Effects of Perceptual Training on the Learning of English Vowels in Non-native Settings

Denize Nobre-Oliveira

Federal Center of Technological Education of Santa Catarina, Brazil

denizenobre@yahoo.com.br

1. Introduction

Many of the studies carried out so far involving training approach a number of difficulties nonnative speakers of English may have depending on their L1 (Strange & Dittmann, 1984; Jamieson & Morosan, 1986; Logan & Pruitt, 1995; Yamada, Tohkura, Bradlow & Pisoni, 1996; Pisoni, Lively & Logan, 1994; Bradlow, Pisoni, Yamada & Tohkura, 1997; Ceñoz Iragui & Garcia Lecumberri, 1999; Hardison, 2002; Wang, 2002; Wang & Munro, 2004; Hawkey, Amitay & Moore, 2004; among others). For instance, a considerable number of studies on the effects of training have focused on the learning of the /l/-/r/ contrast by Japanese learners of English, others on the learning of nonnative vowel contrasts by native speakers of Spanish, Mandarin or Cantonese. Although there have been so many studies on this issue involving learners with various native-language backgrounds, no studies have focused on perceptual training involving Brazilian learners.

In this study I investigated the effect of perceptual training on the perception of three English vowel contrasts (/i-ɪ/, /ɛ-æ/, and /ʊ-u/) by Brazilian EFL students. I also investigated whether Brazilian learners would benefit or not from training involving synthesized speech stimuli with cue enhancement – a specific kind of training in which learners would be exposed to artificial stimuli with an emphasis in the crucial portions of the signal. Previous studies involving synthesized stimuli (Wang, 2002; Wang & Munro, 2004) have shown that vowel synthesis can be useful to enhance subtle but crucial L2 properties not usually perceived by a nonnative listener, and that this enhancement can facilitate perception. In this study, I investigated how the effects of training with artificial stimuli compared to the effects of training with natural stimuli for the purpose of perception improvement of the targets, that is, a better identification rate.

I also checked whether the knowledge acquired by means of synthesized stimuli was transferred to natural listening settings, that is, if training with synthesized stimuli led to improvement in listening to natural speech. Evidence of this kind of generalization was found by Jamieson and Morosan (1986), Yamada et al. (1996), Bradlow et al. (1997), Wang et al. (1999), Fox and Maeda (1999), Wang (2002), Hardison (2003), Wang and Munro (2004), Hazan et al. (2005), and Pruitt et al. (2006).

2. Method

2.1 Participants

Thirty-six undergraduate students of English at the Universidade Federal de Santa Catarina (UFSC) in Brazil participated in this study: 7 in the control group of Brazilian Portuguese native speakers, with no specific phonetic training, and 29 in the experimental group, which received perceptual training. The experimental group consisted of third- and fourth-semester students of the undergraduate English program and fifth-semester students of the undergraduate Executive Secretary program. There was also two control groups: one

consisting of native speakers of American English and one with native speakers of Brazilian Portuguese. The participants in both of the control groups were graduate or undergraduate students with no specific knowledge about phonetics.

All of the participants in the experimental group were enrolled in the "Pronunciation Lab" class, which is taught during the third semester of the undergraduate English program. All of the Brazilian participants reported to be intermediate speakers of English, more specifically of the General American English variety.

The experimental group was then divided into two sub-groups, according to the kind of training they would receive (with natural stimuli – *NatS group* – or with synthesized stimuli – *SynS group*). The participants were assigned to each sub-group according to their results in the perception pretest so that each group would be formed with equivalent pretest scores and the level of proficiency regarding perception of the target contrasts would be balanced in both groups.

2.2 Materials and procedures

2.2.1 Perception test

The perception test consisted of a forced-choice labeling task in which the participants had to identify the American English vowels within 108 CVC words produced by 8 native American English speakers (4 males and 4 females). The words included nine of the American English vowels (the targets /i, ɪ, ɛ, æ, ʊ, u/), totaling 72 words with the targets plus 36 extra-words with /ʌ, ɑ, ɔ/, which were not considered for analysis. The target vowels and the extra-vowels /ʌ, ɑ, ɔ/ appeared twice within the following contexts: /kVt/, /pVt/, /sVt/, /tVk/ and /tVt/.

