slightly better than their monolingual counterparts, with 53.7% vs. 52.1% correct, but the difference is not statistically significant (ANOVA of correct identification scores by listener group, $F(1, 142) = .234$, $p = 629$). Within the L2 group we find three participants performing above 55% (range: 55.5 - 57.4%), and one listener (L#7) with only 45.4% correct responses. The monolingual group showed much more individual variation. Indeed, L#1 and L#3 had approximately 63% correct responses, more than five points above the highest percentage obtained by the most accurate L2 learner. On the other hand, L#2 and L#4 performed worse than the least accurate L2 participant, with only 37% and 44.4% correct responses, respectively.

A very different picture emerges when we consider the overall percent correct identification scores obtained *for the foreign-produced tokens*. The individual and group results appear in Table 8. As we can see, overall the monolingual group correctly identified foreign-produced /p,t,k/ in only 42.6% of the cases. By contrast, the L2 group was correct on 63.2% of the cases –a 20.4% difference that is highly significant (ANOVA of correct identification scores by listener group, F (1, 142) = 30.553, p <.001).

MONOLINGUAL LISTENERS			L2 SPANISH LISTENERS		
Listener	**Mean**	**SD**	**Listener**	**Mean**	**SD**
L#1	36.1	18.3	**L#5**	73.1	19.1
L#2	33.3	22.9	**L#6**	59.2	23.7
L#3	43.5	20.7	**L#7**	58.3	20
L#4	57.4	22.3	**L#8**	62	23.4
Total	42.6**	22.7	**Total**	63.2**	22

Table 8. Overall percent correct identification scores for the foreign-produced tokens (including both ‘definitely’ and ‘possibly’ responses).

Double asterisk (**) indicates that the difference between monolingual and L2 learners is significant at p<.01.

A brief look at the individual results shows that in the L2 group all participants performed well above chance level (range: 58.3 - 73.1% correct). By contrast, in the monolingual group only one participant performed above chance level: L#4, with 57.4% correct. The other three participants performed quite poorly, with scores ranging from 33.3 to 43.5%. Importantly, here there is no overlap in scores between both groups.

A comparison of Tables 7 and 8 reveals two important facts. First, all four L2 learners were more accurate when judging foreign-produced than native-produced tokens, whereas the opposite is true for the monolingual group, with the sole exception of L#4. Second, in the L2 group there were three participants who performed above chance level when judging *both* native- and foreign-produced tokens (L#5, L#6 and L#8). By contrast, none of the monolingual listeners obtained over 50% of correct responses for both native- *and* foreign-produced /p,t,k/.

3.5. Correlational Analysis of VOT and Perceptual Judgments

A (Pearson) correlation analysis of VOT and perceptual judgments was conducted in order to ascertain the role that VOT may have played in foreign accent detection. The monolingual group showed a very weak negative correlation between VOT and total ‘native’ responses ($r = -.14$), but the association is not significant ($p = .088$, 2-tailed). By contrast, in the L2 group we find a moderate positive correlation between both variables ($r = .42$), and the association is highly significant ($p < .001$, 2-tailed). Thus, for the L2 participants, ‘native’ responses tend to increase as mean VOT values increase: the more aspirated the tokens, the more 'native' responses they received.

Further analysis of the correlational data revealed important individual differences between monolingual and L2 learners. The results appear in Table 9 (see next page).As Table 9 shows, in the L2 group three participants show a robust positive correlation between VOT and 'native' responses, ranging from .40 for L#8 to .73 for L#5, and in all cases the association is significant at alpha = .05. The fourth bilingual participant, L#7, also shows a positive correlation between VOT and 'native' responses, but the association is weak and non-significant.

MONOLINGUAL		LISTENERS L2 SPANISH LISTENERS	
Listener	**r**	**Listener**	**r**
L#1	- .17	**L#5**	.73**
L#2	- .73**	**L#6**	.43**
L#3	.41*	**L#7**	.08
L#4	.00	**L#8**	.40*

Table 9. Pearson correlation coefficient (r) between VOT and percentage of total 'native' responses by listener.

Asterisk (*) indicates that r value is significant at $p < .05$. Double asterisk (**) indicates that r value is significant at $p < .01$.

As Table 9 also shows, in the monolingual group we do not find any clear pattern. Two participants (L#1 and L#4) show an extremely weak correlation between VOT and total 'native' responses, and the patterns are not significant. The other two listeners do show a significant correlation between both variables, but in opposite directions. On the one hand, L#3 shows a moderate positive correlation between the two variables (r = .41). Similar to what we saw in the bilingual group, then, this participant tended to give more 'native' responses to more aspirated tokens. On the other hand, L#2 shows a strong negative correlation between VOT and 'native' responses (r = -.73). Thus, for this listener the percentage of 'native' responses decreases as aspiration increases.

