VOICE-ONSET-TIME IN THE PERCEPTION OF FOREIGN ACCENT BY NATIVE LISTENERS OF SPANISH

By Manuela Gonzalez-Bueno

University of Kansas Lawrence U.S.A.

This experiment seeks to determine the extent to which the variation of temporal characteristics of stops in a particular Spanish utterance spoken by an English speaker had a noticeable effect on the perception of foreigness of speech. The experiment was conducted by manipulating an utterance with special computer programs and using native subjects to rate the manipulated utterances as being more foreign or more native. The results were positive; the variation of the temporal characteristics of stops did have a noticeable effect on the perception of foreigness of speech by native listeners. Therefore, instruction directed to shorten such temporal characteristics of the English accented stops toward more Spanish-like values was recommended. It was suggested that this might be achieved through the use of interactive computer techniques.

INTRODUCTION

Joan Morley (1987) states that "intelligible pronunciation is an essential component of communicative competence" (iv). From this point of view, it seems quite easy to learn to pronounce a foreign language well enough to be intelligible. However, sometimes the goal of achieving a quasi-native pronunciation is very desirable for those language students who want to be successful in their learning of a second language. This may not only be a personal goal addressed to fulfill a personal resolution, but may also be addressed to gain acceptance and recognition by the native speakers and listeners of the second language in question.

The author felt that it would be of interest to measure the recognition of one cue of accentedness by native listeners. In the production of Spanish stops by English-speaking learners of Spanish as a second language, long VOT and aspiration have been established as cues of accentedness, since Spanish stops have short VOT and are not aspirated (Lisker & Abramson, 1964). A perception experiment was conducted with naive native monolingual Spanish listeners, presented with English accented productions of a Spanish word, "casa." The word contained a syllable-initial velar stop, [k], with a voice-onset-time modified in 5 ms-steps from native Spanish to native English values. The spectral characteristics of stops include a silent period that corresponds with the vocal tract closure; a transient that corresponds to the burst or the release of the oral constriction; and the voice onset time (VOT) defined as the time interval between the release of the oral constriction and the start of the glottal pulsing (Lisker and Abramson, 1964). Most languages limit VOT to two or three relatively narrow and non-overlapping ranges. For prevocalic stops the range into which a given VOT falls serves the listeners as an important cue to the voicing category of the phone (Lisker & Abramson, 1964). By convention, a VOT of 0 ms signals simultaneous onset of release-burst noise and glottal pulsing. Negative VOT values denote the time before burst noise voice begins (prevoicing or voicing lead), while positive VOT

values show the time of vocal striations appear after burst onset (voice lag). The difference between English and Spanish stops consists of a long lag for the English stops, a short lag and voicing lead for Spanish stops.

The VOTs in Spanish stops are known to be shorter than those in English stops. Lisker & Abramson (1964) determined the distinct ranges into which VOT fell in different languages. The following approximate mean values are for English and Spanish:

	Voiced	Voiceless
English	25 ms (or less)	50 ms (or more)
Spanish	-100 ms	0 ms

Also, Spanish voiceless stops do not contain aspiration during their VOTs, whereas English stops do. Aspiration is the turbulence of air at the vocal folds, caused by the drop of Psg (subglottal pressure) and the rise of Pio (intra-oral pressure) at the closure phase, and the sudden release of the articulators (Revoile, S., James M. Pickett, Lisa D. Holden-Pitt, David Talkin and Fred D. Brandt, 1972).

The experiment arose on the basis of what Flege (1981) calls the "phonological translation hypothesis." This hypothesis states that "an individual may be completely successful in his/her phonetic learning of a second language and yet retain an accent because pronunciation of the foreign language is based on pairs of corresponding sounds (or nonsegmental phonetic dimensions) found in the native and target language" (p. 451). This is particularly true for stops: Spanish and English stops are phonetically similar sounds; hence they are transcribed with the same IPA symbols. However, as Flege warns, this should not obscure the fact that they are realized slightly differently at the phonetic level. The English speaker of Spanish is often not aware of the subtle difference between English stops and Spanish stops. The difference consists basically of the different VOT duration. This experiment was intended to determine the extent to which the modified duration of VOT of a spoken Spanish initial voiceless stop [k] caused this stop to be perceived by Spanish monolingual listeners as "foreign" or "accented," when produced by a native speaker of English whose proficiency level in Spanish was Intermediate (according to ACTFL Guidelines). The initial stop was contained in the Spanish word "casa" ['kasa] ('house'). The investigator manipulated the VOT by removing proportional segments of aspiration (contained in the VOT) and adding proportional silent segments. Listeners were asked to identify the utterance as "near native" or "near foreign."

