Perceptual Training on the Pronunciation of /s/-Clusters in Brazilian Portuguese/English Interphonology

Melissa Bettoni-Techio, Rosana Denise Koerich

Federal University of Santa Catarina, Brazil
melissa@formatto.com.br, rkoerich@matrix.com.br

1. Introduction

The perception of nonnative contrasts is an important topic under investigation in the field of second language acquisition (SLA) (e.g., Baptista, 2006; Best, 1995; Flege, 1988, 1995). However, the teaching of pronunciation in the foreign language (FL) classroom has not been given the necessary importance (Piske, 2007).

Accounting for the necessity of proposing materials and methods to fill in the gap, the present study investigates the effects of a computer-assisted pronunciation teaching approach, namely, perceptual training, which is intended to provide FL learners with substantial native L2 input. The object of training in this study, English word-initial /s/-clusters, has shown to be a structure of difficult acquisition in both first language (L1) and L2/FL (Yildiz, 2005).

In Brazil, studies on the production of these clusters have shown that the /s/ in /s/-stop clusters is frequently epenthesized (prothesis), and the /s/ in /s/-sonorant clusters is both epenthesized and voiced (e.g., Cornelian, 2003; Bonilha & Vinhas, 2005; Rauber, 2006; Rebello & Baptista, 2006). In Brazilian Portuguese (BP), these processes are found in loanwords, as for instance, in *stress* which is pronounced and written with prothesis and paragoge – 'estresse' [isˈtrɛsɪ], and *slogan* pronounced with prothesis and voiced /s/ – [ɪzˈlogã]. Depending on idiolect, palatalization of /s/ in these clusters is frequent; so that 'estresse' may be pronounced as [ɪʃˈtrɛsɪ] and *slogan* as [ɪʒˈlogã].

Silveira (2002) investigated the relationship between perception and production of /sC(C)/ in BP/English interphonology and found a correlation between the two abilities in terms of type of cluster, that is, the more accurately perceived clusters were the ones more accurately produced.

Two current models of cross-language speech perception and acquisition, Flege's (1995) Speech Learning Model (SLM) and Best and Tyler's (2007) Perceptual Assimilation Model for second language learners (PAM-L2), which was originated from the Perceptual Assimilation Model (PAM) (Best, 1995), relate perception and production of nonnative contrasts to some extent, suggesting that errors in production may be motivated by misperception of L2 sounds. As mentioned above, BP learners of English tend to turn word-initial /sC(C)/ into the native language permitted initial sequence /VsC(C)/ and to voice the /s/ in /s/-sonorant clusters, processes which are probably due to L1 phonotactics influence on adults' L2 perception (Flege, 1989; Hallé, Segui, Fraunflender, & Meunier, 1998; Sebastián-Gallés, 2005). In this sense, Beddor, Harnsberger, and Lindemann (2002) have shown that nonnative perception is influenced by L1 coarticulatory patterns, and Harnsberger (2000, 2001) has proposed that it is influenced by allophonic or other phonetic variations.

Both the SLM and the PAM-L2 claim that adults can acquire L2 sound categories through experience with quality and amount of input being essential. Technology has brought advances to computer-assisted learning and these advances can be applied to FL learning; however, much has to be done to develop FL pronunciation tools that provide learners with the input they might need in order to be able to perceive and produce non-native speech with great accuracy.

Students looking for self-study materials can benefit from perceptual training programs (Wang & Munro, 2004) because they are efficient in reinforcing traces and consequently forming robust sound categories (Ellis, 2005). Thus, in order to improve perception and production of nonnative contrasts, L2/FL input has been provided by perceptual training materials using identification tasks. Successful intents have been reported by Bradlow, Pisoni and Akahane-Yamada (1997), Hardison (2004), Hazan, Sennema, Iba and Faulkner (2005), among others. The perceptual training studies have given special attention to the /r/-/l/ contrast regarding Japanese speakers (e.g., Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997), but other nonnative contrasts have been investigated as well (e.g., /θ/-/ð/, as in Jamieson & Morosan, 1986; /t/-/d/, as in Flege, 1989; /s/-/z/, as in Trapp & Bohn, 2000; CV+alveolopalatal versus CV+alveolopalatal+V, as in Yeon, 2004).

2. Perceptual training

Ellis (2005) stated that providing learners with input of words naturally spoken by multiple talkers and with immediate feedback on the identification of these words in an adaptive training fashion leads to rapid learning. The effectiveness of this type of training is due to the formation of traces originated from the attended details of the perceptual event and their storage in long-term memory (Hardison, 2004, 2005).