4. Discussion

As we saw earlier, all four listeners in the L2 group, as well as three monolingual participants (L#1, L#3 and L#4), correctly identified the foreign accent in the input as Spanish- or Romance-accented English. The choice of Spanish or Romance by the L2 participants was to a certain extent predictable, not only because of their own experience as L2 speakers of Spanish, but also because, according to self-report, they were relatively familiar with Spanish-accented English, although to different degrees. The finding that most of the monolingual listeners were also able to identify the foreign accent represented in the input is probably due to the fact that, according to self-report, they were exposed to Spanish accents more often than to any other foreign-accented varieties of English.

Although important, relative exposure to different foreign accents was probably not the only factor. This claim is corroborated by the fact that one of the monolingual listeners (L#2) misidentified the foreign accent in the input as Thai when asked to chose among seven possible options even though she reported that the foreign accent to which she had been typically exposed during the six months prior to the experiment was Spanish. L#2's incorrect choice could have resulted from some inherent properties of the stimuli. In fact, the /p,t,k/ tokens produced by the native English speakers had a great deal of aspiration (an average of 106 ms). It is thus conceivable that L#2 perceived these highly aspirated stops as less natural than the shorter tokens produced by the Spanish natives. This possibility is consistent with the results of Flege and Schmidt's (1995) study of the effects of speaking rate on goodness ratings. Flege and Schmidt found that in the fast rate continuum native English speakers identified English /p/ tokens with VOT values ranging from 110 to 120 ms as exaggerated, atypical realizations. Deprived from any information about the L1 background of the foreigners, and given the lack of any audio vocalic context in which to anchor their perceptual judgments, L#2 could have considered the highly aspirated tokens produced by the English natives as exaggerated, and thus less-native-like, productions. In turn, this confusion could have triggered a response bias favoring /p,t,k/ tokens with shorter VOT values. If correct, this scenario would explain why L#2 gave a majority of 'native' responses to the Spanish speakers and a majority of 'foreign' responses to the English natives.

Regardless of the particular reasons for each listeners' choice of a standard for evaluation, what is important is that those participants who correctly chose Spanish or Romance performed much better in the perception task. This is clearly the case of L#3 (within the monolingual group), and L#5, L#6 and L#8 (in the L2 group), who obtained overall correct identification scores ranging between 53.7 and 64.8%. Granted, identifying the foreign accent in the input did not always guarantee performance clearly above chance level. In fact, L#1, L#4 and L#7 obtained overall correct identification scores between 49.5 and 51.8% even though they correctly chose Spanish or Romance. However, these three participants performed substantially better than L#2, who received only 35.2% correct responses, presumably as a result of her incorrect choice of standard for evaluation.

Another important finding of this experiment is that self-reported criteria for foreign accent detection tended to be consistent with actual perceptual behavior, as evidenced by the results of the correlational analysis of VOT and total 'native' responses. Recall that, during the interview after the experiment, all four L2 learners and one of the monolinguals (L#3) not only produced aspirated [p^h] when imitating what native-produced /p/ sounded like and unaspirated [p] for a foreign-sounding /p/, but also chose [p^h] for the natives and [p] for the foreigners when given a binary choice. These criteria were consistent with the finding that four of these listeners (L#3, L#5, L#6 and L#8) showed a significant moderate positive correlation between VOT and total 'native' responses, so they were giving more 'native' responses to more aspirated tokens. The only exception was L#7, who showed virtually no correlation between both variables.

The general congruence between self-reported criteria and actual perceptual behavior is corroborated by the other three participants (L#1, L#2 and L#4). L#1 and L#4 chose

[p^h] for the natives and [p] for the foreigners in the binary choice, but they produced [p^h] for both natives and foreigners in the imitation task. Apparently, then, these listeners did not uniformly associate more aspiration with a more native-like pronunciation. This is consistent with the fact that they showed no significant correlation between VOT and total 'native' responses.

The case of L#2 is particularly striking. Remember that this listener imitated native-produced /p/ as aspirated [p^h] and foreign-produced /p/ with even more aspiration. Moreover, when given a choice, she incorrectly chose [p^h] for the foreigners and [p] for the natives. In other words, L#2 seemed to associate less aspirated tokens with a native-like pronunciation. This is consistent with the fact that she showed a strong negative correlation between VOT and 'native' responses, so she was giving more 'native' responses to consonants produced with shorter VOT values.

