Perceptual switching in Spanish/English bilinguals

Ocke-Schwen Bohn
English Department, Kiel University, Germany

James Emil Flege
Department of Biocommunication, University of Alabama at Birmingham, Birmingham, Alabama, U.S.A.

Received 30th October 1990, and in revised form 23rd July 1992

Previous studies have shown that bilingual native speakers of languages in which /p t k/ are realized with short-lag voice onset time (VOT) values identify stops with short-lag VOT as voiceless more often in an L1 perceptual “set” than in an L2 (English) set. The purpose of this study was to explore the basis of language set effects. The listeners in Experiment 1 identified nine short-lag Spanish /t/ tokens as “t” or “d”. The Spanish /t/ tokens were identified as “t” more often when they were presented along with short-lag English /d/ tokens than when they were presented along with long-lag English /t/ tokens. Much the same small “phonetic context” effect was observed for Spanish and English monolinguals and for two groups of Spanish/English bilinguals. As expected, native Spanish listeners identified the Spanish /t/ as “t” more often than the native English listeners. The fact that native English listeners often identified the Spanish /t/ as “t” indicated that VOT was not an overriding cue to the voicing feature for them. In Experiment 2, the short-lag Spanish /t/ tokens were presented in Spanish and English perceptual sets. Much the same small “language set” effect was observed for the monolingual and bilingual listeners.

1. Introduction

A persistent question in second language (L2) research is whether bilinguals possess distinct processing systems for their two languages, or whether they process the L2 using a more or less modified native language (L1) system (e.g., Mägiste, 1979; Obler, 1982; Altenberg & Cairns, 1983; Mack, 1984; Grossjean, 1985, 1989; Flege, 1990).
1987, 1988, 1991; Cutler, Mehler, Norris & Segui, 1989, 1992). For example, do L1 speakers of a language that implements the phonological voicing contrast with lead vs. short-lag voice onset time (VOT) use their L1 phonetic criteria for voicelessness when processing voiceless stops in an L2 that has long-lag stops, or do they use L2-appropriate criteria to process stops with non-native modal VOT values?

The present study addressed this issue by investigating whether speakers of two languages use different criteria when identifying the same sounds. The study made use of a set of Spanish and English stops with short-lag VOT. The stops were drawn from Spanish and English because Spanish voiceless stops and English voiced stops are produced with approximately the same (short-lag) VOT values. One might suppose that speakers of English and Spanish have distinct phonetic category representations for stops such as English /d/ and Spanish /t/. If so, then the use of a “Spanish” category for identifying stops would be expected to result in more /t/ than /d/ identifications for a set of Spanish /t/ tokens. The use of an “English” phonetic category, on the other hand, should result in more /d/ than /t/ identifications of the same set of Spanish /t/ tokens.

Five previous studies of voicing perception have addressed this issue through so-called “language set” experiments. In each, an attempt was made to induce listeners to perceive short-lag stops either as do monolingual native speakers of English, or as do monolingual speakers of a language in which short-lag stops are associated with voiceless phonemes (Caramazza, Yeni-Komshian, Zurif & Carbone, 1973; Elman, Diehl & Buchwald, 1977; Williams, 1977a; Flege & Eefting, 1987a; Hazan & Boulakia, 1993). The primary method used to create differing perceptual sets has been to interact with the subjects using their two languages (e.g., English in the English set, Spanish in the Spanish set). In addition, Elman et al. (1977) interspersed English or Spanish filler words among their stimuli, and used English or Spanish carrier phrases as appropriate to the set. Similarly, Hazan & Boulakia (1993) used French or English language precursors in the respective sets. Flege & Eefting (1987a) had their subjects read a long list of sentences in Dutch or English at the beginning of the given language set. They also interspersed Dutch or English questions among the stimuli being identified. The subjects were required to answer the questions to ensure that they actually processed and comprehended the Dutch and English speech material in the appropriate language.

The five studies yielded divergent results. In the Elman et al. (1977) study, English monolinguals heard stops spoken with short-lag VOT values as /b/ more often than did Spanish monolinguals (88% vs. 4%). The Spanish/English bilinguals gave slightly more /b/ responses in the English than in the Spanish set (43% vs. 34%). The authors interpreted this latter finding to mean that the Spanish/English bilinguals used different phonetic criteria in the two languages sets. In support of this interpretation, the authors noted that the magnitude of the shift from the Spanish to the English set was greater for “strong” than for “weak” bilinguals. (“Strong” vs. “weak” was defined in terms of estimated degree of foreign accent in English.) Flege & Eefting (1987a) had native speakers of English and Dutch label the members of a VOT continuum ranging from /da/ to /ta/. Dutch speakers of English identified more stops as voiced in an English than in a Dutch set. The language set effect was highly significant, but the average size of the shift in the location of the /d/-/t/ phoneme boundary was quite small (2.1 ms). Contrary to the finding of Elman et al. (1977), the size of the boundary shift was unaffected by the
subjects’ degree of foreign accent as assessed by a panel of native English-speaking listeners. Hazan & Boulakia (1993) essentially replicated the two studies just cited, showing a small but significant effect of language set on the location of the phoneme boundary. As did Elman et al., Hazan & Boulakia reported some evidence that the boundary shift was larger for strong than for weak bilinguals.

Two other studies failed to show an effect of language set. Neither Caramazza et al. (1973) nor Williams (1977a) found a significant change in voicing judgments as a function of language set. It is uncertain why they did not. One possibility is that the procedures used for creating different language sets, which were less elaborate than those in the studies cited earlier, may have been insufficient to induce the subjects to process stimuli as if they were hearing two different languages. Another possible explanation is that the two studies which reported negative results made use of synthetic stimuli. It may be more difficult to show a shift in stop identification for synthetic stimuli than for naturally spoken words because they do not contain the full range of acoustic cues to stop voicing.

VOT was varied in the synthetic continua used by Caramazza et al. (1973) and Williams (1977a). Even though VOT is an important cue to stop voicing, it is by no means the only—or even the most important—cue to the voicing feature in short-lag stops. The lead vs. short-lag contrast in a synthetic VOT continuum is based on the extent of prevocing. It does not include other cues present in naturally-produced stops, such as differences in burst frequency and amplitude and changes in fundamental frequency (Williams, 1977a, 1979). Similarly, the contrast between naturally-produced stops differing in short-lag vs. long-lag VOT includes cues in addition to VOT, such as aspiration noise after release from stop closure (Winitz, LaRiviere & Herriman, 1975) and the extent and duration of the F1 transition (Stevens & Klatt, 1974; see also Lisker, Liberman, Erickson, Dechovitz & Mandler, 1977).