The enthusiasm researchers have towards identification perceptual training comes from positive results of research on perceptual training regarding nonnative speech contrasts at both segmental (e.g., Jamieson & Morosan, 1986; Flege, 1989; Lively, Logan, & Pisoni, 1994) and suprasegmental levels (e.g., Mandarin tones, as in Wang, Spence, Jongman, & Serena, 1999; Thai tones as, in Wayland & Guion, 2004). These results provide evidence that adults are able to form robust categories of new L2 sounds. The ability has been demonstrated through learners' performance (Bradlow et al., 1997) and change in areas of brain activation due to learning induced neural plasticity (Callan et al., 2003).

Even though Hardison (2000, 2003, 2005) carried out auditory and audiovisual training on videotape and Lambacher and colleagues (2005) carried out auditory training on cassette, both with positive results, most researchers have conducted computer-based perceptual training (e.g., Logan et al., 1991; Bradlow et al., 1997; Hazan et al., 2005; Pruitt, Jenkins, & Strange, 2006). In this type of training learners listen to stimuli and identify the sound as one of two sounds being trained. The stimuli usually consist of phonemes in isolated words produced by multiple talkers in multiple phonological contexts and word positions (High Variability Approach) and containing minimal pairs of the target sounds. If the learner accurately identifies the sound, s/he listens to the following trial; otherwise, the correct word blinks on the screen and the stimulus is repeated. Research has not been consistent on the number of training sessions, usually varying from 3 to 15 sessions, each one lasts around 30 minutes. The training program varies from 1 to 8 weeks.

Among variations of perceptual training procedures are (a) training of three-sound contrasts (e.g., /p/-/v/-/b/, as in Hazan et al., 2005); (b) explicit instruction on the articulation of the to-be-trained sounds (e.g., Pruitt et al., 2006); (c) adaptive training, where learners have to master a structure to proceed to a more difficult one (e.g., Jamieson & Morosan, 1986; McClelland, Fiez, & McCandliss, 2002); (d) use of synthetic stimuli (e.g., Rochet, 1995); (e) fading, where the feature differentiating the sounds is enhanced in the beginning of the training session and gradually fades away (e.g., Jamieson & Morosan, 1986); and (f) audiovisual training (e.g., Hardison, 2000, 2003, 2005).

One particular study carried out by Yeon (2004) is of great interest for this research because it dealt with syllabic structures, investigating effects of perceptual training on vowel epenthesis avoidance. The study focused on final alveolopalatals in Korean/English interphonology and tested transfer and three-month retention of perceptual training to production. She found immediate positive results in perception and improvement in production detected only in the three-month retention test. Yeon claimed that participants at lower levels of proficiency benefited more from the training than those of more advanced levels.

In line with the studies previously mentioned, the present study examined the role of perceptual training on epenthesis and voicing avoidance as well as transfer of training to untrained clusters and unfamiliar talkers.

3. Objectives, research questions, and hypotheses

This exploratory study aimed at investigating effects of identification perceptual training on the pronunciation of /s/-clusters by a BP learner of English. Based on the studies reviewed, the research question and hypotheses set up were the following.
Question: What are the effects of perceptual training on word-initial /s/-clusters?
Hypothesis 1: There will be improvement in perception and in production after the training.
Hypothesis 2: There will be retention of improvement five months after the posttest.
Hypothesis 3: There will be transfer of improvement to an unfamiliar talker.
Hypothesis 4: There will be transfer of improvement to untrained clusters.

4. Method

The participants of the study were two Brazilian learners of English at an intermediate level of general proficiency (300 hours of formal instruction), a twenty-nine-year-old man and a twenty-seven-year-old woman. The research consisted of four phases: pretest, training, posttest, and retention test. Only the female participant received the training. The time lag between pre and posttests was around fifteen days for the two participants. The talkers of the perception test were an American woman (T3) and an American man (T2). In the training material, the talkers were T3 and another American man (T1).

The production test and the perception test of the pretest phase were used in the posttests taken right after the training phase and five months later. Production was assessed by a phrase-reading test (see the Appendix). The participants were recorded reading the phrases used in the perception test. The /s/-voiced consonant clusters included were /sl, sm, sn/ and the /s/-voiceless consonant clusters included were /sk, sp, st, skr, spr/. Four preceding phonological environments were tested and trained - /aʊ/, as in 'how smiles'; /f/, as in 'if smiles'; /v/, as in 'move smiles'; and silence, as in 'smiles'. The total number of tokens per participant was 45. It took the participants less than five minutes to complete the production test.

The perception test applied was a forced-choice identification test similar to the task used in the training and the stimuli consisted of phrases containing three contrasts ([s-is, s-iz, s-z]). The /s/-clusters included were /sm, sl, sn, sk, sp, st, skr, spr/. As mentioned above, the stimuli were produced by two talkers (T3 and T2). The tokens from T2 were not used in the training sessions, but they were preserved for the investigation of Hypothesis 3. The total of tokens in the perception test was 300. It took the participants around 45 minutes to complete the perception test.