Before the perception test started, the participants received instructions in English on the procedure they should follow: They would hear a word and, on the answer sheet, they should circle the word that contained the same vowel sound as in the word they had just heard. The possible answers corresponding to the vowels /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, ʊ, u/ on the answer sheet were, respectively, "sheep, ship, bed, bad, cut, hot, talk, foot, boot". The perception tests (pre- and post-) were administered in the Language Lab at UFSC. To perform the identification task appropriately, the participants used Sony headsets (H5-95). Also, there was a short break after every 18 words.

The pretest was given on the first day of class, and the posttest as soon as vowel training was over, in the fifth week of class. The results of the perception pretest and posttest were compared and statistical tests[1] were run in order to check whether there was any significant improvement in the participants' perception of the targets or not.

2.2.2 Training

Two different sets of stimuli were used, depending on the group learners were in. The stimuli used in the natural-stimuli-based training were recorded by 7 native speakers of American English[2] (3 males and 4 females) from different U.S. states: Kentucky, Massachusetts, Michigan and New York. The Americans recorded nine monophthongs /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, ʊ, u/ within the /bVt/ frame. The words were then edited and organized according to the design of each activity.

The stimuli used in the synthesized-stimuli-based training consisted of isolated vowels generated by a Praat script. The values for F1 and F2 are the means of the values found in

[1] Independent sample t-tests were used to compare means between groups, and paired t-tests were used to compare means within groups.

[2] The native speakers in this Section were borrowed from Rauber (2006) with permission.

Peterson and Barney (1952) and in Ohnishi (1991); F3 values were taken from Peterson and Barney (1952). The means of the two studies were used because Peterson and Barney (1952) is a classical study of a variety of AE accent, whereas Ohnishi (1991) is restricted to the Californian accent, but reports more updated formant values. These mean values were used as a basis to establish the enhanced values for the synthesized vowels used in the training stimuli, shown in Tables 1 and 2.

Table 1. Enhanced F1 and F2 values for the synthesized stimuli generated in Praat for males. F3 values are the same as in Peterson and Barney (1952)

			Males			
	F1	**F2**	**F3**	**Duration**	**F0 initial**	**F0 final**
/i/	253	2280	3310			
/ɪ/	429	1850	2550			
/ɛ/	502	1804	2480		150	
/æ/	662	1692	2410		or	
/ʌ/	594	1280	2390	500ms	180	80
/ɑ/	721	1083	2440		or	
/ɔ/	573	895	2410		200	
/ʊ/	454	1130	2240			
/u/	302	962	2240			

Table 2. Enhanced F1 and F2 values for the synthesized stimuli generated in Praat for females. F3 values are the same as in Peterson and Barney (1952)

			Females			
	F1	**F2**	**F3**	**Duration**	**F0 initial**	**F0 final**
/i/	315	2795	3010			
/ɪ/	475	2283	3070			
/ɛ/	591	2273	2990		350	
/æ/	839	1977	2850		or	
/ʌ/	703	1524	2780	500ms	380	180
/ɑ/	850	1191	2710		or	
/ɔ/	587	943	2710		400	
/ʊ/	483	1249	2680			
/u/	371	1001	2670			

In order to generate the vowel sounds produced by a male speaker and by a female speaker, pitch values were also manipulated. In addition, three different initial pitch values were used in order to prevent the participants from relying in the pitch, instead of in the spectral quality. Pitch variation was also good for minimizing boredom during the task. Thus, as shown in Table 1, initial pitch values were 200Hz, 180Hz and 150Hz, and the final pitch value was fixed in 80Hz[3] for male participants. For females (see Table 2), initial pitch values were 400Hz, 380Hz and 350Hz, and 180Hz for final pitch. All nine AE vowels had tokens with the three initial pitches.

The duration of the vowels was kept constant at 500ms in order to train learners to rely primarily on the spectral properties of the vowels, rather than on duration. Americans rely primarily on spectral quality cues to discriminate the vowels in their L1 (Fox & Maeda,

[3] 80Hz was the standard value in the Praat script.

