THE ACQUISITION OF VOICELESS STOPS  
IN THE INTERLANGUAGE OF SECOND LANGUAGE  
LEARNERS OF ENGLISH AND SPANISH  

Marie L. Fellbaum  

Australia National University  
Department of Linguistics  
Canberra, ACT, Australia  

ABSTRACT  
This paper presents the preliminary results from work in progress of a paired study of the acquisition of voiceless stops by Spanish speakers learning English and American English speakers learning Spanish. For this study the hypothesis was that the American speakers would have no difficulty suppressing the aspiration in Spanish unaspirated stops; the Spanish speakers would have difficulty acquiring the aspiration necessary for English voiceless stops, according to Eckman's Markedness Differential Hypothesis. The null hypothesis was proved. The results also reveal that a simple report of means will not distinguish the speakers and the respective language learning situation; measurements must also include the range of acceptability of VOT for phonetic segments.  

1. INTRODUCTION  
The Markedness Differential Hypothesis (MDH) was proposed by Eckman in 1977 as an explanation for areas of difficulty in learning a second language. Using notions from typological markedness and implicational relations developed by Greenberg, Eckman's model predicts that aspirated stops will be more difficult to acquire than unaspirated stops for language learners, since aspirated stops are marked with respect to unaspirated stops typologically, i.e. the presence of aspirated stops implies the presence of unaspirated stops in a language, but not vice versa. In this study, the Spanish speakers attained acceptable measures of aspiration for English, whereas American speakers were outside the range of acceptability for native-like production of unaspirated Spanish stops. Measurements of VOT in this report are compared to Lisker and Abramson's (L&A) 1964 crosslinguistic study of stops. The L&A study analyzed eleven languages and included both means and ranges in their report, which serves as a convenient comparison for this study.  

2. THE STUDY  
Subjects. The subjects in this pilot study are eight native speakers of Midwestern American English and seven native speakers of Spanish. The American subjects were in the fifth and final quarter of Spanish at the University of Minnesota and had studied Spanish four years in high school. The Spanish speakers were in the intermediate level of the English as a Second Language program at the university and had lived in the U.S. for a total of six months. The control group for this experiment were the subjects themselves.  

Elicitation Materials. Subjects were given the same set of disyllabic real words in both English and Spanish in carrier phrases, which made the tokens closer to natural speech. This maintained control and comparability of the segments in the same environment, since vowel height and syllabic stress are known to affect aspiration (Keating, 1983). The English carrier phrase was "Say the word ___ again," and the Spanish carrier phrase was "Di la palabra ___ otra vez." The tokens were limited to word-initial voiceless stops, in all three places of articulation (labial, dental-velar, and velar) in a stressed syllable followed by a low back vowel.  

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>_a p</td>
<td>propit</td>
</tr>
<tr>
<td>t toddy</td>
<td>type of drink</td>
</tr>
<tr>
<td>k copper</td>
<td>red metal</td>
</tr>
</tbody>
</table>

Figure 1: Tokens  

Procedure. Tokens were randomized and then arranged in a list with the words appearing three separate times. Subjects were asked to read the tokens in the respective carrier phrase from the second language first, followed by the tokens from their native language. To avoid modelling the tokens for the subjects, each subject was given the list of words and told to ask about any words they didn't know. This gave them the opportunity to try pronouncing the word as they asked for information, and the researcher indicated when they pronounced the word correctly. They were instructed to read at a rate that felt natural and comfortable.  

Measurements. Recordings were made on an Ampex reel-to-reel recorder in a sound-proof booth, transferred to a DAT tape and analyzed on a Sun Sparc2 using Enuoric Waves+ software, version 5.0. Although several measurements were made for each token, this report only analyzes length of VOT. VOT was measured from the burst to onset of voicing on the spectrum. If the burst was not completely clear in the English tokens, the presence of aspiration was taken as evidence of the burst after closure.  

3. RESULTS  
Results of three measurements are reported: 3.1- The mean of the three tokens per subject are graphed and differences in mean VOT length per language group for each consonant are analyzed for statistical significance; 3.2-The individual tokens by native speakers are plotted for each consonant; 3.3-The ranges for each consonant by native speakers are determined, the group means
plotted within the ranges, and the results compared with Lisker and Abramson's study.