Previous research suggests that listeners are more sensitive to contrasts between stops with short- vs. long-lag VOT values than between stops with lead vs. short-lag VOT values. The short-lag vs. long-lag distinction may be more robust psychoacoustically than a distinction based on lead vs. short-lag VOT (see Pastore, Ahroon, Baffuto, Friedman, Puleo & Fink, 1977; Williams, 1979; Keating, Mikos & Ganong, 1981). Abramson & Lisker (1972) and Williams (1977b) report that native Spanish listeners may have a secondary discrimination peak in the lag VOT region. This implies sensitivity to a short- vs. long-lag contrast in the absence of language-specific input. Kuhl & Miller (1978) found that the identification functions for synthetic VOT stimuli were nearly identical for chinchillas and native speakers of English, a language in which the phonological contrast between /b d g/ and /p t k/ is based on a distinction between stops with short- vs. long-lag VOT values. Thus the English phonetic contrast between voiced and voiceless stops may have a psychoacoustic advantage over the phonetic contrast found in languages like Spanish.

Based on the above, one might suppose that synthetic VOT continua are inherently biased in favor of a short-lag vs. long-lag phonetic distinction. This may explain why the mean crossovers for bilingual subjects in the studies by Caramazza et al. (1973) and Williams (1977a) were closer to the English phoneme boundary than to the phoneme boundaries in the subjects’ L1s (French, Spanish). An apparent bias in favor of a short-lag vs. long-lag distinction was also evident in the Flege & Eefting (1987a) study. Phoneme boundaries obtained in both the Dutch
and English sets were located near the English rather than near the Dutch boundary (cf. Lisker & Abramson, 1964, 1970; Williams, 1977a,b).

The possibility explored in the present study was that the so-called “language set” effects obtained in previous studies (Elman et al., 1977; Flege & Eefting, 1987a; Hazan & Boulakia, 1993) did not result from the use of two language-specific phonetic criteria for classifying stops as voiced or voiceless. Elman et al. (1977, p. 973) noted that the changes in the frequency of /t/ judgments they observed were much smaller than would be expected if the bilingual listeners had used “separate monolingual category boundaries” in the two language sets. The difference between the rates at which the English and Spanish monolinguals examined by Elman et al. (1977) identified short-lag stops as voiceless (84%) was much greater than the mean shift across language sets for Spanish/English bilinguals (9%). Even the “strongest” bilinguals showed only a 35% shift. Similarly, in the Flege & Eefting (1987a) study, the mean difference between the /d/-/t/ boundaries obtained in the Dutch and English sets was much smaller than the difference in phoneme boundaries for English monolinguals and near-monolingual Dutch subjects (2.1 vs. 17.5 ms).

If the language set effects obtained in previous studies did not result from the application of distinct, language-specific perceptual criteria for identifying stops in the two languages, from what then did they result? Although we have no clear answer at present, it is worth noting that the language set effects observed in previous studies are comparable in magnitude to effects noted in previous perception experiments with monolingual subjects. Such research has shown significant changes in the labeling of speech stimuli as the result of a variety of experimental manipulations:

1. Changes in the relative frequency with which stimuli are presented (e.g., Rosen, 1979).
3. The presence vs. absence of an adaptor or anchor (Eimas & Corbit, 1973; Cooper, 1974a,b,c,d; Cooper & Lauritsen, 1974; Fite, 1977; Miller & Eimas, 1977; Ohde & Sharf, 1979; Roberts & Summerfield, 1981; Samuel, 1982; MacMillan, Goldberg & Braida, 1988).
4. The order in which stimuli are presented (e.g., Fry, Abramson, Eimas & Liberman, 1962; Repp, Healy & Crowder, 1979).

The three experiments with bilinguals reviewed earlier that yielded significant language set effects seem to have had something in common with some of the above-mentioned experimental studies. Changes in monolinguals’ perception of consonants induced through the experimental manipulations listed above have usually been quite small, just like language set effects. For example, the largest difference that Brady & Darwin (1978) obtained by varying the range of VOT values in ensembles of stimuli presented to subjects was a boundary shift of about 7 ms. This is similar in magnitude to the largest language set effect observed by Flege & Eefting (1987a) for any of their subject groups, namely 4.1 ms.
However, while the effects reported by Elman et al. (1977), Flege & Eefting (1987a) and Hazan & Boulakia (1993) were all similar in magnitude to response shifts seen in monolinguals as a result of experimental manipulations, the effects reported in these language set studies cannot be due to the nature of the stimuli. This is because the same set of stimuli was presented in both language sets (i.e., in both the “English” and the “non-English” condition). But, in addition to the stimuli to be judged, the listeners also heard other speech materials in the two conditions that were designed, in part, to create the two language sets. (Elman et al., 1977, used a carrier sentence with short-lag /p/ tokens, Flege & Eefting, 1987a, interspersed questions among their stimuli which contained Dutch and English stops, and Hazan & Boulakia, 1993, used precursors with French /p/ in the French set and with English /p/ in the English set.) The English and the non-English stops in the speech materials may have defined two different “standards”. Potentially, they may have in some way set up different stimulus ranges, or they may have inadvertently created frequency effects. If so, then the possibility arises—which has never before been tested—that much the same effects that have been observed for bilinguals in language set studies might also be obtained for monolinguals. If comparable effects were indeed obtained for monolingual and bilingual listeners, and if the effects were as small as the range and frequency effects seen in studies with monolinguals, it would strongly imply that language set effects need not arise from the application of two different, language-specific norms. Instead, one might conclude that the effects are the results of conceptually determined, top-down processes that are independent of the language-specific criteria.

The purpose of this study was to provide a better understanding of the language set effects obtained in previous research. Experiment 1 tested the effect of varying contextual stimuli on the identification of short-lag Spanish /t/ tokens. In Experiment 2, the same Spanish /t/ tokens were presented in two language sets. Both experiments examined the perception of Spanish/English bilinguals as well as of monolingual control subjects. The monolinguals’ responses were crucial for determining if experimental effects, should they be obtained, were due to the use of phonetic criteria associated with Spanish /t/ and English /t/, or if they were due to post-perceptual, cognitively-based changes in the decision criteria used in labeling stops. A change in the frequency with which Spanish /t/ tokens are identified as voiceless would be considered a genuine “language set” effect only if bilingual but not monolingual subjects gave more /t/ responses in the Spanish than English set.

