Effects of Perceptual Training on the Identification and Production of the English Voiceless Plosives Aspiration by Brazilian EFL Learners

Mara Silvia Reis[1], *Denize Nobre-Oliveira*[2]

[1]Federal University of Santa Catarina, Brazil
[2]Federal Center of Technological Education of Santa Catarina, Brazil
marasreis@hotmail.com, denizenobre@yahoo.com.br

1. Introduction

Within the world's languages, 88.9% (Kessinger & Blumstein, 1997) includes the voicing distinction among consonants, and voice onset time (VOT) is an acoustic cue associated with this distinction. VOT is the length of time measured between the release of a plosive and the beginning of vocal fold vibration (Cho & Ladefoged, 1997). This time, usually expressed in milliseconds (ms), is decisive for accurate perception of the voiceless and voiced stops /p, t, k/ and /b, d, g/, respectively.

Three major VOT phonation types are frequently described in the literature (Lisker & Abramson, 1964): (a) negative VOT, also called voicing lead, in which the beginning of the vocal fold vibration precedes the release of the plosive, a time that ranges from -125 msec to -75 msec; (b) zero VOT, also called short voicing lag, in which the beginning of the vocal folds vibration is simultaneous, or almost simultaneous, with the plosive release, a time ranging from 0 to +35 msec; and (c) positive VOT, or long voicing lag, in which the plosive is released and there is a delay in the beginning of the vocal fold vibration, ranging from +35 msec to +100 msec.

Although there is not an absolute value for the VOT of each plosive, some researchers (Kent & Read, 1992; Toribio, Bullock, Botero, & Davis, 2005) point out that typical measurements in English range from 30 to 80 ms, /k/ having the longest VOT, about 80 ms, /t/ around 70 ms, and /p/ about 55 ms. On the other hand, Istre (cited in Klein, 1999) states that the approximate VOT values in Brazilian Portuguese are 38 ms for /k/, 18 ms for /t/, and 12 ms for /p/.

English and Brazilian Portuguese (BP) demonstrate different realizations of voiceless stop VOT depending on the phonological context. While in BP the voiceless stops tend to be produced with short voicing lag in any context, in English they are realized with long voicing lag in word-initial position and in stressed syllables. Thus, in these contexts, English /p/, /t/, and /k/ are aspirated as [p^h], [t^h] and [k^h].

Bearing in mind the differences between BP and English VOT, the present study aims at verifying whether adult Brazilian learners of English as a foreign language (EFL) perceive these distinctions, and whether perceptual training with English voiceless stops can modify perception and production of /p/, /t/, and /k/.

2. Review of literature

Much of the research on the acquisition of VOT has dealt with the establishment and the maintenance of phonetic categories, either in perception or production. In terms of cross-language studies, research has investigated the realization of VOT among languages that differ in patterns, such as English and Spanish (Flege, 1991; Flege & Munro, 1994; Bohn & Flege, 1993; Flege & Eefting, 1987, 1988; Toribio et al., 2005), English and French (Saerens, Serniclaes, & Beeckmans, 1989; Hazan & Boulakia, 1993; Kessinger & Blumstein, 1997), and English and BP (Major, 1987; Sancier & Fowler, 1997; Rocca, 2003; Cohen, 2004).

Whether adult L2 learners are able to attain native-like VOT production is a controversial issue that researchers have been attempting to answer. While some authors argue that it is not likely to occur (Williams, 1980; Flege & Eefting, 1987, 1988), others, as Major (1987), found that the hypothesis of a sensitive period is not an impediment for native-like VOT production. Toribio and colleagues (2005) investigated the production of VOT by Spanish and English bilinguals and hypothesized that the production of these speakers could not be assessed in terms of monolinguals norms, either in English or in Spanish. Indeed, Flege and Eefting (1987) found that Spanish-English bilinguals produced /p, t, k/ VOT values that were intermediate (57ms) to those observed for monolingual Spanish (26 ms) and monolingual English (87 ms).

