SOME CHARACTERISTICS OF VOT IN PLOSIVES PRODUCED BY SPEAKERS OF ENGLISH AND PORTUGUESE

Paulina D. Artimonte Rocca and Marcello Marcelino
Pontifícia Universidade Católica de São Paulo, Department of Linguistics, São Paulo, Brasil

ABSTRACT

Our present research investigates how speakers with distinct linguistic backgrounds differ in their production of voiceless plosives in both English and Portuguese as far as VOT lag is concerned. A group of 5 speakers were the subjects of the experiment: a Brazilian learner of English at intermediate level, a Brazilian speaker of English with linguistic knowledge, three bilinguals Portuguese/ English, one of them with Portuguese as his first language and the other two with English as their first language. The speech material consisted of 9 Portuguese words and 8 English words in carrier sentences. The recordings were submitted to acoustic analyses using the CSRE (Canadian Speech Research Environment). The VOT of /p, t, k/ were measured and compared. The results of the research are discussed on the factors that seem to lead the distinct speakers to present different VOT patterns in both languages which go from partial phonetic approximation to a complete command of the languages.

1. INTRODUCTION

Any speaker of a given language is endowed with an elaborate time control system which governs the specific duration of speech units as well as the catenation between them. Such control system develops in parallel with the neurological maturation and the language acquisition. It also depends on conditions inherent to the central nervous system as well as on anatomic-physiological characteristics of the speech articulators proper and it is sensitive to the linguistic environment to which the speaker is exposed. The phoneme boundaries will, therefore, vary from language to language along the time dimension.

It is commonly postulated that voiceless contrasts are characterized differently along the Voice Onset Time (VOT) dimension in languages such as English in opposition to Portuguese. In English, voiceless plosives /p, t, k/ are produced with a long-lag VOT followed by aspiration in initial and medial position before a stressed syllable, while in Portuguese, they are realized with a short-lag VOT thus resulting in absence of aspiration. These cross-language oppositions presented the possibility to access how subtle differences in the phonetic realization of a phonological contrast would interfere in the production of the speech sounds of a foreign language by adult language learners thus characterizing various levels of proficiency. As far as perception and production are concerned, Flege [1] gave evidence for the learner's tendency to consider equivalent "new" and "similar" phones in a foreign language, and [2] stated the relevance of phonetic input to L2 phonological learning, whereas Jamieson [3] considered that training to establish non-native speech contrasts is more effective when it focus the student's attention on the acoustic patterns that are relevant for specific groups. Fowler [4], on the other hand, gives indication of phonetic learning based on an inclination to imitate gestures of the language community, suggesting that "to speak is to engage in a kind of activity having linguistic significance that speakers share with members of their language community, and to listen is ... to achieve "parity" in communication". In the light of these findings, the objective of our present research is to investigate how accurately speakers with different linguistic backgrounds can produce the VOT of /p, t, k/ at the beginning of words in both Portuguese and English.

2. METHODOLOGY

2.1 The Corpus

The corpus was built with 9 Portuguese words in the carrier sentence "Eu disse _____ para ele" and 8 English words in the carrier sentence "I said _____ in a modal voice". Control over vocalic environment was exercised. The words selected were medial in the carrier sentences and presented the following characteristics:

- 6 Portuguese words with the voiceless plosives /p, t, k/ in initial stressed position;
- 6 English words with the voiceless plosives /p, t, k/ in initial stressed position;
- 3 Portuguese words with the voiceless plosives /p, t, k/ in medial unstressed position;
- 2 English words with the voiceless plosives /p, k/ in medial unstressed position. The voiceless English plosive /t/ was discarded since it is usually realized as a flap in unstressed position.

The recordings were made at random.