It is crucial to note that the correlation between VOT and total 'native' responses seems to be a reliable predictor of performance in the perception task. In fact, those participants who obtained over 53% correct scores (L#3, L#5, L#6 and L#8) showed a positive correlation between VOT and total 'native' responses. By contrast, those who performed around chance level (L#1, L#4 and L#7, with 49.5 to 51.8% correct) showed no significant correlation, whereas the listener with the worst score (L#2, with only 35.2% correct) showed a negative association. Taken together, these results suggest that, as expected, listeners used VOT differences to discriminate between native- and foreign- produced /p,t,k/.

It is also important to note that even among the most accurate listeners the correlation between VOT and 'native' responses was not perfect. In fact, the highest correlation coefficient between both variables was .73 for L#5, who also obtained the highest correct identification score. This finding suggests that VOT, though critical, was not the only acoustic cue for foreign accent detection. As several researchers have noted, VOT is the dominant cue for the voicing contrast in English and Romance. Nevertheless, this contrast is also cued by differences in (i) the fundamental frequency change immediately following consonantal release, (ii) the acoustic characteristics of the release burst, and (iii) periodic energy in the frequency of F1 (Hazan and Boulakia, 1993; Williams, 1977). It is thus quite likely that these differences may have contributed to the perception of a foreign accent, perhaps in combination with some vowel 'coloring' left in the excised /p,t,k/ tokens.

Perhaps the most important finding of the experiment is that, contrary to what was predicted, the L2 group as a whole was significantly more accurate than its monolingual counterpart in discriminating between native- and foreign-produced tokens (58.4 vs. 47.3% total correct identification scores, respectively). This advantage is impressive considering (a) the intrinsic difficulty of detecting a foreign accent in speech segments that lasted less than a fifth of a second, and (b) the fact that none of the participants had any phonetics training.

Besides its superior overall scores, the L2 group also obtained a significantly higher percentage of correct 'definitely' responses than the monolingual group (21.2 vs. 4.6%, respectively). Moreover, the L2 group was significantly more accurate than its monolingual counterpart when judging foreign-produced tokens, with an average of 63.2 vs. 42.6% correct responses.

L2 participants also outperformed monolinguals in terms of individual results. First, all four L2 learners obtained a higher percentage of correct identification scores than their monolingual counterparts with respect to 'definitely' responses, with no overlap between both groups. Second, whereas the L2 participants obtained from 58 to 73% correct responses when evaluating foreign-produced tokens, only one monolingual participant performed above chance level here (L#4, with 57.4% correct). Third, whereas three bilingual listeners obtained a majority of correct responses for both native- *and* foreign-produced tokens (L#5, L#6 and L#8), none did so in the monolingual group.

The apparent advantage of the L2 participants could be due to two factors. The first one involves their experience as L2 speakers of Spanish. Apparently, such experience provided them with the relevant perceptual criteria to detect the *specific* foreign accent in the input. More specifically, their knowledge of Spanish possibily served to sharpen their awareness of and sensitivity towards the subtle VOT differences distinguishing native English pronunciation of /p,t,k/ from Spanish- or Romance-accented English /p,t,k/.

The second factor that might have contributed to the advantage of the L2 participants has to do with their long-term exposure to Spanish-accented English. Recall that there was no significant difference between the two groups with respect to their daily exposure to Spanish accents during the six months prior to the experiment. However, this study did not consider any possible differences in exposure before that time period. It is possible that in the long run the L2 group had had more sustained contact with Spanish-accented varieties of English than their monolingual counterparts, presumably as a result of their increased opportunities to interact (in English) with Spanish natives.

This study raises two important issues concerning native speakers' ability to detect foreign accents. First, it suggests that our ability to detect foreign accents is based not only on our internalized representation of the native norm for segmental production (cf. Flege, 1984), but also on our tacit knowledge of how certain L1 segments are typically produced by nonnative speakers with a *particular* L1 background. To complete the task successfully, the listeners in this study presumably had to activate their L1 categories *and* the relevant perceptual criteria for the specific foreign accent represented in the input (i.e. Spanish-accented, as opposed to, say, Thai-accented, English). In the absence of any information about the L1 background of the foreign speakers, different listeners resorted to different criteria. L#2's tendency to associate shorter VOT values with a foreign accent proved clearly inadequate. By contrast, the other listeners' association of longer VOT values with a more native-like production provided them with the relevant perceptual standard. As we saw earlier, choosing the appropriate criteria for the particular foreign accent represented in the input yielded a substantially higher percentage of correct responses, but it did not always guarantee performance above chance level. This finding suggests that activating the relevant standard for evaluation is a necessary, but not sufficient condition for foreign accent detection.