Major (1992) reports that American English speakers living in Brazil for a period that ranged from 12 to 35 years had changed the patterns of both English and BP VOT. Likewise, Sancier and Fowler (1997) found that a BP speaker living in the United States modified the VOT values, both in English and Portuguese, according to the language setting in which she was inserted. Rocca (2001, 2003), investigating factors that might influence the production of English VOT values by Brazilian bilingual speakers of Portuguese and English, found that bilinguals produced VOT values different from monolinguals of both languages.

In reality, VOT values are not isolated entities within any phonological system. Instead, the values are inclined to undergo influence from either segmental or suprasegmental factors, such as prosody, phonetic contexts, sentence and syllable stress, age, and speaking rate (Klein, 1999; Cohen, 2004). Perhaps the most influential factor in VOT variation is the speaking rate—the faster it is, the smaller the measurements tend to be (Schmidt & Flege, 1996). Cohen (2004) asserts that at a fast speaking rate there is an overlap in the short and long lag values, so that the identification between voiced and voiceless stops is problematical. Clark and Yallop (1990) claim that, due to speaking rate, the literature is inconsistent in defining VOT categories—while some studies define the long lag at around 40 ms, others define it at 70 ms or more.

With reference to the relationship between perception and production of voiceless stop aspiration, Flege and Schmidt (1995), investigating Spanish bilingual speakers of English, found a significant correlation between the perception and production of VOT values. Rocca (2003) found that early bilinguals of BP and English do not separate the two phonological systems the way monolinguals do. Similarly, Cohen (2004) found a positive correlation between transfer of English VOT values to BP due to perception changes.

As regards studies with VOT perceptual training, Pisoni, Aslin, Perey and Hennessy (1982) examined the influence of training on the perception of a three-way VOT voicing contrast by monolingual speakers of English, a language that has only a two-way contrast. The results demonstrated the effectiveness of perceptual training on the categorization of the contrast not present in English. Other studies also examined adult English speakers training to discriminate a three-way voicing contrast (McClaskey, Pisoni, & Carrell, 1983; Tremblay, Kraus, Carrell, & McGee, 1997), and found similar results of those of Pisoni et al. (1982). Rochet (1995) investigated whether perceptual training could modify Mandarin speakers' perception and production of the French /t/-/d/ contrast. He reports that both perception and production were positively affected by training. Clarke and Luce (2005) examined English speakers categorizing /t/-/d/ produced with short lag and prevoicing leading, respectively, by the use of a perceptual task. They found that perceptual criteria were adjusted for the alveolar contrast and that this adjustment was generalized to a different place of articulation—the velars /k/ and /g/.

Therefore, these findings indicate that adults are able to learn to perceive and produce non-native speech contrasts through the use of perceptual training, and that training effects

may transfer to phonetic contexts not included in the training set. However, to the best of our knowledge, no study has been carried out investigating the effects of VOT perceptual training on perception and production of Brazilian EFL learners. Thus, the present exploratory study aims at shedding some light on the issue and suggests the following hypotheses (H):
H1: The VOT means for /p/, /t/, and /k/ production in the L1 and the L2 are different in the pre and posttests;
H2: The VOT means for the production pretest and posttest will vary according to type of consonant;
H3: The means of the identification pretest and posttest are different;
H4: The means of the discrimination pretest and posttest are different.

3. Method, materials and procedures

3.1 Participants

Eleven undergraduate students at the Universidade Federal de Santa Catarina participated in this study, 9 women and 2 men, their ages ranged from 17 to 44 years. At the time of data collection they had been studying English for an average of 4 years, and identified themselves as intermediate EFL speakers. No trainee had been to an English speaking country, nor did they report having any hearing or speaking impairment. In addition, in order to have a native baseline for the tests, two native speakers of American English took the perception tests and the training tasks.

3.2 Production tests

Two production tests were given, one in Portuguese, in order to measure the participants' voiceless stop VOT values in this language, and a second in English. The latter test was carried out twice, as pretest and posttest, while the former was recorded at the time of the pretest only.