2.2 The Subjects of the Experiment

A group of 5 speakers, all educated speakers of standard varieties of their languages, with different linguistic awareness and competence as far as the foreign languages involved are concerned, were the subjects of the experiment:

- subject **L** - a Brazilian learner of English at intermediate level who does not have contact with an English speaking community.
- subject **P** - a Brazilian speaker of English with linguistic knowledge who has occasional contact with an English speaking community.
- subject **M** - a bilingual Portuguese/ English, educated in English, with Portuguese as his first language, who has always lived in Brazil, and who has daily contact with English and Portuguese speaking communities.
- subject **T** - a bilingual Portuguese/ English with English as his first language, who has been living in Brazil for the last 3 years, and who has daily contact with English and Portuguese speaking communities. The subject has also got

a Master degree in the Spanish Language and lived in Spain for 3 years.

- Subject **S** - a bilingual Portuguese/ English with English as her first language, who has been living in Brazil for the last 7 years, and who has daily contact with English and Portuguese speaking communities.

Henceforth, the subjects will be identified by the letters listed above.

2.3 The Recording

Each subject recorded the corpus individually. They received the sentences written on strips of paper, one strip for each key word. They were instructed to read the sentences as naturally as possible at normal speed. No other attempt to control the rate of the utterance was made. The subjects recorded the sentences with each of the selected words 5 times in a sound proof booth. The order of occurrence of the words varied from one recording to the other.

2.4 The Analysis of the Data

The recordings were submitted to acoustic analyses using the CSRE (Canadian Speech Research Environment) from Avaaz Corporation. Voice onset times for /p, t, k/ were measured by pointing out the interval between the release of the stop, zero-time point of reference, and the onset of glottal vibration. Since oral closure is marked spectrographically by the total or nearly total absence of acoustic energy in the formant frequency range and oral release is marked by a burst, an abrupt onset of energy in the formant frequency range, the point of voicing onset was determined by locating the first of regularly spaced vertical striations which indicate glottal pulsing, while the instant of release was defined by fixing the point where the pattern shows an abrupt change in the overall spectrum.

Average values of voice onset time for the 5 recordings of each word, as produced by the 5 subjects in stressed and unstressed position, were extracted. Average values of voice onset time for the group of words with the same voiceless plosive produced by each of the subjects in stressed position were also extracted. The results obtained were compared among the different subjects and in relation to measures suggested as specific for English and Portuguese. In the comparison of the results, the VOT measures presented by Lisker and Abrahson [6] were taken as reference for English and the ones presented by Behlau [5] were admitted for Portuguese. It should be noted that the referent values for the English language were based on the measurement of words taken in isolation and the Portuguese ones were based on syllable data, while this present research extracted the values from the recording of words inserted in carrier sentences. Average measures of VOT for voiceless plosives in non-stressed position were included in the experiment so as to corroborate the observations driven from the measures taken in stressed position.

3. THE RESULTS OF THE EXPERIMENT

In this section, we present our findings. The values of averages and ranges of voice onset time obtained in the experiment for each word as produced by the subjects are displayed in tables 1-6. The first line registers the words selected for the experiment. The letters in the first column of the tables indicate the different subjects as they are identified in section 2.2. The measures obtained from the recordings of each subject are listed along the line with their respective identification letters. The average value extracted from the 5 recordings of each word is indicated by **A**, and the range of variation among the 5 emissions is represented by **R**.

	Papa		**P**ala	
L	A: 7,57	R: 5,50/12,88	A: 7,42	R: 6,84/8,27
P	A: 12,84	R: 11,19/13,69	A: 17,89	R: 12,35/20,67
M	A: 19,42	R: 13,45/23,39	A: 20,00	R: 16,32/23,19
T	A: 9,71	R: 8,83/10,36	A: 10,04	R: 9,62/10,52
S	A: 22,47	R: 19,03/26,39	A: 21,08	R: 15,70/25,78
	Pot		**P**aper	
L	A: 15,85	R: 13,62/19,12	A: 8,7	R: 6,24/12,42
P	A: 74,46	R: 50,14/87,93	A: 67,16	R: 59,46/74,32
M	A: 86,63	R: 68,84/104,89	A: 49,17	R: 38,56/72,51
T	A: 71,48	R: 57,85/106,83	A: 58,80	R: 44,79/67,97
S	A: 78,48	R: 59,95/95,93	A: 68,23	R: 57,57/77,18