The training consisted of two-alternative-forced-choice (2AFC) identification trials with immediate feedback and replay allowed before hitting the decision key. The stimuli were phrases produced by T3 and T1. The training program was developed in six sessions with numbers of blocks varying from one to twenty-four. The number of trials in each block varied from sixteen to sixty-four. In the first three sessions, the level of complexity of the material presented increased gradually to ensure that participants could engage in the tasks smoothly. Thus, the number of talkers, and types of clusters, contrasts, and contexts were presented with increasing complexity beginning with clusters of /s/-stops before /s/-sonorants; with the context 'silence', then 'consonants', and then 'vowels'; with clusters presenting epenthesis, and then voicing; and, with material recorded by a single talker, and then by two talkers. Adapting Rochet (1995), in the first three sessions the participants were required to obtain 93% of accuracy to be allowed to advance to the following block, and in the other sessions they were required to reach 96% of accuracy. This difference in minimum accuracy requirements was due to the different number of trials in the blocks across sessions. At the end of each session, the participants had a repetition block featuring some of the phrases included in the training. The repetition block was included with the objective of triggering the participants' awareness regarding the relationship between perception and production, since manipulation of attention has proved to be an efficient strategy (Guion & Pederson, 2007). In the fourth session, contexts were presented in randomization. In the fifth session, contexts as well as contrasts were presented in randomization. Finally, in the sixth session, clusters, contrasts, and contexts were randomized.

The posttest happened immediately after the last training session and was identical to the pretest. The retention test happened five months after the posttest and was also identical to the pretest.

5. Procedure

The data was collected individually in a laptop computer. The first test administered was the production test and the second test administered was the perception test. Before each test, the participants received instructions in their native language. The production test lasted less than five minutes. The untrained participant (UP) took around 40 minutes to complete the perception pretest and posttest. The trained participant (TP) took 55 minutes to complete the perception pretest and 40 minutes to complete the perception post- and retention tests.

The production data was acoustically (in Praat) and perceptually analyzed. The results for production were tabulated by word produced. There were thirteen words per participant. The number of productions for each of the words containing /s/-sonorant clusters was four and for each of the words containing /s/-stop clusters was three because silence was not tested for this type of cluster. Four variables comprised the results for pre and posttests – accurate (/sC/), with prothesis (/isC/, /ɪsC/), voiced (/zC/), voiced and with prothesis (/izC/, /ɪzC/).

The results for perception were also tabulated by word perceived. There were thirteen words per participant. The number of tokens for each word containing /s/-sonorant clusters was thirty-six – eighteen for each talker – and for each of the words containing /s/-stop clusters was twelve – six for each talker because the only contrast tested was /s vs. is, ɪs/. The results were analyzed according to four items – total accurate identification (/sC/), accurate identification of prothesis (/isC/, /ɪsC/), accurate identification of voicing (/zC/), and accurate identification of voicing with prothesis (/izC/, /ɪzC/).