3.1 t-Test of Means - Graphs 1 - 6

For the first statistical analysis, alpha was set at .01. The method of analysis was the "correlated samples" t-Test, since each speaker was compared first with their own native language pronunciation of the token and then with the respective second language token. The results for both sets of speakers are displayed in Graphs 1-6. The horizontal axis gives the subject and the vertical axis gives the mean number of msec. of VOT. The mean VOT per speaker of the English tokens are in the top graph and the mean for the Spanish tokens are on the bottom graph. The dark squares represent the second language means.

**Spanish speakers.** Graph 1 shows the mean VOT for word-initial /p/ for each Spanish speaker's pronunciation of the English word 'posi' and the Spanish word 'pago'. It is to be observed that only one subject has achieved the mean VOT of native speakers in this study, 61 msec. (cf. L&A, 58 msec.)

![Graph 1: Mean VOT for word initial /p/ in English and Spanish by Spanish Subjects](image1)

The overall group mean by Spanish speakers for the English 'posi' is 36 msec., and for the Spanish 'pago' is 9 msec., a difference of 27 msec. A Paired Two Sample t-Test shows that the difference between the English and Spanish pronunciation by Spanish speakers is significant: t=3.365, p<.01.

Graph 2 shows the mean VOT for word-initial /θ/ by Spanish speakers' pronunciation of the English word 'taddy' compared with their mean VOT for the Spanish word 'taza'. It is to be observed that no speaker has achieved the mean VOT of native speakers in this study, 75 msec. (cf. L&A, 70 msec.) The overall group mean for /θ/ in English 'taddy' is 44 msec., and for the Spanish 'taza' is 8.9 msec., a difference of 35.1 msec. The t-test result, t=4.9, again shows that the difference between the English and Spanish pronunciation is significant: p<.01.

Graph 3 gives the mean VOT for word-initial /k/ by Spanish speakers for the English word 'koper' and the Spanish word 'cada'. Two speakers achieved the mean group mean by native speakers, 71 msec. (cf. L&A 80 msec.)

![Graph 3: Mean VOT for word initial /k/ in English and Spanish by Spanish Subjects](image2)

The overall group mean for VOT of /k/ in English 'koper' is 63 msec., and for the Spanish 'cada' is 24 msec. The 39 msec. difference in group means is the largest for the three consonants; and the value of t=9.175, shows that the difference between the English and Spanish pronunciation is significant at the .001 level. Thus, all three voiceless stops change in mean length of VOT by Spanish speakers is statistically significant.

**American speakers.** Graph 4 gives the mean VOT for word-initial /p/ for the English word 'posi' compared with the Spanish word 'pago'. The group mean for Spanish native speakers, 9 msec. (cf. L&A, 4 msec.) has not been achieved by any subject.

![Graph 4: Mean VOT for word initial /p/ in English and Spanish by American Subjects](image3)

For the American subjects, the overall group mean for /p/ in English 'posi' is 61 msec., and for Spanish 'pago' is 42 msec. The difference in means of 19 msec., using the Paired Two-Sample t-Test, is significant: t=4.54, p<.01.

Graph 5 gives the mean VOT for word-initial /θ/ by American speakers for the English word 'taddy' and the Spanish word 'taza'. Again no American subject has achieved the mean of 8.9 msec. (cf. L&A, 9 msec.)
The overall group mean for the Americans for 'toddle' is 75 msecs.
and Spanish 'lazá' is 37 msecs., a difference of 38 msecs. The t-
Test results show the difference between the VOT of English and
Spanish /l/ is significant with $t=6.354$, $p<0.001$.
Graph 6 gives the mean VOT for word-initial /l/ by American
speakers for the English word 'copper' and the Spanish word
'cada'. Again no American subject has achieved the group mean
for native speakers, 24 msecs. (cf. L&A, 29 msecs.)

The overall group mean for English 'copper' is 71 msecs. and
Spanish 'cada' is 51 msecs. The t-Test shows the difference in
mean of 20 msecs. between the English and Spanish pronunciation of /l/ by Americans as a group is also significant at
$p<0.01$, $t=3.586$.
The results show, as with the Spanish speakers, the changes in
VOT for all three voiceless stops are statistically significant for
the American subjects on all three consonants.

3.2 Ranges of VOT - Graphs 7-9
By calculating the range of VOT in the native speaker control
group of this experiment, the ranges and means for the tokens can
be established for each language. Graphs 7 - 9 show the ranges
and individual variation in patterns for VOT on the three tokens
by native speakers in their respective native language, Spanish or
English. The lower portion of each graph shows the VOT
measurements on all three tokens by the Spanish speakers, and
the upper portion of each graph shows the individual pronunciation for native speakers of English for the respective
stops tested in this experiment.