Second, this study argues against Long's (1990) and Flege et al.'s (1995b) speculation that monolinguals are better at detecting foreign accents than individuals who have been exposed to a second language. As we saw in the Introduction, Long, Flege and others hypothesize that exposure to foreign languages and different varieties of our native language, including foreign-accented varieties, decreases our sensitivity to any departures from the native norm, possibly by expanding (and hence 'contaminating') our L1 phonetic categories. The present experiment suggests that this hypothesis may be incorrect. Even if we accepted the premise that (late) exposure to a second language leads to an expansion of our internalized segmental representations, it does not necessarily follow that monolingual listeners should be more accurate in detecting foreign accents. Such a conclusion would follow *if and only if* listeners evaluated speech sounds occurring in their native language only by reference to their representation of the native norm for segmental production (i.e. the L1 phonetic category). As we just saw, listeners also seem to tap on their knowledge of how certain L1 sounds are typically realized by nonnative speakers with a specific L1 background. If this conclusion is correct, then individuals who have been exposed to a second language should not necessarily be less sensitive to foreign accents than monolinguals. In fact, as this study suggests, such exposure may sharpen our sensitivity to specific departures from the native norm by providing us with the relevant perceptual criteria to evaluate speech sounds produced by L2 speakers of our native language.

The findings of this study must be qualified in several respects. First, they are based on a small sample size, since the need to control for a wide variety of background factors considerably reduced our participant pool, thus raising the possibility that individual differences may have affected the overall group results. Second, listeners made their judgments based on a very limited amount of perceptual data (i.e. consonants that lasted less than a tenth of a second), so it is unclear whether L2 learners would also outperform monolinguals in tasks including vowel segments or longer speech samples. Third, the study did not include listeners with L2s that differed from the foreign accent represented in the input. Hence, we do not know if the advantage of our L2 participants was due to an increased awareness of crosslinguistic phonetic differences in general, or to familiarity with the specific foreign accent represented in the input. To address these issues, future research should (a) include a larger sample size, (b) use a wider variety of tasks and stimuli, both at the segmental and suprasegmental level, (c) involve listeners with different L2 backgrounds, and (d) probe the relative role of L2 experience and long-term exposure to specific foreign accents.

5. Conclusions

This study examined English listeners' ability to discriminate between native- and foreign-produced English stop consonants. The experiment found that self-reported criteria for foreign accent detection were typically consistent with actual perceptual behavior, thus supporting the claim that VOT differences provide a perceptually salient cue for foreign accent detection. The study also found that L2 Spanish listeners not only obtained a significantly higher percentage of correct identification scores than their monolingual counterparts, but also performed consistently better, both at the group and at the individual level. These results suggest that, contrary to what is often assumed, monolinguals are not necessarily more sensitive to foreign accents than L2 learners.

Works Cited

Abramson, Arthur and Leigh Lisker (1972): "Voice-timing perception in Spanish word-initial stops". *Haskins Laboratories Status Report on Speech Research* 29/30: 15-25.

Asher, James (1969): "The optimal age to learn a foreign language". *Modern Language Journal* 53: 334-341.

Bongaerts, Theo, Brigitte Planken and Erik Schils (1995): "Can late starters attain a native accent in a foreign language? A test of the Critical Period Hypothesis". In D. Singleton and Z. Lengyel, eds., *The Age Factor in Second Language Acquisition: A Critical Look at the Critical Period Hypothesis*. Clevedon, UK: Multilingual Matters, 30-50 .

Caramazza, Alfonso, Grace Yeni-Komshian, Edgar Zurif and E. Carbone (1973): "The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals". *Journal of the Acoustical Society of America* 54(2): 421-428.

Flege, James Emil (1984): "The detection of French accent by American listeners". *Journal of the Acoustical Society of America* 76(3): 692-707.

— (1991): "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language". *Journal of the Acoustical Society of America* 89(1): 395-411.

— (1995): "Second language speech learning: theory, findings, and problems". In W. Strange, ed., *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 237-277.

Flege, James Emil and Wieke Eefting (1987): "The production and perception of English stops by Spanish speakers of English". *Journal of Phonetics* 15(1): 67-83.

Flege, James Emil and Kathryn L. Fletcher (1992): “Talker and listener effects on degree of perceived foreign accent”. *Journal of the Acoustical Society of America* 91(1): 370-389.

Flege, James Emil, Elaina Frieda and Takeshi Nozawa (1997): “Amount of native-language (L1) use affects the pronunciation of an L2”. *Journal of Phonetics* 25(2): 169-186.