1999), even though AE vowels differ both in duration and in spectral quality. Furthermore, although previous research found no evidence that Brazilian advanced speakers of English always use L1 cues to identify L2 vowels (Rauber, 2006), speech perception models suggest that perception is a language-specific property and the learners' L1 will *guide* L2 speech perception.

As for the difference in linguistic context of the targets between groups (vowel within a word for the NatS group and isolated vowel for the SynS group), research has found that consonants may interfere in vowel duration, which would either facilitate or complicate vowel identification (Garcia Lecumberri & Ceñoz Iragui, 1997). Peterson and Lehiste (1960) found that although preceding consonants do not affect the duration of vowels, vowel duration is significantly affected by the following consonants. The extent to which vowel duration is affected depends on the voicing and the natural class of the following consonant. Thus, voiced following consonants make vowels longer than their voiceless counterparts. Considering their natural classes, voiced fricatives lengthen the syllable nucleus most, followed by nasals and voiceless plosives. The consonants that affect vowel duration the least are the voiceless plosives.

Since the lack of a consonantal context is also a way to enhance the stimuli, isolated vowels were selected for the SynS group training stimuli. Conversely, it was decided to use words for the natural stimuli tokens because naturally produced vowels extracted from words are too short and, thus, more difficult to perceive and potentially inappropriate for training materials.

Moreover, although 500ms is a rather long duration for vowels, it was decided to keep it for the synthesized vowels in order to facilitate the perception of the different spectral properties of each vowel during the training phase, which would hopefully help learners to improve their ability of categorizing L2 vowels successfully.

Training was divided into two phases: theoretical and practical. In the first phase, which took 40 minutes, the learners were introduced to some basic articulatory vowel properties, such as vowel height and backness, and the relation between vowel articulation and their representation in vowel charts. In the second phase of the in class training, which took 50 minutes, the learners went to the Language Lab, where they performed listening activities in which they listened to specific stimuli according to the group they were in (natural words with the target vowels or just synthesized vowels) – and had to choose one option in their answer sheet. In the first week of training they practiced only front vowels, in the second week they practiced back vowels, and the third week of training was dedicated to practicing all vowels together. All activities performed in lab (the second phase of training) were divided into two blocks ("Part 1" and "Part 2") and they consisted of identification tasks. In the first block, the learners were asked to listen to a vowel and say whether it was vowel "X" or not. For instance, when they were practicing the front vowels, they had to say if the vowel they heard was /i/ or another vowel. In the second block, learners had two vowel choices. For instance, they heard the vowel /i/ and they had to say if it was an /i/ or an /ɪ/. The teacher provided immediate feedback after each trial.

3. Results and discussion

Comparing the results of the pretest with those of the posttest in each group, statistical tests confirmed that there was a significant difference in the performance of the participants in the experimental group. This difference indicates that they performed much better after training. The control group, however, did not show any significant difference in its performance from pretest to posttest, although a slight improvement was found. These results (Figure 1), suggest that training indeed has positive effects on the improvement of the perception of L2 sounds.