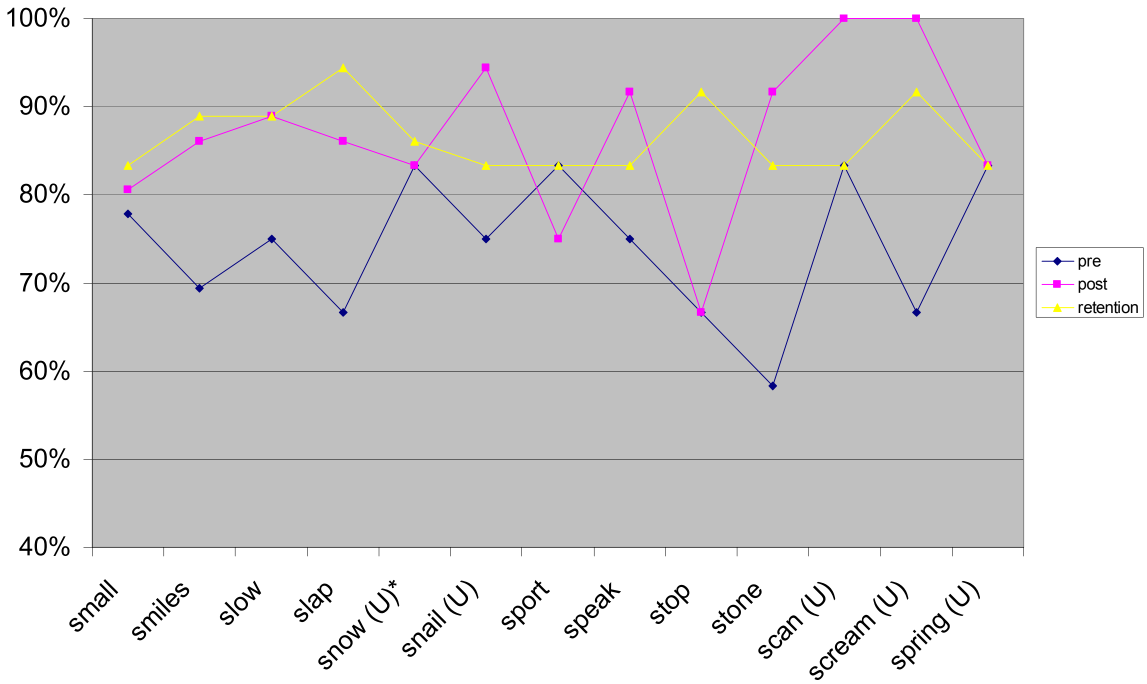
The training lasted around seven hours. In order for the participant to be able to reach the minimum accuracy required to advance to a following block, the number of times each block was used varied from one to 21 times, and the time spent to advance to a following block varied from one to 71 minutes. The participant took around 150 minutes to complete the first session and around 90 minutes to complete the second session. The first and second sessions differed only in the talker who produced the tokens. The length of each session varied according to the number of times necessary for reaching the level of accuracy required, and the number of times the participant listened to each trial before reaching the decision key. The difference in the time spent in the first and second sessions is explained by the facts that (a) the participant reached the minimum accuracy to advance faster in the second session probably because she had already been trained on the target contrasts, and (b) in the second session, the participant did not need to listen to each trial as many times as she needed when she was introduced to the contrast. The only exception was block 23 ('how slow' vs. 'how's low') which was not demanding in the first session – the participant was 100% accurate the first time she performed it – and took more than 40 minutes to reach the minimum accuracy required in the second session probably because the male talker produced a less salient contrast. The third session lasted around 40 minutes. The fourth and fifth sessions lasted around 30 minutes each. Finally, the sixth session lasted twelve minutes.

Statistical tests were run using SPSS. Non-parametric tests were used because the data was tabulated in percentages. The tests run were Wilcoxon Sign Ranked Test – for comparing means between two interval variables, Mann-Whitney Test – for comparing means with an interval and a nominal variable, and Kruskal-Wallis Test – for comparing means between two intervals and one nominal variable. Following Bradlow et al. (1997), measures of gain score and proportion of the room for improvement were calculated.

6. Results and discussion

This study was motivated by the observation that the production of word-initial /s/-clusters by BP learners of English may be problematic and mere awareness tends not to lead to better production. Also, some learners claim they cannot perceive the difference between erroneous and accurate word-initial /s/-cluster productions. Perceptual training has been successfully used regarding several non-native contrasts such as voicing and tones. However, /s/-clusters have not been the object of any of the studies on perceptual training published so far. To verify whether perceptual training may help improve pronunciation of word-initial /s/-clusters, four hypothesis were investigated.

Hypothesis 1 predicted that results from the perception and production pretests would be different from results from the perception and production posttests considering the trained participant. Figure 1 illustrates the results for pre and posttests considering perception. It also corroborates Hypothesis 2 concerning retention of improvement in perception.



* - (U) stands for untrained word.

Figure 1. Perception pretest and posttest scores for the trained participant by word

Table 1 shows the percentages of accurate identification by word and familiar (TT) and unfamiliar talkers (UT).