Flege, James Emil and Robert Hammond (1982): “Mimicry of non-distinctive phonetic differences between language varieties”. *Studies in Second Language Acquisition* 5(1): 1-17.

Flege, James Emil and James Hillenbrand (1984): “Limits on phonetic accuracy in foreign language speech production”. *Journal of the Acoustical Society of America* 76(3): 708-721.

Flege, James Emile and Murray Munro (1994): “The word unit in second language speech production and perception”. *Studies in Second Language Acquisition* 16(4): 381-411.

Flege, James Emile, Murray Munro and Ian MacKay (1995a): “Effects of age of second-language learning on the production of English consonants”. *Speech Communication* 16(1): 1-26.

— (1995b): “Factors affecting strength of perceived foreign accent in a second language”. *Journal of the Acoustical Society of America* 97(5): 3125-3134.

Flege, James Emil and Anna Schmidt (1995): “Effects of speaking rate changes on native and nonnative speech production”. *Phonetica* 52(1): 41-54.

Guiora, Alexander, Benjamin Beit-Hallahmi, Robert Brannon, Cecelia Dull and Thomas Scovel (1972): "The effects of experimentally induced changes in ego states on pronunciation ability in a second language". *Comprehensive Psychiatry* 13(5): 421-428.

Hazan, Valerie and Georges Boulakia (1993): "Perception and production of a voicing contrast by French-English bilinguals". *Language and Speech* 36(1): 17-38.

Krashen, Stephen (1975): "The critical period for language acquisition and its possible bases". In D. Aaronson and R. Rieber, eds., *Developmental Psycholinguistics and Communication Disorders*. New York: New York Academy of Sciences, 211-224.

— (1981): *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

Lenneberg, Eric (1967): *Biological Foundations of Language*. New York: Wiley.

Lisker, Leigh and Arthur Abramson (1964): "A cross-language study of voicing in initial stops: acoustical measurements". *Word* 20: 384-423.

Long, Michael H. (1990): "Maturational constraints on language development". *Studies in Second Language Acquisition* 12(3): 251-285.

Major, Roy (1987): "English voiceless stop production by speakers of Brazilian Portuguese". *Journal of Phonetics* 15(2): 197-202.

Nathan, Geoffrey S. (1987): "On second-language acquisition of voiced stops". *Journal of Phonetics* 15(4): 313-322.

Nathan, Geoffrey S., Warren Anderson and Budsaba Budsayamongkon (1987): “On the acquisition of aspiration”. In G. Ioup and S. Weinberger, eds., *Interlanguage Phonology*. Cambridge: Newbury, 204-212.

Oyama, Susan (1976): “A sensitive period for the acquisition of a non-native phonological system”. *Journal of Psycholinguistic Research* 5(3): 261-283.

Patkowski, Mark (1990): “Age and accent in a second language: a reply to James Emil Flege”. *Applied Linguistics* 11(1): 73-89.

Penfield, Wilder and Lamar Roberts (1959): *Speech and Brain Mechanisms*. Princeton: Princeton University Press.

Piske, Thorsten, Ian MacKay and James Emil Flege (2001): “Factors affecting degree of foreign accent in an L2: a review”. *Journal of Phonetics* 29(2): 191-215.

Riney, Timothy, and Naoyuki Takagi (1999): “Global foreign accent and Voice Onset Time among Japanese EFL Speakers”. *Language Learning* 49(2): 275-302.

Schumann, John (1975): “Affective factors and the problem of age in second language acquisition”. *Language Learning* 25(2): 209-235.

Scovel, Thomas (1969): “Foreign accents, language acquisition, and cerebral dominance”. *Language Learning* 19(3-4): 245-253.

— (1988): *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. New York: Newbury House.

Tahta, Sonia, Margaret Wood and Kate Loewenthal (1981): "Foreign accents: factors relating to transfer of accent from the first language to a second language". *Language and Speech* 24(3): 265-272.

Thompson, Irene (1991): "Foreign accents revisited: the English pronunciation of Russian immigrants". *Language Learning* 41(2): 177-204.

Walsh, Terrence and Karl Diller (1981): "Neurolinguistic considerations on the optimum age for second language learning". In K. Diller, ed., *Individual Differences and Universals in Language Learning Aptitude*. Rowley, MA: Newbury House, 3-21.

Williams, Lee (1977): "The voicing contrast in Spanish". *Journal of Phonetics 5: 169-184.*

— (1979): "The modification of speech perception and production in second-language learning". *Perception and Psychophysics* 26(2): 95-104.