2. Experiment 1

Stop consonants were identified as “t” or “d” by listeners in four groups differing in L1 and/or L2 experience: English monolinguals, Spanish monolinguals, and two Spanish/English bilingual groups differing in age of first contact with English (i.e., Early vs. Late L2 Learners). The listeners participated in two conditions, in which different sets of stimuli were presented. In both conditions, the listeners heard Spanish /t/ tokens with short-lag and Spanish /d/ tokens with lead VOT values. The two conditions differed in that, in addition to the stimuli just mentioned, one contained tokens of short-lag English /d/, and the other contained tokens of long-lag English /t/. For convenience, we will refer to this as the manipulation of
“phonetic context”, although it should be clear that it is much akin to the manipulation of stimulus range.

An assumption underlying this experiment was that native speakers of Spanish, even those with little or no previous exposure to English, would identify the long-lag English /t/ tokens as voiceless and that, in comparison, the short-lag Spanish /t/ tokens would somehow sound less voiceless. If so, the Spanish /t/s should be labeled “t” less often when juxtaposed to the English /t/s than when juxtaposed to the English /d/s.

We hypothesized that if Spanish speakers of English develop two different perceptual representations for voiceless stops owing to their familiarity with two ways of realizing /t/ (i.e., as a short-lag stop in Spanish, or as a long-lag stop in English), they might show a larger Phonetic Context effect than the native English subjects. If much the same context effect were obtained for the monolingual subjects and for the bilingual subjects, on the other hand, we would be forced to conclude that the Phonetic Context effect was the result of a post-perceptual bias similar to the stimulus range effects seen in many speech perception experiments with monolinguals, not the result of applying two different phonetic criteria. If the Early L2 Learners showed a significantly larger Phonetic Context effect than the Late L2 Learners, it would support the hypothesis (Flege & Hillenbrand, 1984; Flege & Eefting, 1988) that they, but not the Late L2 Learners, had established a phonetic category for the long-lag /t/s of English.

2.1. Method

2.1.1. Stimuli

The 36 stimuli were consonant-vowel (CV) syllables excised from Spanish and English words spoken by adult monolingual native speakers of those two languages. Nine monolingual speakers of Spanish (eight males, one female) from Puerto Rico each contributed a /d/-initial word with lead VOT and a /t/-initial word with short-lag VOT. These words were selected from the corpus for a previous study (Flege & Eefting, 1987b). Most of the English CVs were also taken from words recorded for the earlier study but, to ensure the availability of a short-lag /d/ token and a long-lag /t/ token for each talker, several new talkers had to be recorded. Nine monolingual native speakers of American English (six males, three females) each contributed a /d/-initial word with short-lag VOT and a /t/-initial word with long-lag VOT. The native English speakers were living in Birmingham, Alabama, at the time they were recorded, and none spoke English with a marked regional accent.

The Spanish monolinguals read a randomized word list that included multiple tokens of the Spanish words doma and toma; the English monolinguals read a comparable list with toner and donor tokens. One token of each word was low-pass filtered at 8 kHZ and digitized at 20 kHZ with 12-bit amplitude resolution. The first syllables of the words (phonemically /to/ or /do/) were edited out and normalized for peak intensity. The segmentation criterion was the decrease in amplitude and change in waveform shape that accompanied constriction for the intervocalic nasal consonant (/m/, /n/). All vowels were partially nasalized. No attempt was made to segment the oral from the nasal portions of the vowels.

We did not expect potential variations in the degree of nasalization to influence
voicing judgments. Nevertheless, a preliminary experiment was conducted to determine if the vowels produced by the Spanish and English monolinguals differed in degree of nasalization. The nine short-lag Spanish /t/ stimuli and the nine short-lag English /d/ stimuli were randomly presented six times each to three trained judges (native speakers of English). They rated degree of nasalization in each vowel using a nine-point scale ranging from "not nasalized" (1) to "extremely nasalized" (9). The judges found no difference in nasality between the English and the Spanish stimuli.

The VOT of the stop in each CV was measured to the nearest millisecond using a waveform editor. VOT was measured from the beginning of the release burst to the zero crossing that marked the first glottal pulse of the following "vowel" portion. The mean VOTs for the four sets of nine stimuli were:

1. prevocal Spanish /d/: -94 ms (SD = 30)
2. short-lag English /d/: +17 ms (SD = 5)
3. short-lag English /t/: +21 ms (SD = 5)
4. long-lag English /t/: +84 ms (SD = 12)

All Spanish /d/ tokens had lead VOT values, and all English /d/ tokens had short-lag VOT values. These two sets of CVs will be referred to here as the "S- /d/" and "E- /d/" stimuli, respectively. The voiceless Spanish and English stops will be referred to as the "S- /t/" and "E- /t/" stimuli.

2.1.2. Subjects

Four groups of 10 listeners each participated as paid subjects. Each listener filled out a language-background questionnaire and passed a pure-tone hearing screening (0.5–4.0 kHZ, 20 dB HL). The mean age of listeners in the "English monolingual" group, which consisted of native speakers of American English who reported no knowledge of Spanish, was 22.6 years (SD = 4.9).

Characteristics of the three groups of Spanish listeners are presented in Table I. Some listeners in the "Spanish monolingual" group had received formal instruction in English (M = 3.5 years, range = 0–12 years; SD = 3.9). They came from different countries and had been residing in the U.S.A. for an average of 0.3 years (SD = 0.2) at the time of the study. Although it may not be strictly accurate to do so, we will refer to these subjects as "monolingual" because they were unable to carry on a simple conversation in English. We supposed that the relatively small amount of exposure they may have had to English would not influence their perception of stop consonants.

The two groups of Spanish/English bilinguals differed principally according to the age at which they were first massively exposed to English. These listeners were born in a variety of countries. The Early L2 Learners were native speakers of Spanish who were first exposed to English at a mean age of 3.0 years (SD = 3.0). The Late L2 Learners were first exposed to English in the U.S.A. at an average age of 24.6 years (SD = 7.6). The Early and Late L2 Learners differed little in mean chronological age at the time of the study (28.5 vs. 29.7 years), but the Early L2 Learners reported using English somewhat more on a daily basis (78% vs. 60).