Table 1. Perception scores for the trained participant

Word		N				%		
			pre	pos	RET	pro	pos	RET
Small	**TT**	18	14	15	14	77.8	83.3	77,8
	UT	18	14	14	16	77.8	77.8	88,9
		36	**28**	**29**	**30**	**77.8**	**80.6**	**83,3**
smiles	**TT**	18	13	17	15	72,.2	94.4	83,3
	UT	18	12	14	17	66.7	77.8	94,4
		36	**25**	**31**	**32**	**69.4**	**86.1**	**88,9**
Slow	**TT**	18	13	15	18	72.2	83.3	100,0
	UT	18	14	17	14	77.8	94.4	77,8
		36	**27**	**32**	**32**	**75.0**	**88.9**	**88,9**
Slap	**TT**	18	12	15	17	66.7	83.3	94,4
	UT	18	12	16	17	66.7	88.9	94,4
		36	**24**	**31**	**34**	**66.7**	**86.1**	**94,4**
Snow (N)	**TT**	18	15	16	15	83.3	88.9	83,3
	UT	18	15	14	16	83.3	77.8	88,9
		36	**30**	**30**	**31**	**83.3**	**83.3**	**86,1**
Snail (N)	**TT**	18	12	16	17	66.7	88.9	94,4
	UT	18	15	18	13	83.3	100	72,2
		36	**27**	**34**	**30**	**75.0**	**94.4**	**83,3**
Sport	**TT**	6	4	4	5	66.7	66.7	83,3
	UT	6	6	5	5	100	83.3	83,3
		12	**10**	**9**	**10**	**83.3**	**75**	**83,3**
Speak	**TT**	6	3	6	5	50.0	100	83,3
	UT	6	6	5	5	100	83.3	83,3
		12	**9**	**11**	**10**	**75.0**	**91.7**	**83,3**
Stop	**TT**	6	4	4	5	66.7	66.7	83,3
	UT	6	4	4	6	66.7	66.7	100,0
		12	**8**	**8**	**11**	**66.7**	**66.7**	**91,7**
Stone	**TT**	6	4	5	5	66.7	83.3	83,3
	UT	6	3	6	5	50.0	100	83,3
		12	**7**	**11**	**10**	**58.3**	**91.7**	**83,3**
Scan (N)	**TT**	6	6	6	4	100	100	66,7
	UT	6	4	6	6	66.7	100	100,0
		12	**10**	**12**	**10**	**83.3**	**100**	**83,3**
Scream (N)	**TT**	6	3	6	4	50.0	100	66,7
	UT	6	5	6	5	83.3	100	83,3
		12	**8**	**12**	**9**	**66.7**	**100**	**91,7**
Spring (N)	**TT**	6	5	4	6	83.3	66.7	100,0
	UT	6	5	6	4	83.3	100	66,7
		12	**10**	**10**	**10**	**83.3**	**83.3**	**83,3**
Total		**300**	**223**	**252**	**259**	**74,4**	**84,0**	**86,3**

As Table 1 illustrates, the overall results show that there was 12.3% of gain score in the perception test. A measure of the proportion of the room for improvement showed that the 12.3%

of improvement corresponds to 48.1% of the total possibility of improvement (100% - pretest) which was of 25.7%. A Wilcoxon Signed Ranks test run on the percentages of pretest and posttest scores by word yielded a significant result (Z=-2.86, p = .004). Thus, H1 was confirmed considering perception. Figure 1 also illustrates that the second hypothesis, which claims that there would be retention of improvement after five months, was corroborated concerning perception. A Wilcoxon Signed Ranks test yielded a non-significant difference for accuracy in perception in the posttest and in the retention test (Z = .535, p>.05), indicating that improvement in perception was retained but no further improvement was detected after five months.

Table 2 shows the scores obtained by the trained participant for production in the pretest and in the posttest.

Table 2. Production scores for the trained participant

	N	/s/		%		/is/		%		/z/		%		/iz/		%	
		pre	Pos	pre	pos	pre	pos	pre	pos	pre	pos	pre	pos	pre	pos	Pre	pos
Small	4	0	1	**0**	**25**	0	3	**0**	**75**	1	0	**25**	**0**	3	0	**75**	**0**
Smiles	4	0	0	**0**	**0**	0	3	**0**	**75**	1	1	**25**	**25**	3	0	**75**	**0**
Slow	4	0	0	**0**	**0**	1	1	**25**	**25**	0	1	**0**	**25**	3	2	**75**	**50**
Slap	4	0	1	**0**	**25**	1	0	**25**	**0**	0	2	**0**	**50**	3	1	**75**	**25**
Snow (U)	4	0	0	**0**	**0**	0	0	**0**	**0**	0	2	**0**	**50**	4	2	**100**	**50**
Snail (U)	4	0	1	**0**	**25**	0	1	**0**	**25**	1	1	**25**	**25**	3	1	**75**	**25**
Sport	3	0	1	**0**	**33**	2	1	**67**	**33**	0	1	**0**	**33**	1	0	**33**	**0**
Speak	3	0	2	**0**	**67**	2	1	**67**	**33**	0	1	**0**	**33**	1	0	**33**	**0**
Stop	3	0	3	**0**	**67**	2	0	**67**	**0**	0	0	**0**	**0**	1	0	**33**	**0**
Stone	3	1	3	**33**	**100**	2	0	**67**	**0**	0	0	**0**	**0**	0	0	**0**	**0**
Scan (U)	3	0	2	**0**	**67**	3	1	**100**	**33**	0	0	**0**	**0**	0	0	**0**	**0**
Scream (U)	3	0	3	**0**	**100**	3	0	**100**	**0**	0	0	**0**	**0**	0	0	**0**	**0**
Spring (U)	3	0	2	**0**	**67**	3	1	**100**	**33**	0	0	**0**	**0**	0	0	**0**	**0**
TOTAL	**45**	**1**	**19**	**2.2**	**42**	**19**	**15**	**42**	**33**	**3**	**9**	**6.7**	**20**	**22**	**6**	**49**	**13**

The columns labeled with /s/ correspond to word-initial /s/-clusters accurately produced. As can be verified from raw scores and percentages by word, as well as from the total, the rate of misproductions was very high in the pretest with only one out of 45 tokens being accurately produced. In the posttest, there was nearly 40% of improvement. A Wilcoxon Signed Ranks Test yielded a significant difference between pre and posttests (Z=-2.84, p =.004). The proportion of the room for improvement was 41%. Therefore, H1 was also confirmed considering production. The improvement in production was retained after five months. Table 3 and Figure 2 also show that accuracy in production increased from the posttest to the retention test.