In both Portuguese and English production tests the participants recorded a list of 21 CVC words, 7 for each voiceless stop in word-initial position. Total recordings resulted in 231 tokens for each production test (21 tokens X 11 participants), 77 for each voiceless stop.

3.3 Perception tests

Two perception tests were applied as pre and posttests—a discrimination and an identification test. The recordings of both tests were made by four adult American speakers of English, two male and two female, and four adult proficient Brazilian EFL speakers, two male and two female. The native speakers (NS) were instructed to produce the tokens with typical English VOT values, while the nonnative speakers (NNS) were asked to realize the tokens with typical BP values, although the words were in English. The stimuli of each test consisted of 48 randomized CVC words, 8 with each phoneme in word-initial position, and each word was repeated twice. As a result, each pre and posttest provided 528 tokens to be analyzed (48 tokens X 11 participants).

Concerning the discrimination test, the 48 words were equally divided in terms of aspiration and gender. The number of same or different trials was also controlled; i.e., each /p/, /t/, and /k/ had 24 *different* and 24 *same* trials. The test lasted about 6 minutes and was given before the identification test in both times, pre and posttest. The tokens were played with 1.5 second of inter-stimulus interval, and 3 seconds of between trials interval. The participants heard the two words once and were asked to mark whether the sound of the first consonant of the two words belonged to the same category or if they were different.

As regards the identification test, which took around 4 minutes to be completed, the randomized words were presented in isolation, with an interval of 2 seconds between trials. The participants heard the word once and had to mark *yes* or *no* for native-likeness.

3.4 Perceptual training

Four other adult American NS of English recorded the tokens for the perceptual training, two men and two women. Each talker recorded a set of 30 CVC words with the voiceless stops in word-initial position, 10 with each stop. Originally, these 10 words had very high VOT values, even for the English pattern—a mean of 108 ms for /p/, 130 ms for /t/, and 145 ms for /k/. The values were then manipulated with the use of the software Audacity 1.2.4, all shortened to the standard BP values (as described by Istre, cited in Klein, 1999), a mean of 14 ms for /p/, 25 ms for /t/, and 39 ms for /k/.

A discrimination and an identification perceptual training tasks were given. The procedures and the design of the tasks were exactly the same as those of the tests. The difference is that now the tokens were all produced by NSs, the participants had feedback on the correct response after 3 seconds of the tokens' presentation, they were given an eight-trial-familiarization practice before they carried out the training tasks, and there were 60 trials in each task. Thus, a total of 660 tokens were collected (60 tokens X 11 participants) for each task. The discrimination training lasted about 10 minutes, while the identification, 8 minutes.

The data collection originally consisted of a set of 231 words in the BP production test (21 words x 11 participants), and 231 words in the English production tests, both pre and posttest. However, the analysis of the production of /t/ demonstrated that when this consonant was followed by /i/ or /ɪ/, it was produced with long lag voicing values due to its palatalization. Because the affricate value of /tʃ/ could cause misleading results for the /t/ VOT value, the word was removed from the production analysis. Thus, each production test had 20 words analyzed, 7 for /p/ and /k/, and 6 for /t/, resulting in a total of 220 words per production test.

The data set of the perception tests consisted of 1056 trials for each test (528 in the discrimination test and 528 in the identification test—48 trials x 11 participants), while the perceptual training had 1320 trials (660 in the discrimination task and 660 in the identification task). The results were statistically analyzed in order to verify whether there was any significant difference in the participants' perception and production of English VOT values.

4. Results and discussion

As regards production tasks, Table 1 displays the results in the BP test, and English pre and posttests, as well as the expected values in each language task. As can be seen, there was a general increase in VOT values from pre to posttest, although the Portuguese ones are not consistent with the literature.