2.1.3. Procedure

Ten randomizations each of the nine S- /d/ and the nine S- /t/ stimuli were presented in two conditions. The listeners in all four groups participated in both
TABLE I. Characteristics of the listeners in the three Spanish speaking groups in Experiment 1. (POB = place of birth; AOA = age of arrival in the United States, in years; EDU = years of formal education in English; LOR = length of residence in the United States, in years; %USE = self-estimated percentage daily use of English; md = missing data)

<p>| Characteristics |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>S#</th>
<th>Sex</th>
<th>Age</th>
<th>POB</th>
<th>AOA</th>
<th>EDU</th>
<th>LOR</th>
<th>%USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>22</td>
<td>Colombia</td>
<td>22</td>
<td>0</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>30</td>
<td>Colombia</td>
<td>md</td>
<td>6</td>
<td>md</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>36</td>
<td>Puerto Rico</td>
<td>36</td>
<td>2</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>20</td>
<td>Mexico</td>
<td>20</td>
<td>md</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>md</td>
<td>Uruguay</td>
<td>md</td>
<td>0</td>
<td>md</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>32</td>
<td>Chile</td>
<td>32</td>
<td>3</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>44</td>
<td>Chile</td>
<td>44</td>
<td>6</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>25</td>
<td>Colombia</td>
<td>24</td>
<td>6</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>30</td>
<td>Colombia</td>
<td>30</td>
<td>2</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>31</td>
<td>Colombia</td>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td>(SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(7.3)</td>
<td></td>
<td>(0.2)</td>
</tr>
</tbody>
</table>

“Late” bilinguals

<table>
<thead>
<tr>
<th>S#</th>
<th>Sex</th>
<th>Age</th>
<th>POB</th>
<th>AOA</th>
<th>EDU</th>
<th>LOR</th>
<th>%USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>31</td>
<td>Mexico</td>
<td>16</td>
<td>5</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>28</td>
<td>Puerto Rico</td>
<td>28</td>
<td>9.5</td>
<td>0.2</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>24</td>
<td>Venezuela</td>
<td>23</td>
<td>4</td>
<td>1.5</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>26</td>
<td>Colombia</td>
<td>18</td>
<td>6</td>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>18</td>
<td>Puerto Rico</td>
<td>18</td>
<td>12</td>
<td>0.1</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>34</td>
<td>Mexico</td>
<td>32</td>
<td>2</td>
<td>2</td>
<td>md</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>31</td>
<td>Mexico</td>
<td>31</td>
<td>9</td>
<td>0.5</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>31</td>
<td>Guatemala</td>
<td>26</td>
<td>4</td>
<td>5.5</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>41</td>
<td>Colombia</td>
<td>38</td>
<td>4</td>
<td>3</td>
<td>md</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>33</td>
<td>Guatemala</td>
<td>16</td>
<td>5</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.7</td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td>(SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6.2)</td>
<td></td>
<td>(17.7)</td>
</tr>
</tbody>
</table>

“Early” bilinguals

<table>
<thead>
<tr>
<th>S#</th>
<th>Sex</th>
<th>Age</th>
<th>POB</th>
<th>AOA</th>
<th>EDU</th>
<th>LOR</th>
<th>%USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>39</td>
<td>USA</td>
<td>(birth)</td>
<td>39</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>21</td>
<td>USA</td>
<td>(birth)</td>
<td>21</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>29</td>
<td>Cuba</td>
<td>2</td>
<td>27</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>30</td>
<td>Cuba</td>
<td>4</td>
<td>26</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>32</td>
<td>Cuba</td>
<td>5</td>
<td>26</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>28</td>
<td>USA</td>
<td>3</td>
<td>25</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>24</td>
<td>USA</td>
<td>(birth)</td>
<td>15</td>
<td>md</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>30</td>
<td>Spain</td>
<td>0.5</td>
<td>29</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>25</td>
<td>Honduras</td>
<td>7.5</td>
<td>(14)*</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>27</td>
<td>Honduras</td>
<td>8</td>
<td>(13)*</td>
<td>13</td>
<td>99</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>28.5</td>
<td></td>
<td>3.0</td>
<td>23.1</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>(SD)</td>
<td></td>
<td>(4.9)</td>
<td></td>
<td>(3.0)</td>
<td>(8.6)</td>
<td>(25.1)</td>
<td></td>
</tr>
</tbody>
</table>

* Subjects 9 and 10 in the early bilingual group started to go to an English-speaking school at the age of 7.5 and 8 years, respectively, and arrived in the U.S.A. at the age of 14 and 13 years, respectively
conditions, in counterbalanced order. In the E-/d/ condition, the 18 Spanish CVs were presented along with the nine E-/d/ stimuli. In the E-/t/ condition, the same 18 Spanish CVs were presented with the nine E-/t/ stimuli. Thus, a total of 270 stimuli (10 repetitions × 3 sets of CVs × 9 tokens in each set) were presented in each condition. Instructions were given in English to all listeners except the Spanish monolinguals, for whom instructions translated into Spanish were read aloud. The listeners identified the stop in each CV as “t” or “d” by pressing one of two buttons (labeled “t” and “d”) on a response box. The interval between each response and the next CV was fixed as 1 s.

The frequency of “t” responses for the nine E-/t/ stimuli, and of “d” responses to the nine S-/d/ and nine E-/d/ stimuli was tabulated for each of the four groups. These data were not analyzed in detail, however, because the focus of the study was on the short-lag Spanish /t/ tokens. For the S-/t/ stimuli, the percentage of times each listener gave “t” responses in the two conditions was tabulated. The percent “t” scores were then submitted to a (4) Group × (2) Context (E-/d/ vs. E-/t/) × (9) Token ANOVA, with repeated measures on the last two factors.

2.2. Results

Figure 1 shows the frequency with which listeners in the four groups identified the nine long-lag English /t/ tokens as voiceless (i.e., as “t”), and the frequency with which the listeners identified the nine short-lag tokens of English /d/ and the nine prevocalized tokens of Spanish /d/ as voiceless.

![Figure 1. The frequency with which the subjects in four groups identified as voiceless the nine long-lag tokens of English /t/, the nine short-lag tokens of English /d/, and the nine prevocalized tokens of Spanish /d/. The error bars bracket ± one standard error.](image-url)
The listeners in all four groups labeled the English /t/ tokens as “t” consistently, even the Spanish monolinguals. This confirmed our assumption that, from the standpoint of Spanish, the English /t/ tokens would be identified in terms of the Spanish /t/ phoneme (even though it may have sounded “distorted” or “accented”).

Not surprisingly, most listeners in all four groups identified the prevocal Spanish /d/ tokens as “d”. The rate for the Early L2 Learner group was somewhat lower (91%) than for the other groups because two listeners labeled the prevocal /d/ tokens as “t” in 42% and 33% of instances, respectively. They may have been confused about the nature of the task, which called for phonemic-level identification. Their anomalous responses were evident in just one condition.