Table 1. Voice onset time in Msec- /p/ initial stressed position

	Tala		**T**apa	
L	A: 14,27	R: 12,89/15,88	A: 13,81	R: 13,14/15,33
P	A: 14,08	R: 12,02/15,66	A: 21,89	R: 10,84/44,61
M	A: 16,97	R: 16,21/17,99	A: 16,45	R: 15,57/17,74
T	A: 14,46	R: 12,57/15,43	A: 13,68	R: 11,98/15,71
S	A: 30,05	R: 23,08/35,19	A: 15,47	R: 11,51/19,01
	Top		**T**ape	
L	A: 13,61	R: 12,26/14,62	A: 14,61	R: 10,13/22,45
P	A: 72,02	R: 60,37/83,09	A: 79,60	R: 74,98/88,73
M	A: 73,30	R: 53,02/103,37	A: 60,14	R: 51,84/85,13
T	A: 70,22	R: 56,76/87,75	A: 52,69	R: 39,20/89,62
S	A: 59,48	R: 44,27/69,26	A: 88,02	R: 71,31/101,22

Table 2. Voice onset time in Msec- /t/ initial stressed position

	Cala		**C**apa	
L	A: 52,28	R: 45,27/67,15	A: 42,52	R: 35,13/45,97
P	A: 48,31	R: 42,51/53,10	A: 45,20	R: 41,15/50,75
M	A: 55,64	R: 45,40/67,28	A: 47,03	R: 40,70/55,06
T	A: 27,72	R: 25,12/30,16	A: 30,69	R: 26,62/37,29
S	A: 46,32	R: 36,29/53,26	A: 44,21	R: 36,88/55,27
	Cot		**C**ake	
L	A: 43,70	R: 31,61/62,37	A: 50,29	R: 35,81/60,66
P	A: 93,48	R: 69,10/118,06	A: 80,87	R: 75,80/84,62
M	A: 98,24	R: 79,90/111,87	A: 80,84	R: 70,95/89,54
T	A: 85,39	R: 119,95/77,47	A: 72,45	R: 63,28/83,17
S	A: 58,75	R: 47,87/70,44	A: 72,41	R: 51,62/94,23

Table 3. Voice onset time in Msec- /k/ initial stressed position

Tables 4 - 6 refer to the measures obtained for the VOT of /p, t, k/ in medial unstressed position. Since this position is exemplified by single words in both languages, the Portuguese and English reference values (**V**) are added with the target words on the first line.

	A**p**agou V: 10,37		Ta**pp**ing V:34	
L	A: 14,35	R: 11,37/19,85	A: 24,86	R: 18,29/31,12
P	A: 11,76	R: 10,74/13,14	A: 24,15	R: 21,18/30,24
M	A: 16,69	R: 13,14/21,49	A: 20,48	R: 11,60/29,69
T	A: 9,95	R: 8,28/10,67	A: 16,05	R: 11,69/32,08
S	A: 28,48	R: 23,37/30,41	A: 20,24	R: 19,07/23,02

Table 4. Voice onset time in Msec- /p/ medial unstressed position

	A**t**acou V: 14,93		
L	A: 13,99	R: 12,84 / 16,80	
P	A: 12,23	R: 10,67 / 14,43	
M	A: 16,94	R: 16,61 / 17,18	
T	A: 14,36	R: 12,72 / 17,67	
S	A: 13,15	R: 11,59 / 14,47	

Table 5. Voice onset time in Msec- /t/ medial unstressed position

	A**c**abou V: 43,73		pa**ck**ing V:49	
L	A: 50,33	R: 41,31/59,44	A: 64,69	R: 58,55/82,95
P	A: 26,96	R: 23,66/32,65	A: 65,94	R: 61,18/70,89
M	A: 50,17	R: 48,11/54,00	A: 50,27	R: 42,86/66,98
T	A: 26,26	R: 25,18/26,96	A: 27,12	R: 22,45/33,33
S	A: 29,12	R: 23,78/34,82	A: 52,25	R: 43,96/60,83

Table 6. Voice onset time in Msec- /k/ medial unstressed position

Tables 7 - 9 present the average values of voice onset time for the group of words with the same voiceless plosive produced by each of the subjects in initial stressed position in both languages. The letters on the first line of the tables indicate the different subjects. The measures related to them are listed under their respective identification letters. Reference values (**V**) for Portuguese (**P**) and English (**E**) are added in the first column.