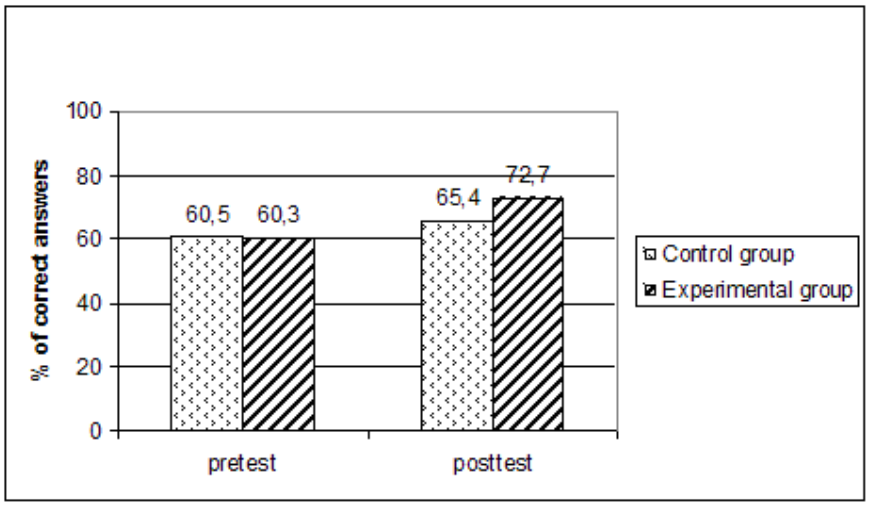

Figure 1. Overall pretest and posttest results per group

Significant improvement in the production domain was also found for the high front vowel contrast (/i-ɪ/) in the experimental group. An improvement tendency for the other two contrasts (/ɛ-æ/ and /u-ʊ/) was also found although this improvement did not reach the level of significance (Figure 2).

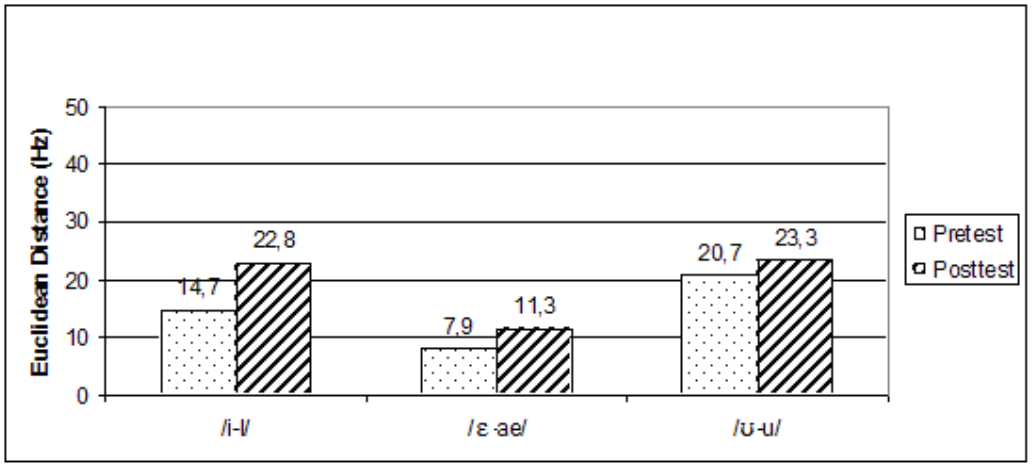

Figure 2. Production improvement for each vowel contrast

Comparing the pretests of the three goups, a Kruskal-Wallis test confirmed that there was no significant difference between groups before training (X^2 = -.812, p > .05). Similarly, no significant difference was found between groups after training (X^2 = -.114, p > .05). Still, Figure 3 shows an improvement tendency from pretest to posttest for the three groups, but the highest improvement rate was achieved by the SynS group (14 percentage points), followed by the NatS group (10.5 p.p.) and the Control group (5 p.p.).

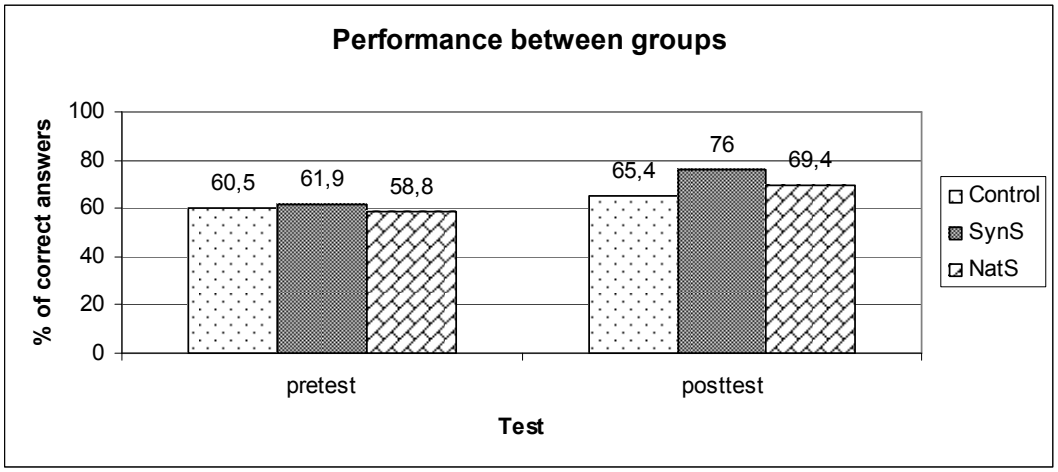

Figure 3. Comparison of overall pretest and of posttest results of the three groups (between-group performance)