Figure 1 also shows the frequency with which the four groups identified the nine short-lag English /d/ tokens as “t”. It came as no surprise that the native English listeners consistently judged the E-/d/ stimuli to be voiced. They gave “t” responses in just 3% of instances. The rate of the Early L2 Learners was somewhat higher (11%) because one of them labeled the short-lag English /d/ tokens as “t” in 90% of instances. The rate for the Late L2 Learners was substantially higher (22%) because all ten listeners in that group identified the English /d/ stimuli as “t” in at least 10% of instances, suggesting that their phonetic criterion for voicelessness may have differed from the native English speakers’. The Spanish monolinguals gave the largest proportion of “t” responses (35%).

Figure 2 shows the frequency with which the four groups judged the Spanish short-lag /t/ tokens to be voiceless (i.e., “t”) in the English /d/ and in the English /t/ contexts. The frequency of “t” judgments was higher on the average in the /d/
than in the /t/ context (76% vs. 54%) for all four groups, which resulted in a significant main effect of Phonetic Context \([F(1, 36) = 50.6, p < 0.01]\). When the listeners heard long-lag stops, it seemed to have made the Spanish /t/ sound less “voiceless”. The lack of a significant Group \(\times\) Phonetic Context interaction \([F(3, 36) = 1.14]\) indicates that the Phonetic Context effect was much the same for all four groups (although there was a significant three-way interaction—see below).

We expected the Spanish monolinguals to identify the short-lag Spanish /t/ tokens as voiceless far more often than the English monolinguals. Indeed, we expected them to label Spanish /t/ tokens as “t” without exception. However, the Spanish monolinguals labeled the Spanish /t/ tokens as “t” in only 81% of instances. The main effect of Group was nevertheless significant \([F(3, 36) = 5.61, p < 0.01]\). Newman–Keuls’ post hoc tests revealed that the Spanish monolinguals gave significantly more “t” responses than the Early L2 Learners and the English monolinguals (56%, 52%; \(p < 0.01\)). The Late L2 Learners’ rate (70%) did not differ significantly from that of any other group.

A significant main effect of Token was obtained \([F(8, 288) = 41.4, p < 0.01]\) because some of the Spanish /t/ tokens were labeled “t” more often than others. The existence of a significant three-way Group \(\times\) Phonetic Context \(\times\) Token interaction \([F(24, 288) = 1.63, p < 0.05]\) suggested that listeners in the various groups may have responded differently to particular tokens. To test this, separate Phonetic Context \(\times\) Token ANOVAs were carried out for each group. The two-way interaction was significant for every group except the Early L2 Learners \((p < 0.05)\). The simple main effect of Phonetic Context was tested for all 36 Group \(\times\) Token combinations with the per-comparison alpha level set at 0.0028 to give an overall experiment error rate of 0.10 (Bonferroni adjustment). These tests showed that only three of the nine Spanish short-lag /t/ tokens were affected significantly by the Phonetic Context manipulation.

The three-way interaction apparently reached significance mainly because of the responses obtained for just one of the Spanish /t/ tokens. Token 4 was judged to be voiceless significantly more often in the context of English /d/ than in the context of English /t/ by the Spanish monolinguals, the English monolinguals, and the Late L2 Learners, but not by the Early L2 learners. It is worth noting that, of the nine Spanish /t/ tokens examined, Token 4 was among the most ambiguous. It was labeled “t” in 55.3% of the instances, on average. Two other Spanish /t/ tokens were labeled “t” significantly more often in the English /d/ than English /t/ context by one of the listener groups, one (Token 6) by the English monolinguals and the other (Token 1) by the Early L2 Learners. Possible acoustic differences between the Spanish /t/ tokens that were ambiguous and Spanish /t/ tokens that were usually judged to be voiceless will be examined below (Section 4).

2.3. Discussion

The listeners in all four groups labeled the short-lag Spanish /t/ tokens “t” (i.e., judged them to be voiceless) more often when these stops were heard along with short-lag English /d/ tokens than when the same CVs were presented with long-lag English /t/s. The two groups of monolingual listeners could not be expected to have had two different phonetic criteria for voiceless stops. Yet, they showed much the same effect as the bilingual listeners, who could be reasonably expected to have had
two criteria. This Phonetic Context effect might reasonably be regarded as a range effect.

The results of Experiment 1 therefore raise doubts as to whether the effects reported by Elman et al. (1977), Flege & Eefting (1987a) and Hazan & Boulakia (1993) were genuine language set effects. By “genuine” we mean that the effects were solely due to the processing of stimuli according to the phonetic criteria of one language in one set, and according to different phonetic criteria of the other language in the other set. As noted earlier, naturally produced short-lag stops from Spanish, Dutch or French in the studies just cited were included in the speech materials used in those experiments. This was done to help induce differing language sets. This procedure may have made the short-lag stops sound relatively more voiceless to the listeners in much the same way that the co-occurrence of short-lag English /d/s in the present experiment made the short-lag Spanish /t/s sound more voiceless.

We expected that the English monolingual listeners in the present experiment would sometimes label the short-lag Spanish /t/ tokens as “d” because English /d/ is often produced with short-lag VOT values. This was indeed often the case. It came as a surprise, however, that none of the three native Spanish groups consistently labeled the Spanish /t/ tokens as “t” even though all of the Spanish /t/ tokens had been produced by monolingual native speakers of Spanish with VOT values appropriate for Spanish /t/.

The bilingual listeners may have sometimes judged the Spanish /t/ tokens as “d” as the result of familiarity with the long-lag realizations of /t/ in English. Previous research has shown that learning an L2 may influence the production of sounds in the L1 (Schouten, 1977; Garnes, 1978; Williams, 1979; Flege & Hillenbrand, 1984; Flege & Eefting, 1987a,b; Yeni-Komshian & Bhathal, 1987). Perhaps the bilinguals’ phonetic representation for (Spanish) /t/ changed as a consequence of their experience with long-lag English /t/, as predicted by Flege’s Speech Learning Model (e.g., Flege, 1992). This might be called the “phonetic category modification hypothesis”.

The Spanish monolingual listeners also gave a sizeable percentage of “d” responses to some of the Spanish /t/ tokens. The L1 phonetic category modification hypothesis might appear to be less plausible for them because they had been in the U.S.A. for only four months on average. However, Williams (1979) found that native Spanish children who had been exposed to English for six months or less showed an effect of L2 exposure on their perception of stops. Further research is needed to determine how much exposure to English stops native speakers of Spanish must receive in order for their perception of /p t k/ to change.