Table 1. Mean VOT values for the Portuguese production task (Port.), the English production pretest (Pre), and the English production posttest (Post). *SD*: Standard deviation

	/p/			/t/			/k/		
	Port.	**Engl.**		**Port.**	**Engl.**		**Port.**	**Engl.**	
Literature	12	55		18	70		38	80	
		Pre	**Post**		**Pre**	**Post**		**Pre**	**Post**
Mean VOT	17,27	18,55	31,36	23,55	32	36,91	46,55	46,36	55,18
SD	6.03	9.22	11.16	8.38	7.44	11.75	9.05	11.10	11.80

Regarding H1, whether perceptual training could affect the production of individual voiceless stops, the means of each consonant were compared with their own means, in each test and in each time. Thus, the means of the production of /p/ in BP, for example, were compared with the means of /p/ in the English pre and posttests. Overall, the values for /p/, /t/, and /k/ seem to indicate that training had a positive effect on production, since the values are heading toward a more English-like pattern. However, not all consonants had a statistically significant improvement. On the one hand, an ANOVA reveals that for /p/, training seem to have improved its production [$F(2,11)= 8.41$, $p<.003$], and the post hoc confirmed that there was a significant difference from the BP to the English posttest and from the English pre to posttest. Similarly, the improvement in the production of /t/ was statistically significant [$F(2,11)= 7.7$, $p<.011$], although only between BP and English pretest, and BP and English posttest.

On the other hand, the production of /k/ did not show a significant effect of training [$F(2,11)= 2.62$, $p<.103$]. Although there seems to be a movement toward the English values, no mean comparison yielded a significant difference. Thus, H1 is partially supported—perceptual training seemed to have significantly affected the production of /p/ and /t/ only. Although the results did not achieve the English VOT pattern, after training all consonants had different values from BP to English posttest.

As regards H2, or whether the production in the pretest and posttest would vary according to type of consonant, no voiceless stop seemed to be more difficult, or easier, to be produced than the others (see Table 2).

Table 2. Mean Z-scores the English production pretest (Pre) and the English production posttest (Post). *SD*: Standard deviation. df: degrees of freedom. *F*: ratio for ANOVA

	/p/		/t/		/k/	
	pre	**post**	**pre**	**post**	**pre**	**post**
Mean Z-score	2.60E-16	-2.77E-16	4.43E-16	1.837E-16	-3.35E-16	-4.54E-16
SD (df, *F*)	1 (2, 0)	1 (2, 0)	1 (2, 0)	1 (2, 0)	1 (2, 0)	1 (2, 0)

Since the individual values of each consonant are intrinsically different, it would not be possible to compare their means. The raw scores were then converted into Z-scores and, subsequently, the means of each consonant were compared with the means of the two other consonants, in each time. The results demonstrated that no consonant, in either pre or posttest, imposed more or less difficulty to be produced than others. Thus, H2 was not corroborated.

Concerning the results of the two perception pretests, Table 3 demonstrates that only identification seemed to be positively affected by treatment. While in the identification pretest the participants achieved a mean of 32.82, out of 48, in the posttest it increased to 42.82. A paired-sample *t*-test revealed that the difference is statistically significant ($t(10)= -4.94$, $p<.001$), which suggests that even short perceptual training can affect the identification of nativelike VOT in the voiceless stops. These results support H3, which suggests that the means of the identification pretest and posttest are different.

Table 3. Number of correct responses, out of 48, in the identification pretest (Ident. pre), identification posttest (Ident. post), discrimination pretest (Disc. pre), and discrimination posttest (Disc. post). *SD*: standard deviation

	Ident. pre	Ident. post	Discr. pre	Discr. post
Mean	32.82	42.82	42.82	32.09
SD	7.22	2.4	2.4	6.69

The discrimination test, however, showed a significant negative effect of training ($t(10)= 4.96$, $p<.001$). Not only did the participants attain a lower number of accurate responses in the posttest, but they also varied more. Thus, H4, which predicts that means of the discrimination pretest and posttest would be different, is not confirmed.