The hypothesis of the experiment was that listeners' judgements of accentedness for each stimulus would be correlated with long VOT duration. It would also indicate that the presence of aspiration caused these rOTs to be perceived as more accented.

On the basis of results of this experiment, it was hoped that it would be possible 1) to establish the range in duration of the VOT of velar voiceless stops that would be perceived as native-like by Spanish listeners, and 2) to determine the effect of the presence or absence of aspiration on the perceived nativeness by Spanish-speaking listeners. If such a range in duration of the VOT of [k] could be established, the same could be done with the other places of articulation, bilabial and dental. This information would make it easier to implement the formal teaching of the

pronunciation of Spanish stops. This could be achieved through the design of techniques addressed to shorten the VOT of students' English accented stops. Through periodical measurements of the VOTs produced by the learners, it could be determined how close, in milliseconds, they were to the values perceived as more native by the listeners of this experiment. It would also be possible to determine how much instruction a particular learner would need according to the duration of her or his VOT.

SUBJECTS

The native speaker of American English (female, age 20) whose utterance was used as stimuli had been analyzed in a previous study of pronunciation. After being administered an OPI (Oral Proficiency Interview) in Spanish, she had been rated as having and Intermediate proficiency level in Spanish, according to ACTFL (American Council of the Teachers of Foreign Languages).

The listeners were 18 monolingual native Spanish speakers, nine males and nine females, with ages ranging from 19 to 57 years old. The subjects were all volunteers residing in Seville (Spain). The Seville dialect differs from standard Castilian Spanish, but there is no evidence that stops are produced differently from those in standard Spanish. None of the subjects were very familiar with a foreign accent nor were they familiar with pronunciation issues. They reported no speech or hearing disorders, and all of them had a universitary education in various fields or were college students at the time of the experiment.

STIMULI PREPARATION

The stimulus used was the Spanish word "casa" ['kasa] ('house'), taken from an OPI administered to the American-English speaker. This word was chosen simply for its clarity and perceptibility, and for the absence of other acoustic characteristics that would lead to additional accented cues (e.g., other accented vowel or consonant productions). The original VOT for English accented [k] was measured in a wide-band spectogram and was found to have a value of 75 ms, which is typical for the English voiceless velar stop. Lisker and Abransom (1964) reported an average of 80 ms for English velar stops. This original VOT was modified in two ways via digital editing of the speech waveform. For the manipulation of the VOT waveform, the original tape containing the OPI from which the word-stimulus was extracted was played in a Realistic SCT-84 High Speed Dubbing Cassette Deck. This recorder was attached to an IBM/XT personal computer through a series of other stereo equipment, including a frequency equalizer, an amplifier, and D-A/A-D converters. The computer is equipped with a program designed and developed for the analysis and manipulation of speech waveform. This program will analyze a speech sample by converting the original analog signal into a digital signal read and understood by the computer. The word "casa" ['kasa] was input into the computer. The different options of the waveform analysis program allow one to look at the utterance screen-by-screen to examine the waveform and to edit it at the desired points. Cuts ("splices") were made immediately after the burst of the velar stop [k] and immediately before the transition into the following vowel, and the remaining two segments of the waveform were pasted together. Periods of silence of different durations were inserted between the different sections resulting from the cuts previously made, Twenty-eight stimuli were created in this way.

First, 65 ms of VOT duration were deleted; the remaining 10 ms corresponding to the burst were retained. The burst was left intact so that not all cues for the perception of a stop would be lost, since there is no previous silent period for a syilable-initial stop, which is an important cue for stops at intervocalic position (Lisker, 1957). Then, intervals of silence were added to the remaining 10 ms, in 5 ms steps, up to the original of 75 ms. That is, 14 stimuli of different durations were created in addition to the original stimulus. Stimuli are listed in Table 1.

A second set of stimuli was created by cutting back duration of the original English VOT in 5 ms steps. This way, the original aspiration was kept intact. These stimuli are listed in Table 2.

There were, then, 14 stimuli with aspiration (s A1 to s A14), corresponding to 14 without aspiration (s U1 to U14).

A software randomization routine randomized the 28 stimuli into three different blocks. There were two repetions, with a 4 s interval between repetitions and an 8 s interval between stimuli. The complete stimulus set was recorded onto a Sony 60-min. Type I audiotape.