	L	**P**	**M**	**T**	**S**
P **V**:10,37	7,49	15,36	19,71	9,8	21,77
E **V**: 58	12,27	70,81	67,90	65,14	73,35

Table 7. VOT average values in Msec- /p/ initial stressed position in English and Portuguese

	L	**P**	**M**	**T**	**S**
P **V**:14,93	14,04	17,98	16,71	14,07	22,76
E **V**: 70	14,11	75,81	66,72	61,45	73,75

Table 8. VOT average values in Msec- /t/ initial stressed position in English and Portuguese

	L	**P**	**M**	**T**	**S**
P **V**:**4**3,73	47,40	46,75	51,33	29,20	45,26
E **V**: 80	46,99	87,17	89,54	78,92	65,58

Table 9. VOT average values in Msec- /k/ initial stressed position in English and Portuguese

4. DISCUSSION OF RESULTS

From a comparison of the mean values and ranges given in the tables just presented, it is quite clear that measure of voice onset time is, to a certain extent, sensitive to the place of stop closure; velars having consistently higher values than the other stops in both Portuguese and English. Portuguese has voiceless unaspirated stops, except in the velar position, where a longer VOT lag results in aspiration and in a burst with stronger intensity corroborating Lisker and Abrahnson's findings [6] that unaspirated voiceless plosives are those in which voicing is delayed (20-25 msec) in relation to the burst release, whereas aspirated voiceless plosives are characterized by a delay of more than 25 msec.

There is some effect to embedding the stops in medial unstressed position: voice onset time values tend to be a bit compressed in comparison to the values measured in word initial stressed position. In the several cases where this compression occurs, as table 5 shows, there is a reduction of the gap separating those categories from each other along the voice onset time dimension, without overlapping between distinct categories, though. Therefore, both vowel duration and voice onset time are likely to be affected by temporal compression due to word stress. The average values displayed on the tables 4-9 seem to indicate that the compression of VOT rates is not automatic for all the subjects thus requiring training in the various phonetic environments.

It is generally assumed that when exposure to a second language is postponed untill adulthood, the learner's ability to perceive and produce second language speech sounds will be limited because the ability to discriminate non-native contrasts decreases. All the subjects who took part in this experiment, except for subject **M,** acquired their second language when adults. Nevertheless, the results displayed in the tables point at distinct levels of command of VOT in both languages.

The measures referring to subject **L** in tables 8 and 9 confirm Flege's [1] statement that the learner has a tendency

to consider similar phones of a foreign language as identical. Subject **L** transfers the short-lag VOT of Portuguese /t/ and /k/ to English. In spite of bringing in a longer VOT for the English plosive /p/ than for the equivalent Portuguese one, the values achieved in table 7 are below the English references. The Portuguese and English voiceless stops of this subject are similar but the data do not seem to show overlapping of the measures taken as reference. The short-lag VOT results in the absence of the typical aspiration of the English /p, t/ and in a burst of weak intensity in the aspiration present in /k/ with the subsequent strong foreign accent in the subject **L**'s recordings. Such problems remain after 3 years of exposure to formal foreign language teaching,which seems to bear little influence in this case, once training did not improve performance. Many production difficulties have an underlying perceptual difficulty which may be overcome with appropriate training techniques which include varied tasks, phonetic environment, so as to simulate the conditions found in a speaking community.