The contrasting pictures of the identification and discrimination results suggest a number of pitfalls in the use of discrimination training. To start with, as several researchers point out (Jamieson & Morosan, 1986; Akahane-Yamada et al., 1999; Logan & Pruitt, 1995; Hardison, 2003), discrimination tasks may not be an appropriate perceptual training technique. The writers argue that discrimination tasks tend to enhance intraphonemic sensitivities, which may be processed as perceptually irrelevant features.

Furthermore, even though the design of the test and the training was the same, the fact that talkers from different L1 backgrounds were employed in each task may have complicated discrimination task and, eventually, confused trainees' capacity of judgment. Table 4 presents participants' percentages of accurate answers in the identification and discrimination pre and posttest, as well as in the identification and discrimination training.

Table 4. Trainees' percentage of accuracy in the identification pretest (Ident. pre), identification posttest (Ident. post), identification training (Ident. train), discrimination pretest (Disc. pre), discrimination posttest (Disc. post), and discrimination training (Disc. train)

Tasks	Ident. Pre	Ident.train	Ident.post	Dis. pre	Dis. train	Disc. post
% of accuracy	68	69	89	89	62	66

As can be seen again, not only had the discrimination results been poorer after training, but they also were lower in the training task itself. These results seem to support the argument that discrimination tasks may not be adequate for perceptual training of cross-language differences in voiceless stop VOT values. An aspect that may be even more important is that the design used in the testing and training tasks was not adequate. First, because in the testing tasks NSs and NNSs talkers provided the stimuli, while in the training tasks only NSs recorded the stimuli. When asked to produce the English words without aspiration, the talkers reported having great difficulties to produce the accurate vowels without the previous aspiration of the consonants. Baptista (personal communication, December, 2006), suggests that maybe the talkers had to access their BP phonological system in order to shorten the VOT values, which may have caused them to produce the vowels in a more BP manner as well. Thus, the present study's stimuli provided by NNSs for the testing tasks not only had unaspirated consonants, but also the entire word sounded less native-like. It seems that such differences turned out the tests much simpler to identify or discriminate than the training tasks.

Second, the NS who took the perceptual training tasks (which had only NS recordings) reported that the discrimination of aspiration is not an easy task, since short VOT values for voiceless consonants may appear in ordinary conversations, depending on the speech rate, without any harm for native-likeness. The trainees, on the other hand, reported that the discrimination task was very difficult and frustrating. Since they were receiving feedback immediately after their choice, they could accompany their performance and, given that the performance was, in their opinion, unsatisfactory, it seems that they became discouraged and confused for the following discrimination posttest.

A last aspect of the training tasks that is worth mentioning is the trainees' and NSs' report on the something else than VOT values. They argued that even when the words had short values, they still could feel an intensity in the consonant that they associated with aspiration, a realization that is different from words which are naturally produced with short voicing lag. Given that the original stop consonants were produced slowly and with exaggerated values, an average of 108 ms for /p/, 130 ms for /t/, and 145 ms for /k/, maybe the synthetic shortening was not able to reduce this trace of exaggeration, which was felt as aspiration. Thus, even though the NS stimuli were manipulated in order to approximate the English VOT values to the BP ones (a final average of 14 ms for /p/, 25 ms for /t/, and 39 ms for /k/), it seems that shorter values did not satisfy the trainees' necessity of cues in the perception training tasks.

In spite of all these problems found with the use of discrimination in the test and training task, since this was an exploratory study we decided to maintain the task in order to investigate how the voiceless stops would behave with this type of technique. As Lieberman and Blumstein (1993:154) state, the isolation of a relevant linguistic acoustic cue demands the use of psychoacoustic experiments "even when the acoustic cue seems to be very 'simple'."

To sum up, despite all the limitations of the perceptual training, it seems that it was able to affect, in different degrees, both perception and production of voiceless stops. With the exception of /t/ followed by /ɪ/, in the production tests, all other contexts seem to be heading toward the English VOT patterns. Within the three hypotheses that involved effects of training, two were corroborated—the production of /p/ and /t/, and the identification of the consonants improved after training.