A response sheet was prepared for listeners to complete, which contained a 7-point scale of "foreign accentedness," as well as a 5-point confidence rating scale (sure to not sure judgment) for each of the 84 stimuli (28 x 3 randomizations). A questionnaire was created, to obtaine some information about the listeners (e.g., age, education). Both the response sheet and questionnaire are given in Appendix 1 and 2, respectively.

The stimulus tape was played to listeners through the headphones of a Tandberg 600 language laboratory (Tandberg Educational System, Norway, 1988). After a short training session to acquaint listeners with the experiment and the stimuli, the actual experiment was run, in the language laboratory of the Philology College at the University of Seville. The listeners had previously answered the questionnaire, and had been trained to fill out the response sheet. The procedure took approximately 30 minutes. At the end, the response sheets were collected.

RESULTS

The 18 answer sheets were coded, and mean value for the three randomizations calculated, for each stimulus and each listener. Then, the mean values and standard deviations for the 18 subjects' responses for each stimulus and confidence ratings were calculated. The results are shown in Table 3.

The graph of coordinates shown in Figure 1 depicts the different degrees of foreigness for each stimulus. The two sets of stimuli, aspirated and unaspirated, are represented on the horizontal double axis, from s 1 (A1 and U1), with the shortest VOT (10 ms), to s 14 (A14 and U14), with the longest VOT (75 ms). The seven point rating scale is represented on the vertical axis, from 1 (most native) to 7 (most foreign).

Paired t-tests were performed for each pair of stimuli to determine whether they were significantly different or not. Stimulus pairs that showed no significant difference tended to show a value less than 4 (the mid point along the stimulus ranking scale) and are shown in Table 4.

DISCUSSION

Interestingly, the stimuli with VOTs shorter than 20 ms (10 and 15 ms) were perceived as less native than some of the longer ones. S A1 and s U1, for instance, with a VOT of 10 ms, were bound to have a "nativeness" mean value of 4.12 and 3.96, respectively, whereas s U7, with a VOT of 40 ms, was assigned a "nativeness" mean value of 3.41. More interestingly, they were perceived as It] (personal communication with some of the subjects after the experiment). According to the literature (Lisker & Abramson, 1964), the different duration of VOT is one of the cues for place of articulation: VOTs of alveolars [t, d] (and similarly, dentals) are shorter than VOTs of velars ([k, g]). It could be that the VOT of the stimulus was shortened so much that it was perceived as the VOT of a [t]. Martinez Celdran (1991) states that even though VOT is used in English to justify the voiced/voicelessness of stops, it does not play the same role in Spanish, and it is very likely that it is used in Spanish as a supplementary cue for place of articulation. If the listeners perceived stimuli s A1 and s U1 (each with a VOT of 10 ms) as "more foreign" than s U7 (with a VOT of 40 ms), and sometimes as [t], it may be due to the fact that the shortened VOT crossed the threshold -perceptual boundery - between [k] and [t]. Citing Castaneda (1986), Martinez Celdran (1991: 127) points out that the VOT for Spanish velars is 25.7 ms, whereas that for bilabials is 6.5 ms. He does not mention dentals, but it is reasonable to assume that the VOT for dentals must fall in the range between 20 and 10 ms, and this is precisely the length of the VOT of those stimuli perceived as [t]. Mfijica, Santos and Herraiz (1990: 112) also point out the correlation in the perceptive confusion between [t] and [k], especially when they occur before vowels [i, e] and before the vowel [a] (as in this experiment with ['kasa]).

Stimuli U4 and U5, with a VOT of 25 ms and 30 ms, respectively, both unaspirated, were perceived as more foreign (or less native) than their aspirated counterparts (stimuli s A4 and s A5). Beyond 35 ms, the longer the VOT of the stimuli, the more foreign they were perceived, although judgments of "foreigness" did not increase in a linear fashion, parallel to increasing VOT duration.

The presence or absence of aspiration did not seem to affect the responses of the listeners. The responses to stimuli with a VOT duration of less than 35 ms for the two series (aspirated and unaspirated), were parallel. For VOTs of more than 35 ms, the responses were ambiguous: Sometimes aspirated stimuli were perceived as more native, while other cases perceived as more foreign. Aspirated stimuli with VOTs greater than 55 ms tended to be perceived as more foreign than their unaspirated counterparts. A paired t-test was performed, to compare aspirated and unaspirated stimuli. The results showed that, except for s9 (i.e., s A9 and s U9), no stimuli showed a significant difference in nativeness rankings. Confidence ratings were not affected greatly by the presence/absence of aspiration. Only those confidence ratings for sl, s3, s7, and s14 showed a significant difference ($p<0.05$). The remainder of the stimuli were not significantly different.