3. Experiment 2

To help determine if stops might be identified differently as the result of listeners’ use of two different phonetic categories, a language set experiment was carried out. The experiment presented here differed from the experiments of Elman et al. (1977), Flege & Eefting (1987a) and Hazan & Boulakia (1993) in two important respects. It included a monolingual control group, and it included two groups of bilinguals differing in age of first contact with English (Early vs. Late L2 Learners). Inclusion of two bilingual groups was motivated by the hypothesis (e.g.,
Perceptual switching in bilinguals

Flege, 1988) that Spanish speakers who learn English in early childhood may establish separate categories for Spanish and English /t/. If so, one might expect that a language set effect—if one were obtained—would be greater for the Early L2 Learners than for the Late L2 Learners. The comparison was also motivated by Diehl's observation (pers. comm., 1988) that listeners in the Elman et al. (1977) study who showed the largest language set effect may have been Early L2 Learners.

3.1. Method

The stimuli used in this experiment were the nine short-lag Spanish /t/ tokens presented in Experiment 1. Three listener groups from Experiment 1 participated: Early L2 Learners, Late L2 Learners and English monolinguals. The listeners identified the initial consonant in each CV stimulus as “t” or “d” by pushing one of two buttons on a response box. The S-/t/ stimuli were randomly presented 10 times each in Spanish and English sets, the order of which was counterbalanced across listeners.

The stimuli were followed by the carrier phrase “_ es la palabra” in the Spanish set, and by “_ is the word” in the English set. This was done to induce differing language sets. (The software used to collect identification responses made it impossible for the listeners to respond before having heard the entire carrier phrase that followed each CV.) To further help induce differing language sets, we had the listeners give answers to questions posed in Spanish or English, as appropriate, during the presentation of stimuli in the two language sets. Twelve English sentences and their Spanish translation equivalents (see Appendix 1) were interspersed at regular intervals among the CV stimuli in the Spanish and English sets, respectively. The listeners had to write down their answers to the aurally presented questions in the appropriate language before pressing a button marked “Q” for Question. The monolingual English listeners did not, of course, understand the Spanish questions. They were told to put a check on the answer sheet after hearing each Spanish question rather than writing down an answer. They were required to respond meaningfully to the English questions, however. An examination of the responses after the experiment indicated that all listeners gave sensible responses to the questions. From this we can infer that the bilinguals had entered into different processing modes in the two language sets.

As in Experiment 1, the dependent variable was the percentage of times the nine S-/t/ stimuli were labeled “t”. Separate scores were calculated for the English and the Spanish sets. The 540 percent “t” scores (3 groups × 10 listeners × 2 language sets × 9 tokens) were submitted to a (3) Group × (2) Language Sets × (9) Token ANOVA with repeated measures on the last two factors.

3.2. Results

Averaged over the three groups and nine tokens, the short-lag Spanish /t/ tokens were labeled “d” in about 35% of instances. This does not necessarily mean that the listeners actually heard a /d/ in 35% of the instances. Listeners tend to make use of however many response categories are made available to them. They may have felt compelled to push the “d” button at least some of the time.

The most important question, therefore, was whether the relative frequency of “t”
Figure 3. The average frequency with which subjects in three groups identified nine short-lag Spanish /t/ tokens as voiceless in the English and Spanish "sets" of Experiment 2. The error bars bracket ± one standard error.

Figure 4. The mean difference in the frequency of "t" responses in the English and Spanish "sets" for the 10 subjects in each of three groups in Experiment 2. A value of "0" indicates no difference between the two language sets.
responses varied as a function of language set. Figure 3 shows the average frequency with which the Spanish /t/ tokens were labeled “t” in the two language sets by the three listener groups. More “t” responses were obtained in the Spanish set than in the English set (68% vs. 63%), resulting in a significant Language Set effect \( [F(1, 27) = 8.20, p < 0.01] \). Much the same effect was evident for all three groups, so the Group \( \times \) Language Set interaction was non-significant \( (p > 0.05) \).

Figure 4 shows the mean difference in the frequency of /t/ responses in the English and Spanish sets for individual listeners. A few Early L2 Learners showed larger language set effects than the Late L2 Learners. The difference was not so large, however, that it would lead one to conclude that these listeners, unlike the majority of the bilinguals, had made use of different phonetic categories when judging short-lag stops in the two language sets. It is worth noting that a few listeners in each group showed no language set effect, and a few showed small changes in a direction opposite from that expected.

A significant Group \( \times \) Token interaction \( [F(16, 216) = 2.89, p < 0.01] \) was obtained because the frequency of “t” responses for the various tokens varied across the listener groups. This interaction was explored by examining the Group effect separately for all nine S-/t/ tokens. The Bonferroni adjustment was used to obtain an overall experiment error level of 0.10. (The error level for each individual test was set to 0.0037, a number which we arrived at by dividing 0.10 by 27, the number of pair-wise comparisons.) The Group effect was significant in two instances. Newman-Keuls’ tests revealed that English monolinguals gave significantly fewer “t” judgments for Token 2 than the Early L2 Learners (28% vs. 61%); and they gave significantly fewer “t” judgments for Token 3 than both the Early and the Late L2 learners (31% vs. 64%, 61%).

### 3.3. Discussion

This experiment differed from previous language set experiments in that it included a monolingual control group. We attempted to manipulate language set by using two different carrier phrases and by having subjects answer questions in Spanish and English. A significant language set effect was obtained, but it does not appear to have been a genuine language set effect because much the same effect was noted for monolingual speakers of English and two groups of Spanish/English bilinguals. As far as we know, the English monolinguals had not been exposed to Spanish or to any other language in which /p t k/ are implemented as short-lag stops. They—unlike the bilinguals—could therefore not be expected to have developed two different phonetic criteria for voicelessness in short-lag stops. Thus one would hardly expect the English monolinguals to have used different phonetic criteria to identify stops in the Spanish and English sets. If so, one might question if the bilinguals, who showed much the same effect, did use different phonetic criteria.

Another reason to question if the bilinguals’ language set effect was genuine was that the effect obtained was small in magnitude. In this experiment it was roughly comparable to the effects reported in previous experiments (Elman et al., 1977; Flege & Eefting, 1987a; Hazan & Boulakia, 1993), but it was actually smaller than the Phonetic Context effect obtained in Experiment 1 (5% vs. 22%). Taken together with the lack of difference between monolingual and bilingual listeners, the results suggest that a post-perceptual change in the criterion used in labeling stops as
voiced or voiceless was responsible for the effect noted here, not the application of two different phonetic criteria for voicelessness.

We originally hypothesized that the Early L2 Learners would show a larger language set effect than the Late L2 Learners. This was based on the assumption that the Early L2 Learners, but not the Late L2 Learners, would have distinct phonetic category representations for the /t/ of Spanish (which specifies short-lag stops) and the /t/ of English (which specifies long-lag stops). Given the absence of a genuine language set effect, this hypothesis could not be evaluated on the basis of data provided here.