Furthermore, some of the data presented in the tables corroborate Jamieson's assertion [3] that reduction in the learner's discrimination ability, attributable to a change in attention to acoustic cues, may be overcome by effective training.This would focus attention on the acoustic cues that are relevant for sorting out the phoneme contrasts of the foreign language. Subject **P**'s VOT measures are, in general, consistently higher than the values expected for both languages, which seems to be a personal characteristic. Nevertheless, the distance from the referent values is bigger for the English plosive /p/, which is characterized by the shortest VOT lag among the Portuguese voiceless plosives. The subject's linguistic background seems to have interfered with her production leading to a phonetic approximation. Her awareness of the phonetic features of Portuguese and English plosives seems to impose a longer lag on the production of the English /p/, which results in a burst of high intensity that sounds unnatural.The perception of non-native phoneme contrasts persists throughout her data emphasizing the importance of phonetic hints to L2 phonogical learning.

Subject **T**, native speaker of English, nearly always presents VOT values slightly lower, but still very close to the referent ones in both languages, indicating a surprising good command of the Portuguese plosive time dimension. Nonetheless, the failure in producing Portuguese /k/ with the expected VOT measures leads us to the conclusion that the subject, who is also a speaker of Spanish with a master degree in the area, and who lived in Spain for 3 years, is, in fact, reproducing the Spanish values in Portuguese. Considering that the referent value for Spanish /k/ is in between 25 and 29 msec [6], the subject's average value (29,20) for Portuguese /k/, as well as the general tendency to present lower values for all Portuguese plosives, could be explained in terms of Spanish interference. The values presented for Portuguese indicate that the subject has a complete mastery of the Spanish voiceless plosives and that, when adults are appropriately trained attending to the cues which are relevant to phonemic distinctions, their abilities to perceive non-native speech contrasts can improve substancially.

Subject **M**, early-bilingual and native speaker of Portuguese, and Subject **S**, native speaker of English, exhibit unstable values in both languages. Their Portuguese plosives /p, k/ characterized by higher measures than the referent ones could be explained by English interference. The occurrence of unstable values could be justified through accomodation toward the speech of the two language communities to which the speakers are daily exposed. Fowler suggests[4] that such accomodation occurs because listeners have a predisposition to imitate what they hear and also for social reasons, for purposes of identification with a social group.

5. CONCLUSION

The measures of voice onset time, besides being a device for distinguishing stop categories in a specific language, enable us to pinpoint at some factors which seem to be influencing the proficiency achieved by the subjects in the production of non-native voiceless plosives in English and Portuguese such as: different linguistic background, degrees of exposure and social identification with speaking communities, teaching which includes phonetic input and relevant cues in the assigning of phonemic contrasts. It has to be emphasized that the data employed in this study are rather limited and only permit tentative conclusions.

ACKNOWLEDGMENTS

We would like to thank Sandra Madureira for her orientation and motivation throughout the development of our research, Eduardo Vanso Kawanishi and Leila Houf Marreiro for the technical support. We would also like to thank the subjects who agreed in participating in the experiment.

REFERENCES

[1] Flege, J. E., and Eefting, W. 1986. "Linguistic and Developmental effects on the production and perception of Stop consonants", *Phonetica*, vol. 43, pp. 155-171.

[2] Flege, J. E. 1990. Perception and Production: The Relevance of Phonetic Input to L2 Phonological Learning, in *Crosscurrents in Second Language Acquisition and Linguistic Theories* (C. Fergusson and T. Huebner, editors). Philadelphia: John Benjamins.

[3] Jamieson, D. G. 1995. "Techniques for Training Difficult Non-Native Speech Contrasts". *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Stockholm, Sweden, vol. 4, pp. 101-107.

[4] Fowler, C. A. 1995. "A Realist Perspective On Some Relations Among Speaking, Listening and Speech Learning". *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Stockholm, Sweden, vol. 3, pp. 330-334.

[5] Behlau, M. S. 1986. *Análise do Tempo de Início de Sonorização na Discriminação dos Sons Plosivos do Português*. Tese de Doutorado. Escola Paulista de Medicina. São Paulo.

[6] Lisker, L., and Abramson, A. S. 1964. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements", *Word*, vol. 20, pp. 384-422.