Graph 7 shows the ranges for VOT for native speakers for /p/ in
the word 'page' for Spanish speakers and 'posit' for the American
speakers.

For the Spanish subjects the range is 0 - 17.7 msecs., SD = 4.1
For the American subjects, the VOT ranged from 25.9 msecs. to
98.5 msecs., SD =16.
Graph 8 shows the ranges and individual variation for VOT for
native speakers for /l/ in the word 'lazá' for Spanish speakers and
'toddle' for the English speakers.

For the Spanish subjects the range is 0 - 16.4 msecs., SD=3.7.
For the American subjects, the VOT ranged from 46.6 msecs. to
120 msecs., SD=16.
Graph 9 shows the ranges and individual variation for VOT for
native speakers for tokens of /k/ in the word 'cada' for Spanish
speakers and 'copper' for the English speakers.
For the Spanish subjects the range begins at 7 msecs., not 0 as in the two previous consonants, and has a maximum of 38 msecs., SD=4. For the American subjects, the VOT ranged from 36.9 msecs. to 99 msecs., SD=17.

3. 3 Means and Ranges of VOT – Chart 1
The ranges plotted in Graphs 7-9 are displayed in Chart 1 below, and compared with the ranges from the L&A study. The bars on the left side represent the Spanish ranges, and the bars on the right are the English ranges. The mean VOT of each language group and their location within these ranges are noted with S for Spanish and A for American second language learners. Results indicate that the Spanish speakers, as a group, were able to reach the accepted native language VOT ranges of English found in both the L&A study and this study. In comparison, English speakers, as a group, were outside the accepted VOT range for Spanish /p/ and /n/ in both studies. The mean VOT for /k/ was outside the range for Spanish /k/ by native speakers in this controlled study, but 1 msecs. inside the range in the L&A study. After plotting the group mean VOT within the acceptable ranges displayed in the chart, the two groups of language learners are clearly differentiated.

<table>
<thead>
<tr>
<th>Chart 1: VOT Ranges per consonant by native speakers and Mean VOT of second language learners</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>/n/</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>/k/</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>mean: 10 20 30 40 50 60 70 80 90 100 110 120 130 140</td>
</tr>
</tbody>
</table>

| VOT Ranges (Fellbaum, 1996) | VOT Ranges (Lisker & Abramson, 1964) |

A = American VOT Means  S = Spanish VOT Means

4. CONCLUSION
The statistically significant changes of p<0.01 or p<0.001 for both groups of learners on all three steps did not distinguish the two groups of learners. Until ranges were incorporated in the analysis of the results, the use of group means, as currently reported in second language learning studies were not useful in understanding the acquisition of aspirated and unaspirated stops. After acceptable ranges of VOT were analyzed for each language, the null hypothesis was proved and the need for a more complete description other than markedness relations as predicted in the MDE to understand language was demonstrated. Suggestions for further study include the refinement of the carrier sentence used for controlled speech sampling; collection and completion of measurements of more tokens and more subjects in a similar controlled environment, necessary for languages which exhibit as much variability in VOT as is found in English; elicitation of free speech be compared with controlled speech as studied in this experiment, devising of a more careful method for controlling the speed of each speaker's pronunciation. This study points to the need for more careful measures of the ranges of other languages; languages differ in the acceptance of speaker variation, the extent of ranges, and the starting points of each range; e.g., Spanish starts at 0 msecs. for /t/ but Chinese /t/ starts at 5 msecs. in the L&A study. Researchers in multi-language recognition systems which seek to develop evaluate and compare multi-language recognition algorithms, need to incorporate the ranges of a language's aspiration and speaker variation, and not simply calculate the mean duration of individual stop consonants. This is especially important as spoken language corpora expand to the public domain and are used by researchers in various disciplines.

ACKNOWLEDGEMENTS
The author is grateful to Dr. Julie Vowwiher and Chris Clerigh of the Speech Group Engineering Lab, Brian Taylor, James Ng, and Paul Cimenti of the Language Center, and G. Oliphant of the Psychology Dept. at University, and Dr. Harold Koch of the Dept. of Linguistics, and Dr. B. Turf of the Dept. of Applied Mathematics at ANU.

REFERENCES