Table 3. Accuracy in production in pre, post, and retention tests

		N			%		
		pre	Pos	RET	pre	pos	RET
Small	4	0	1	2	0	25	50
Smiles	4	0	0	3	0	0	75
Slow	4	0	0	1	0	0	25
Slap	4	0	1	3	0	25	75
Snow (N)	4	0	0	4	0	0	100
Snail (N)	4	0	1	3	0	25	75
Sport	3	0	1	3	0	33	100
Speak	3	0	2	1	0	67	33
Stop	3	0	3	3	0	67	67
Stone	3	1	3	3	33	100	100
Scan (N)	3	0	2	2	0	67	67
Scream (N)	3	0	3	2	0	100	67
Spring (N)	3	0	2	1	0	67	33
TOTAL	**45**	**1**	**19**	**31**	**2.2**	**42**	**69**

Figure 2 illustrates the improvement in production from pretest to posttest as well as the improvement from posttest to the retention test.

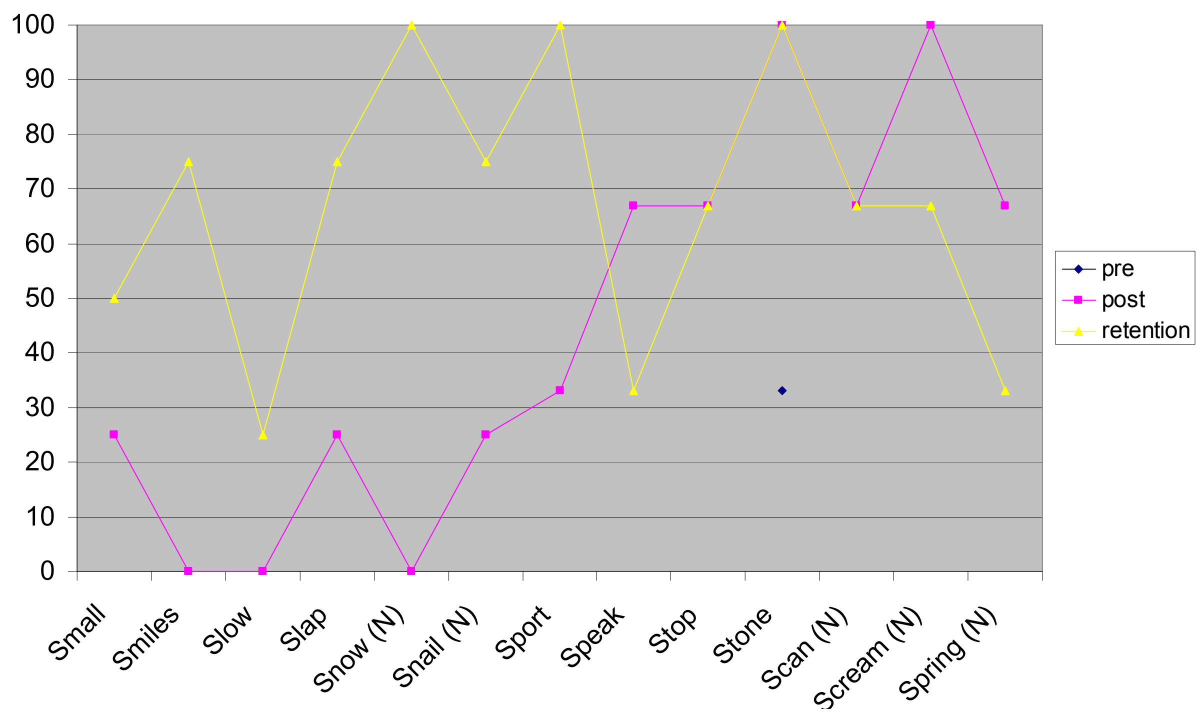


Figure 2. Improvement in production in different tests

Thus, the second hypothesis was confirmed for production as well. Moreover, a Wilcoxon Signed-Rank test showed that the improvement in production from posttest to retention test was statistically significant (Z= 2.692, p = .007). Several studies (e.g., Bradlow et al.,1997; Yeon, 2004) demonstrated that whereas improvement in perception reaches its peak at the posttest, production remains improving for some period probably because production self-monitoring strategies are triggered by perceptual training tasks.