5. Conclusion, limitations and suggestions for further studies

As could be expected from an exploratory study, the results demonstrated a considerable number of drawbacks. A major concern is the lack of a control group and the limited number of participants. Any claim in favor of training effects may be disregarded as due to testing practice effect. Randomization of the stimuli in the pre and posttests could possibly minimize the practice effect. Also, unexpected moderator variables, such as participants' fatigue or eagerness, and the influence of learning maturation, may also have affected the results of a small group such as the one of the study. Also, a follow-up test could be applied in order to verify whether the effects of perceptual training were observable after a certain period.

Another important limitation is the inconsistency between the testing and training tasks. More than simply rejecting the appropriateness of discrimination test or training, perhaps, as discussed above, the use of talkers from different L1 backgrounds in the perception tests and perceptual training probably made the discrimination task a puzzling exercise. Thus, further studies could work only with native stimuli in testing and training tasks, since nonnative stimuli seemed to have provided other cues rather than aspiration only. Also, these stimuli could be produced with different, but native-like, length of VOT lag. Then, if manipulation of the values were necessary, it could avoid leaving behind the trace of exaggeration that the trainees felt as aspiration.

Another possibility is the use of only identification as testing and training tasks. This could guarantee, as some authors argue (Jamieson & Morosan, 1986; Akahane-Yamada, et al., 1999; Logan & Pruitt, 1995; Hardison, 2003), that trainees would not face irrelevant intraphonemic categories, which in turn can mislead the establishment of new L2 phonemes.

However, there seems to be some other possibilities in the use of discrimination tasks that make it worth investigating in future studies. For example, perceptual training could start with identification tasks up to a point in which the trainees demonstrated a high level of accurate answers, say 90% or 95%. It would be expected, perhaps, that at this level of accuracy the establishment of the new phonetic category had occurred. After this point, then, discrimination tasks could be implemented in stages. They could start with trials comparing only exaggerated differences, and gradually these differences could be diminished. Again, trainees could only advance to other stages in the discrimination task if they had obtained a high number of accurate responses. A design like this could, maybe, make it possible to verify generalization, i.e., whether the effect of one training task is transferred to other type of task.

Another aspect of generalization that could be further studied is the phonetic context. The present study did not use all vowel contexts possible either in the BP or English perception and production tests. Equally ignored, since it was not the aim of the study, was the analysis of whether the words that appeared in the perceptual training attained a higher perception score or were more accurately produced than words not present in the training.

Still regarding phonetic contexts, the influence of the following vowel on the production of the VOT has been investigated in several studies (Klatt, 1975; Chang, 1998), which demonstrated that aspiration tends to be shorter if followed by low vowels, and longer if followed by high vowels. Further studies could verify whether the type of preceding voicing lag can interfere in the production of vowels.

Future studies could still have short training tasks as the one of the present study, but in order to avoid the evident confusion that the discrimination training caused, maybe it could be extended in the number of training sessions, using only one consonant per session. Also, fewer but repeated words could be used instead of the eight words with each consonant as in the present experiment. Another procedure that could be investigated is the employment of the perception posttest on a different day from the perceptual training. The trainees of the present study reported that not only were they confused by the training task, but also their memory was exhausted.

In terms of L2 learning, a study like this could investigate whether it would be more profitable for native speakers of languages that have only the short and long voicing lag VOT, such as English, to perceive and produce VOT values of languages that have only prevoicing and short voicing lag, such as BP. As Bohn and Flege (1993) claim, English speakers seem to show more difficulty in distinguishing a short voicing lag from a prevoicing lag than a Spanish or Portuguese speaker to distinguish between short and long voicing lags.

Finally, in terms of L1, this type of perceptual training could be used by speech therapists with patients who misunderstand and misproduce the plosives in any word-position. Perhaps, the confronting of /p/ and /b/, for instance, in discrimination and identification tasks could, eventually, improve both perception and production of these consonants in natural speech. The training task could be used as home-reinforcement for the sessions, which could, in turn, reduce the costs of the treatment, an aspect especially important in a country like Brazil.