4. Post hoc analyses

The nine S-/t/ tokens were labeled four times, twice in Experiment 1 and twice in Experiment 2. As noted earlier, some of these tokens were generally labeled /t/, but some were often identified as /d/. The purpose of the analyses reported in this section was twofold: (1) to determine whether the between-token differences were maintained across the four sessions and across the four listener groups; and (2) if voicing judgments were reliable, to attempt to identify which acoustic properties of the S-/t/ stimuli made them more /t/-like or more /d/-like.

We addressed the question of judgment reliability for the nine S-/t/ stimuli by carrying out a series of correlation analyses. The results showed that the listeners were highly consistent in identifying the nine S-/t/ stimuli. The correlation between percent “t” responses in the two conditions of Experiment 1 ranged from $r = 0.648$ to $r = 0.867$ for the four groups who participated in it. Fairly high inter-group correlations also were obtained for Experiment 1. The correlations between all different pairs of the four groups ranged from $r = 0.580$ to $r = 0.911$ (determined separately for the /d/ condition and for the /t/ condition). The correlations between pairs of the three listener groups who participated in Experiment 2, determined separately for the English and the Spanish set, ranged from $r = 0.803$ to $r = 0.934$. These results suggest that the S-/t/ tokens that were relatively /d/-like for native English monolingual listeners were also relatively /d/-like for the native Spanish listeners. It thus appears that, in judging stops with short-lag VOT, native speakers of both English and Spanish may use similar criteria in making voicing judgments, even when they are confronted with stops from another language.

To compare the results obtained in the two experiments, we computed the overall frequency of “t” responses for the nine S-/t/ tokens in Experiment 1 and in Experiment 2. Figure 5 shows the frequency of voiceless judgments for the nine S-/t/ tokens in the two experiments. The high correlation between the two experiments ($r = 0.946$) suggests that, like the listeners examined by Forrest & Rockman (1988), subjects in the present study were very consistent in how they labeled short-lag /t/ tokens.

As shown in Fig. 5, four of the nine short-lag Spanish /t/ tokens were labeled “t” in 75% to 87% of instances on the average, but five other tokens were labeled “t” less often (45% to 57% of instances). These two subsets of the S-/t/ tokens might be designated the “consistently voiceless” and the “ambiguous” tokens. Acoustic analyses were carried out in an attempt to identify acoustic properties that might distinguish these two subsets of short-lag /t/ tokens. We examined several parameters which have been shown to influence voicing judgments for stops in general, and for
short-lag stops in particular: VOT (see Munro, 1987; but also Forrest & Rockman, 1988); amplitude rise time (Munro, 1987; Darwin & Pearson, 1982); fundamental frequency (F₀) contour (Fujimura, 1971; Haggard, Summerfield & Roberts, 1981; Ohde, 1984); burst intensity (Klatt, 1975; Williams, 1977b; Repp, 1979); burst duration (Klatt, 1975); and duration of the following vowel (see Summerfield, 1981; Miller, Dexter & Pickard, 1984; Miller et al., 1984).

There was little evidence of a relationship between any of the acoustic variables and the frequency of “t” judgments (see also Trent, 1992). The only acoustic variable that seemed to have some predictive value was the duration of the release bursts, which were measured from spectrograms. The correlation coefficient for release burst duration and frequency of /t/ judgments was \( r = 0.493 \), which nearly reached significance. All other correlational analysis yielded clearly non-significant correlation coefficients with \( r < 0.260 \).

5. General discussion

The aim of this study was to provide a better understanding of language set effects. The primary question addressed here was whether Spanish speakers of English might apply two differing criteria for judging short-lag stops as voiced or voiceless as the result of their experience with the long-lag realizations of /t/ in English. Before testing the possible effect of language set, we took the precaution of carrying out an experiment in which phonetic context was manipulated. The most important finding of Experiment 1 was that listeners identified Spanish short-lag /t/ tokens as “t” significantly more often when they were presented with English /d/ tokens than when they were presented with English /t/ tokens (76% vs. 54%). The lack of a significant Group × Context interaction suggested that the context effect was much
the same for Spanish monolinguals, English monolinguals and two groups of Spanish/English bilinguals. It thus appears that the context effect was a kind of range effect due to post-perceptual decision processes, not the result of applying differing phonetic criteria during the process of perception.

Language set was manipulated in Experiment 2. As expected from three previous studies (Elman et al., 1977; Flege & Eefting, 1987a; Hazan & Boulakia, 1993), the subjects identified the short-lag Spanish /t/ tokens as "t" more often in a Spanish set than in an English set (68% vs. 63%). The size of the effect was much smaller than one would have expected had truly "Spanish" and "English" perceptual criteria been applied. Moreover, the Group × Language Set interaction was non-significant, suggesting that the same factor(s) influenced the voicing judgments of English monolinguals and of the two groups of Spanish/English bilinguals who were tested. It thus appears that the voicing judgments of subjects in Experiment 2 were influenced by post-perceptual processes that operated in much the same way regardless of previous linguistic experience. This calls into question the belief that the effects reported by Elman et al. (1977), Flege & Eefting (1987a) and Hazan & Boulakia (1993) were genuine language set effects.

The language set manipulation may have invoked conceptually determined, top-down processes (see Lindblom, 1980; 1986). Even though the English listeners examined here could not speak Spanish, they may have been familiar with a Spanish accent from movies and TV. If so, they may have had tacit awareness that Spanish /t/ is produced with shorter VOT values than the /t/ of English (Flege & Hammond, 1982), and may have applied this knowledge during the experiment (see also Flege & Munro, 1992).

A number of Spanish questions were interspersed among the test stimuli presented in the Spanish set. They contained 10 tokens of /d/, two tokens of /p/, 15 tokens of /k/ and 16 tokens of /t/. The Spanish /p t k/ tokens were realized as short-lag stops. We did not expect the English monolinguals to process the Spanish sentences. (They were told to simply put a check on the answer sheet after each question was output, not to answer the questions.) However, it remains possible that the English monolinguals noted the presence of Spanish stops in the Spanish questions, or the short-lag /p/ in the Spanish carrier phrase (in palabra). One might argue that this was sufficient to induce a language set. However, for this to be true, we believe two things would have been necessary. First, the English monolinguals would have had to very rapidly derive a formulation of how the voiceless Spanish stops differed from their English counterparts. They would then have needed to apply this formulation to their judgements. If this actually happened, we believe the effect would more realistically be viewed as an instance of post-perceptual biasing of responses rather than a genuine change in the processes involved in the actual identification of stops.