The third hypothesis predicted that improvement in the perception test for the familiar talker would not differ from improvement for the unfamiliar talker. Table 1 shows the score for pre and post perception tests by talker. The results indicate that there was no difference between the gain scores of both talkers. In order to verify whether there is a difference between the talkers, a Mann-Whitney Test was run on gain scores for both talkers (Z = -.311, p=.756), and on the proportion of the room for improvement for both talkers (Z = -.367, p=.714). Both results support Hypothesis 3 in that they show that the difference between improvement for the familiar and unfamiliar talker is not statistically significant. There would hardly be a valid reason for training a FL learner to perceive the speech of one specific FL talker without expecting the improvement to be transferred to other talkers and consequently to real-life situations. However, since previous studies have shown that lack of variability in the stimuli can lead to lack of generalization, transfer to an unfamiliar talker remains a very important topic of investigation to attest training efficacy.

The fourth hypothesis concerned trained and untrained clusters and stated that improvement in the perception and production tests for the trained clusters would not differ from improvement for the untrained clusters, characterizing transfer of improvement.

In Table 1 – perception – and Table 2 – production – words containing the untrained clusters are marked with a capital U. By analyzing both tables it can be inferred that there is no significant difference between trained and untrained clusters either for perception or production. To assure that such assumption was correct, a Kruskal-Wallis Test was run for gain score in perception and Gain score in production with "cluster" as a grouping variable. The result yielded was a non-significant chi-square for production and for perception. Thus, the fourth hypothesis was confirmed in that trained and untrained clusters had similar improvement rates.

In order to verify whether training and not task effect or mere exposure to the target structures was responsible for the improvement, the results of pre and posttests for a non-trained participant (control) were analyzed. A Wilcoxon Signed Ranks Test yielded the following result for production: Z = -.36, p = .716. Even though in percentage there was a reduction in accuracy – from 51% to 42% – it was not significant, supporting the hypothesis concerning production. For perception, conversely, results from a Wilcoxon signed ranks test (Z = 3.090) were significant at .01 level. Thus, there was probably a task effect on the results concerning perception even though no carry over effect to production was found. Moreover, this analysis may also indicate that massive exposure without feedback is efficient to improve perception.

7. Conclusion

Hypothesis 1, which claimed that scores from pretest and posttest would be different for the trained participant, was confirmed, showing that training affected perception and improvement was generalized to production.

Hypothesis 2 was confirmed, showing that the improvement obtained from training was retained after five months. Moreover, it indicated that by improving perception, self-monitoring can lead to better production, which may result in continuous improvement for at least a few months after the training.

The confirmation of Hypothesis 3 indicated that there was transfer to an unfamiliar talker. This means that although the perceptual training was carried out with only two talkers there was transfer to a third talker and even though this does not guarantee good perception from word-initial /s/-clusters produced by other talkers, it at least indicates that the training carried out in the study provided enough variability to achieve generalization.

Hypothesis 4 was also corroborated and showed that there was transfer to perception and production of untrained words beginning in untrained two-member clusters and to the two three-member clusters tested (/spɾ, stɾ/).

Finally, the fact that the control had no improvement concerning production shows that it is very likely that training was actually the factor for the improvement detected in the trained participant's result for the production test. In order to verify if this was also true for perception, a different perception test should be administered. Another way of seeing the improvement in perception undergone by the untrained participant is that awareness can actually help to improve perception of word-initial /s/-clusters. Even though training with feedback is necessary to improve production, perception is improved even when no feedback is available through guided exposure alone.

This study was exploratory and that was one of the reasons for the emergence of several limitations. The main limitation concerns the number of participants. It is important to remember that the results observed here comprise an important contribution to the study of perceptual training; however, they cannot be generalized in that the study was carried out with only two participants. Moreover, the participants had very different starting points in relation to both perception and production of word-initial /s/-clusters. Therefore, the conclusions drawn here only apply for the two participants of the present study. Even though this was a drawback, it does not seem problematic to overcome by collecting data and training a larger sample.

The further limitation concerns the tests administered. Concerning production, besides the phrase reading test, a sentence reading test and a guided "free" speech task could be administered. For perception, besides the two-alternative forced choice identification test, a native-likeness identification test and a discrimination test could be used.

In sum, the results show that perceptual training is a good tool for enhancing the learning of word-initial /s/-clusters. These results along with those from the studies reviewed above indicate that identification and production of several non-native contrasts can be improved by the employment of perceptual training programs, which are easy to use and efficient as computer assisted language learning tools.