CONCLUSION

From the results of this experiment, shown in Table 4, it can be concluded that the Spanish voiceless velar stop [k] is perceived as more native by the monolingual Spanish listeners when VOT duration ranged from 15-35 ms, and that the presence or absence of aspiration has little

effect. In the future, this experiment should be replicated for stops having the other two places of articulation, dental [t] and bilabial [p], if these results are to be generalized.

Results of this study have implications for teaching pronunciation. Periodic measurements of the VOTs produced by learners of Spanish can determine how close (in milliseconds) they are to the values that are perceived as "more native". Thus, the amount of instruction required by a particular English-speaking learner in order to shorten VOTs can be established. However, a complete reduction to the Spanish values may not be necessary, since according to existing phonetic studies, "the values of phonetic parameters measured in the speech of second language learners are often intermediate to those typical of monolingual speakers of the native and target language" (Pinkerton 1972, Suomi 1976, Flege 1980). An acceptable pronunciation of Spanish stops can be established by calculating the midpoint between the English and the Spanish values, of ap-proximately 50 ms for English voiceless stops, and 14 ms for Spanish stops, according to Lisker and Abramson 1964, i.e., 32 ms. This experiment showed that VOTs within a range of 15 and 35 ms were acceptable to the 18 participating native Spanish listeners, i.e., the mean duration for VOT of Spanish stops and the midpoint between English and Spanish values.

Various techniques that might help implement teaching pronunciation of Spanish stops should involve training in perception as well as in production. Chun (1991) suggests the use of interactive techniques with computers, to which data from the present study could be applied for teaching perception of VOT using modified stimuli. Research involving techniques is too broad an area to be addressed here. However, the following should be emphasized: (1) There is a clear need to deal with pronunciation in the second language classroom; (2) pronunciation should be examined through the use of commonly accepted methodological procedures involving instrumental analysis (e.g., spectrographic and waveform analyses); and (3) second language teachers must be familiar with the phonetic and phonological components of the target language, as well as with techniques designed to develop perception and articulation of this language. After all, the ultimate goal - perhaps unattainable for some - is, nonetheless, to "sound like a native speaker" in all aspects of the language.

Table 1,: Unaspirated Stimuli

s U1	:	VOT of 10 ms (75 ms original - 65 ms)
s U2	:	VOT of 15 ms (10 ms original + 5 ms silence)
s U3	:	VOT of 20 ms (10 ms original + 10 ms silence)
s U4	:	VOT of 25 ms (10 ms original + 15 ms silence)
s U5	:	VOT of 30 ms (10 ms original + 20 ms silence)
s U6	:	VOT of 35 ms (10 ms original + 25 ms silence)
s U7	:	VOT of 40 ms (10 ms original + 30 ms silence)
s U8	:	VOT of 45 ms (10 ms original + 35 ms silence)
s U9	:	VOT of 50 ms (10 ms original + 40 ms silence)
s U10	:	VOT of 55 ms (10 ms original + 45 ms silence)
s U11	:	VOT of 60 ms (10 ms original + 50 ms silence)
s U12	:	VOT of 65 ms (10 ms original + 55 ms silence)
s U13	:	VOT of 70 ms (10 ms original + 60 ms silence)
s U14	:	VOT of 75 ms (10 ms original + 65 ms silence)

s = stimulus

U - unaspirated

ms = milliseconds

Table 2,: Aspirated Stimuli

s A14	:	VOT of 75 ms (75 ms original)
s A13	:	VOT of 70 ms (75 ms original - 5 ms)
s A12	:	VOT of 65 ms (75 ms original - 10 ms)
s A11	:	VOT of 60 ms (75 ms original - 15 ms)
s A10	:	VOT of 55 ms (75 ms original - 20 ms)
s A9	:	VOT of 50 ms (75 ms original - 25 ms)
s A8	:	VOT of 45 ms (75 ms original - 30 ms)
s A7	:	VOT of 40 ms (75 ms original - 35 ms)
s A6	:	VOT of 35 ms (75 ms original - 40 ms)
s A5	:	VOT of 30 ms (75 ms original - 45 ms)
s A4	:	VOT of 25 ms (75 ms original - 50 ms)
s A3	:	VOT of 20 ms (75 ms original - 55 ms)
s A2	:	VOT of 15 ms (75 ms original - 60 ms)
s A1	:	VOT of 10 ms (75 ms original - 65 ms)[*]