References

Akahane-Yamada, R., Tohkura, Y., Bradlow, A., & Pisoni, D. (1999). Does training in speech perception modify speech production? *Proceedings of the International Congress of Phonetic Sciences* (pp. 117-120). San Francisco.

Bohn, O.-S., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics 21*, 267-290.

Bradlow, A. R., Yamada, R. A., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977-985.

Chang, S. S. (1999). Vowel dependent VOT variation. *Proceedings of the International Conference on Spoken Language Processing*. San Francisco.

Cho T., & Ladefoged, P. (1997).Variations and universals in VOT. *Fieldwork studies of targeted languages V: UCLA Working Papers in Phonetics*, *95*.

Clark, J., & Yallop, C. (1990). *An introduction to phonetics and phonology.* Oxford: Basil Blackwell.

Clarke, C. M., & Luce, P. A. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception* (pp. 23-26). London.

Cohen, G. (2004). The VOT dimension: A bi-directional experiment with English and Brazilian-Portuguese stops. Unpublished Master's thesis, Universidade Federal de Santa Catarina, Florianopolis, Brazil.

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native speakers of Spanish. *Journal of Phonetics, 15*, 67-83.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America. 83*(2), 729-740.

Flege, J. E., & Munro, M. J. (1994). The word unit in L2 speech production and perception. *Studies in Second Language Acquisition*, *16*, 381-411.

Flege, J. E., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants*. Phonetica, 52,* 90-111.

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

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*, 495–522.

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech, 36*(1), 17-38.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants in Japanese learners of English. *Speech Communication, 47*, 360-378.

Istre, G. L. *Um estudo do VOT em monolíngües brasileiros***.** Unpublished manuscript, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America,* 118, 3267-3278.

Jamieson, D., & Morosan, D. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43*, 88-96.

Jamieson, D., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics, 40*(4), 205-215.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech.* San Diego: Singular.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*, 143-168.

Klatt, D. H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18,* 686-706.

Klein, S. (1999). Estudo do VOT no Português Brasileiro. Unpublished Master's thesis. Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word* , *20*, 384-422.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp.351-376). Timonium, MD: York.

Major, R. C. (1987). English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics. 15,* 197-202.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal, 76*, 190-208.
McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics, 34*, 323–330.
Ortega-Llebaria, M., Faulkner, A., Hazan, V. (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Speech, Hearing and Language: UCL Work in Progress*, *13*, 39-51.
Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology. Human perception and performance, 8*, 297–314.
Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Acoustical Society of America 119*(3),1684-1696.
Rocca, P. D. A. (2001). Fala e local de origem: Uma questão de sotaque. *Intercâmbio: Uma Publicação de Pesquisas em Lingüística Aplicada, X,* 215-225. São Paulo: EDUC.
Rocca, P. D. A. (2003). O desempenho dos falantes bilíngües evidências advindas da investigação do VOT de oclusivas surdas do inglês e do português. *Documentação de Estudos em Lingüística Teórica e Aplicada, 19*(2), 303-328.
Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Timonium. MD: York Press.
Saerens, M., Serniclaes, W., & Beeckmans, R. (1989). Acoustic versus contextual factors in stop voicing perception in spontaneous French. *Language and Speech, 32*(4), 291-314.
Sancier, M. I., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25,* 421-436.
Schmidt, A. M., & Flege J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica, 53*, 162-179.
Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics, 36,* 131–145.
Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K.A. (2005). Perseverative phonetic effects in bilingual code-switching. In R. S. Gess, & E. J. Rubin (Eds.), *Theoretical and experimental approaches to Romance linguistics* (pp. 291–306). Salt Lake City: John Benjamins.
Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America, 102*, 3762-3773.
Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *34*, 539-552.
Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child Phonology* Vol. 2: *Perception* (pp. 185-216). New York: Academic Press.