References

Baptista, B. O., & Silva Filho, J. L. A. (1997). The influence of markedness and syllable contact on the production of English consonants by EFL learners. In J. Leather, & A. James (Eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition Of Second Language Speech* (pp. 26-34). Klagenfurt, Austria.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Timonium: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Bettoni-Techio, M. (2005). *Production of final alveolar stops in Brazilian Portuguese/English interphonology*. Unpublished Master's thesis, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Bettoni-Techio, M., & Koerich, R. D. (2004). A Percepção de alveolar plosivas finais na interfonologia de brasileiros aprendizes de Inglês. Unpublished paper.

Bradlow, A. R., Pisoni, D. B.,Akahane-Yamada, R., & Tohkura,Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299 – 2310.

Callan, D., Tajima, K., Callan, A., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Nature NeuroImage, 19,*113-124.

Eckman, F. R. (1987). Markedness and the contrastive analysis hypothesis. In G. Ioup, & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 55-69). New York: Newbury House.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America, 84*(1)*,* 70-79.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *The Journal of the Acoustical Society of America*, *86*(5), 1684-1697.

Flege, J. E. (1995). Second Language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Timonium, MD: York Press.

Flege, J. E. (1997). The role of phonetic category formation in second-language speech learning. In J. Leather, & A. James (Eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second Language Speech* (pp. 79-89). Klagenfurt, Austria.

Hardison, D. (2000). Transfer of second-language perceptual training to production improvement: Focus on /r/ and /l/. In A. James, & J. Leather (Eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech* (pp. 166-173). Klagenfurt, Austria.

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

Hardison, D. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics, 26,* 579-596.

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.

Guion, S., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James Emil Flege* (pp. 57-77). Amsterdam: John Benjamins.

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

Kluge, D. (2004). *Perception and production of final nasals by Brazilian Portuguese learners of English*. Unpublished Master's thesis, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Koerich, R. D. (2002). *Perception and production of word-final vowel epenthesis by Brazilian EFL students*. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26,* 227-247.

Lively, S. E., Logan, J. S., & Pisoni, D. S. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

Lively, S. E., Logan, J. S., & Pisoni, D. S. (1994). Training Japanese listeners to identify English /r/ and /l/. III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076-2087.

Logan,J. S., Lively, S. E., & Pisoni, D. S. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America, 89,* 874-886.

Moore, D. (2004). *Perception and production of final /l/ by Brazilian Portuguese learners of English*. Unpublished paper, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Pennington, M. C. (1994). Recent Research in L2 phonology: Implications for Practice. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New ways, new directions* (pp. 92-107). Illinois: Pentagraph Print.

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. *Journal of the Acoustical Society of America, 119*(3), 1684-1696.

Rauber, A. S. (2006). Production of initial /s/-clusters by speakers of Brazilian Portuguese and Argentine Spanish. In B. O. Baptista, & M. A. Watkins (Eds.), *English with a latin beat: Studies in Portuguese/Spanish-English interphonology* (pp. 155-167). Amsterdam: John Benjamins.

Rebello, J. T., & Baptista, B. O. (2006). The influence of voicing on the production of initial /s/-clusters by Brazilian learners. In B. O. Baptista, & M. A. Watkins (Eds.), *English with a latin beat: Studies in Portuguese/Spanish-English interphonology* (pp. 1139-154). Amsterdam: John Benjamins.

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: York Press.

Silveira, R. (2002). Perception and production of English initial /s/clusters by Brazilian learners. *Revista Brasileira de Lingüística Aplicada, 2*(1), 95-119.

Silveira, R. (2004). *The influence of pronunciation instruction on the perception and production of English word-final consonants*. Unplublished doctoral dissertation, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Spolsky, B. (1989). *Conditions for Second Language Learning*. Oxford: Oxford University Press.

Trapp, N. L., & Bohn, O. S. (2000). Training Danish listeners to identify English word-final /s/ and /z/: Generalization of training and its effect on production accuracy. In A. James, & J. Leather (Eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech* (pp. 343-350). Klagenfurt, Austria.

Wang, X., & Munro, M. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539-552.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*, 681-712.

Winkler, I. Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi R.J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638-642.
Yeon, S. (2004). Training English word-final alveolopalatals to native speakers of Korean. Unpublished doctoral dissertation. University of Florida, Florida, United States.

Appendix: Phrases read by the participants

How smiles
If smiles
Move smiles
Smiles
How small
If small
Move small
Small
How slow
If slow
Move slow
Slow
How slap
If slap
Move slap
Slap
How snail
If snail
Move snail
Snail
How sport
Move sport
If sport
How stone
Move stone
If stone
How stop
Move stop
If stop
How speak
Move speak
If speak
How scan
Move scan
If scan
How scream
Move scream
If scream
How spring
Move spring
If spring
How snow
If snow
Move snow
Snow