The changes in voicing judgments that resulted from the Phonetic Context manipulation in Experiment 1 and the Language Set manipulation in Experiment 2 seemed to operate independently of previous linguistic experience. However, the present study provided direct and compelling evidence for language-specific differences in speech perception. The Spanish monolinguals were more likely to identify short-lag Spanish /t/ tokens as "t" than the English monolinguals. Conversely, the English monolinguals were more likely to identify short-lag English /d/ tokens as "d" than the Spanish monolinguals.
A consideration of previous research suggests that post-perceptual factors may interact with language-specific perceptual factors in the identification of stop consonants. Keating (1979; Keating et al., 1981) varied the ranges of stimuli from a VOT continuum that were presented to listeners. She interpreted the larger range effects obtained for native Polish than English subjects in psychoacoustic terms. She suggested that the English voiced/voiceless phoneme boundary was less easily influenced by range effects than the Polish phoneme boundary because it is based on a contrast between stops with short- vs. long-lag VOT values, rather than on the less salient contrast between stops with lead vs. short-lag VOT values. Foreit (1977) examined the perception of stops by native speakers of English and Thai, a language which has a three-way voicing distinction between stops (lead vs. short-lag vs. long-lag VOT). Presentation of an adapting stimulus with lead VOT had a significant effect on the voiced/voiceless phoneme boundary for the native English speakers but not for the Thai subjects. This led Foreit (1977, p. 349) to conclude that the effects of acoustic manipulations are constrained by “linguistically determined phoneme boundaries”.

We saw much the same effect of the experimental manipulations in Experiment 1 and Experiment 2 for native speakers of both Spanish and English. It would thus appear that Foreit’s observation does not apply to all experimental manipulations. Perhaps differences between listener groups would have been obtained in the present study had we used synthetic stimuli like those employed by Foreit (1977) and Keating (1979). As noted in the Introduction, differences between synthetic short-lag vs. long-lag stops may be more salient to listeners than ones between synthetic stops with lead and short-lag VOT values. This is because the former contain a wide range of acoustic cues (see also Williams, 1977a; 1979). Interestingly, Elman et al. (1977) observed that the language set effect they obtained with naturally spoken short-lag stops could not be replicated with synthetic stimuli.

In Experiment 1, the Spanish monolinguals identified short-lag Spanish /t/ tokens as “t” significantly more often than did English monolinguals (81% vs. 52%). Even though the native English listeners were not consistent, it is noteworthy that they identified the short-lag Spanish /t/ tokens as voiceless (i.e., as “t”) in slightly more than half of instances even though these stops had VOT values similar to the VOT values in the short-lag English /d/ tokens that were examined (Spanish /t/: 21 ms; English /d/: 17 ms). An analogous finding was obtained for the Spanish monolinguals. They identified the short-lag English /d/ tokens as voiced in 65% of instances, even though these stops had VOT values appropriate for /t/ in Spanish. These results clearly indicate that VOT was not an overriding cue to the voicing feature in short-lag stops. A similar result was obtained by Forrest & Rockman (1988), who found only a weak correlation between VOT and voicing judgments for short-lag stops spoken by phonologically disordered children (see also Eilers & Oller, 1976; Monsen, 1976).

This study showed that the frequency with which short-lag Spanish /t/ tokens were identified as voiceless varied considerably. VOT did not seem to be related to the frequency of voiceless judgments for the Spanish /t/ tokens. The listeners were nevertheless quite consistent in terms of how frequently they labeled the various stimuli as “t”. Although VOT was shown not to be a dominant cue for the short-lag Spanish /t/ tokens, some acoustic attributes of the stimuli must have accounted for the consistency of the listeners’ judgments. A wide range of acoustic parameters
were measured in the short-lag Spanish /t/ tokens (VOT, amplitude rise time, rate of $F_0$ decrease, burst intensity and duration), but we were unable to identify an acoustic dimension that could distinguish tokens that were consistently identified as "t" from tokens that were ambiguous between "t" and "d". Similar negative results were obtained by Trent (1992), who tried to determine which acoustic parameters of short-lag stops produced by Spanish/English bilinguals were related to listeners' perception of some of these stops as voiced or voiceless.

6. Summary

The frequency with which Spanish/English bilinguals and Spanish and English monolinguals identified short-lag stops as "t" was affected significantly by a Phonetic Context manipulation in Experiment 1. A smaller but still significant Language Set effect was obtained in Experiment 2. The fact that monolingual and bilingual speakers showed much the same effects for both experimental manipulations suggests that post-perceptual, language-independent decision strategies biased the results of their language-dependent perceptual processing of the stimuli. That is, the results of the two experiments reported here undermine the belief that language set effects arise from the use of two phonetic criteria in identifying short-lag stops found in two languages. However, the possibility remains that certain bilinguals do use two phonetic criteria when processing stops at a phonetic-category level. Phonetic level processes might not be evident when responses are elicited at a phonemic-category level, as in the present study (see Flege, 1992). Additional research is needed to test this, and to provide insight into how the multiple acoustic cues listeners use in making stop voicing judgments are interrelated in perception.

This study was supported by NIDCD grant 20963. Portions of this paper were presented at the 115th meeting of the Acoustical Society of America in Seattle, Washington (May, 1988) and at the Annual Meeting of the German Linguistic Society in Saarbrücken, Saarland (February, 1990). The authors thank Laurie Skelton for help in gathering data, and Mary Beckman, Charles Darwin, Molly Mack and an anonymous reviewer for comments on previous drafts of the article.

References


Perceptual switching in bilinguals


Williams, L. (1979) The modification of speech perception and production in second-language learning, Perception and Psychophysics, 26, 95–104.


Appendix 1

Questions interspersed among the CV stimuli in Experiment 2.

**English set**

1. What is your first name?
2. Where were you born?
3. What is the capital of the U.S.?
4. How many sisters do you have?
5. What is your favorite color?
6. What is the month following April?
7. How many wheels does a car have?
8. How many fingers do you have?
9. What day follows Sunday?
10. How many ears do you have?
11. What color is your hair?
12. In what month were you born?

**Spanish set**

1. ¿Cómo te llamas?
2. ¿Dónde naciste?
3. ¿Cuál es la capital de Estados Unidos?
4. ¿ Cuántas hermanas tienes?
5. ¿ Cuál es tu color favorito?
6. ¿ Qué mes sigue a abril?
7. ¿ Cuántas ruedas tiene un carro?
8. ¿ Cuántos dedos tienes?
9. ¿ Qué día sigue al domingo?
10. ¿ Cuántas orejas tienes?
11. ¿ De qué color tienes el pelo?
12. ¿ En qué mes naciste?