* This stimulus coincides with s U1

s = stimulus

U = unaspirated

Ms = milliseconds

Table 3

STIMULUS	Mean	s.d.	Confidence	s.d.
A 14	5,26	1.32	1.83	1.08
U 1	3,96	1.57	1.99	1.08
U 2	3.16	1.63	1.93	0.94
U 3	2.84	1.39	1.90	1.05
U 4	3.18	1.09	2.25	1.18
U 5	3.22	1.04	2.27	0.92
U 6	3.48	1.63	2.21	1.09
U 7	3.41	1.54	2.25	1.05
U 8	4.26	1.74	2.09	0,91
U 9	3.78	1.54	2.31	1.08
U 10	4.23	1.65	2.40	1.20
U 11	3.97	1.57	2.16	1.03
U 12	3.95	1.43	2.02	1.08
U 13	4.50	1.39	2.27	0.86
U 14	4.52	1.28	2.21	1.07
A 13	4.81	1.32	2.23	1.25
A 12	4.79	1.25	2.29	1.13
A 11	4.82	1.53	2.03	1.08
A 10	4.38	1.69	2.18	0.89
A 9	4.98	1.52	1.96	1.13
A 8	3.43	1.42	2.15	0.87
A 7	4.10	1.64	1.95	1.03
A 6	3.33	1.74	2.15	1.07
A 5	2.80	1.22	2.09	1.06
A 4	2.95	1.62	1.94	1.01

A 3	2.97	1.44	2.29	1.09
A 2	3.58	1.91	2.08	1.23
A 1	4.12	1.78	2.56	1.20

sd = standard deviation A = aspirated U = unaspirated

Table 4

Legend for Chart:

A - Stimulus#
B - VOT
C - Value[*]
D - Standard deviation
E - Confidence

A	B	C	D	E
U2	15 ms	3.16	1.63	1.9
U3	20 ms	2.84	1.39	1.9
U4	25 ms	3.18	1.09	2.2
U5	30 ms	3.22	1.04	2.2
U6	35 ms	3.48	1.63	2.2
A6	35 ms	3.33	1.74	2.2
A5	30 ms	2.80	1.22	2.1
A4	25 ms	2.95	1.62	1.9
A3	20 ms	2.97	1.44	2.3
A2	15 ms	3.58	1.78	2.6

* PR > ITI = 0.4486 (>0.05) = Non-significant difference.

U = unaspirated

A = aspirated

ms = milliseconds

GRAPH: Figure 1

REFERENCES

American Council on the Teaching of Foreign Languages. The ACTFL Oral Proficiency Interview. Yonkers, NY: ACTFL, 1989.

Chun, D. M. 1991. The State of the Art in Teaching Pronunciation. Paper presented at the GURT, Washington, DC.

Flege, J. E. 1980. Phonetic Aproximation in Second Language Acquisition. Language Learning, 30: 117-34.

Flege, J. E. 1981. The Phonological Basis of Foreign Accent: A Hypothesis TESOL Quarterly, 15, 443-453.

Lisker, L. 1957. Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English. Language, 33, 42-49.

Lisker, L. and Abramson, A. S. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. Word, 20, 384-422.

Martinez Celdran, E. 1991. Fonetica Experimental: Theoria y Practica. Madrid: Sentesis.

Morley, J. (Ed.) 1987. Current Perspective on Pronunciation: Practices Anchored in Theory. Washington, DC: TESOL.

Mujica, E., Santos, Maria del Mar, and Herraiz, Jose 1990. Duracion de las Transiciones en las Oclusivas Sordas del Castellano. Estudios de Fonetica Experimental. Barcelona: PPU.

Pinkerton, S. 1973. The Learning of English Suprasegmental Rules for Stress and Final Syllables by Spanish Speakers. Paper presented at the Mid-America Linguistics Conference, October, 1973.

Revoile, S., James M. Pickett, Lisa D. Holden-Pitt, David Talkin and Fred D. Brandt 1987. Burst and Transition Cues to Voicing Perception for Spoken Initial Stops by Impaired and Normal-Hearing Listeners. Journal of Speech and Hearing Research, 30, 3: 311.

Suomi, K. 1976. English Voiceless and Voiced Stops as Produced by Finnish and Native speakers. Jyuaskyla Contrastire Studies 2. Jyvaskyla University, Department of English, Finland.