Generalization effects were also assessed by comparing the performance of the SynS group in the pretest and posttest, as shown in Figure 3. Since the participants in this group were trained exclusively with synthesized stimuli consisting of isolated vowels, and were tested with natural stimuli in the CVC frame, an eventual perceptual improvement in the posttest would constitute evidence for the transfer of the knowledge acquired through synthesized stimuli to natural speech settings and to a new syllable structure. This hypothesis was confirmed by means of a Wilcoxon test, which showed that there was a significant difference from pretest to posttest.

Secondary evidence of generalization was found in the NatS group. Although the participants in this group were trained and tested with natural stimuli, most of the native speakers whose utterances were recorded for the test stimuli were different from the native speakers who recorded the training stimuli. Thus, an improvement from the pretest to the posttest of the participant in the NatS group would indicate that they generalized the new knowledge to new speakers. Figure 3 also shows that there was a significant improvement from pretest to posttest, confirming the generalization predictions to the NatS group.

The results presented here showed that both training methods (training with synthesized stimuli and training with natural stimuli) led to perceptual improvement. Based on the findings of previous research, it was hypothesized that learners trained with enhanced stimuli would benefit more than learners trained with natural stimuli in terms of perceptual improvement. Although the difference in improvement rate was not significant between training groups, the SynS group performance in the posttest was slightly better than the performance of the NatS group.

4. Conclusion

Several studies involving perceptual training as a means to help learners from different parts of the world were reported and corroborated the positive effects of this kind of training. The present study supports the results of previous research: after training, there was significant improvement in the learners' perception and the new knowledge was transferred to new talkers (for the NatS group) and to a new kind of stimuli in terms of nature and syllable structure (for the SynS group).

The results of the SynS group suggest that training with enhanced stimuli is more effective than training with natural stimuli, since the rate of improvement in both skills

(perception and production) was higher for the SynS group. Much of the effectiveness of the synthesized training in helping learners to identify L2 sounds is that subtle and crucial cues of the signal are enhanced, drawing learners' attention to them (and the less important features attenuated). Thus, the results of the present study suggest that enhanced stimuli help learners to develop selective attention to the crucial phonetic cues of certain sounds in a given L2, although more research on this issue is needed.

Perceptual training should also be implemented in EFL classrooms. Nowadays, in the Brazilian context, pronunciation courses foreground training the oral modality, giving less importance to listening practice. As pointed out by Rochet (1995:396), perception training "is easier to administer," since it can be done without the presence of the teacher (it can be done at home), and learners can be provided with immediate feedback.

References

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & 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*, 2299-2310.

Ceñoz Iragui, J., & Garcia Lecumberri, L. (1999). The effect of training on the discrimination of English vowels. *International Review of Applied Linguistics in Language Teaching*, *37*, 261-275.

Fox, M. M., & Maeda, K. (1999). Categorization of American English vowels by Japanese speakers. *Proceedings of the International Conference of Phonetic Sciences* (pp. 1437-1440). San Francisco.

Garcia Lecumberri, L., & Ceñoz Iragui, J. (1997). Identification by L2 learners of English vowels in different phonetic contexts. In J. Leather, & A. James (Eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second Language Speech* (pp. 196-205). Klagenfurt: University of Klagenfurt.

Hardison, D. (2002). 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: University of Klagenfurt.

Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*, 1055-1056.

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

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

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: Issues in cross-language research* (pp. 351-377). Timonium, MD: York Press.

Ohnishi, M. (1991). *A spectrographic investigation of the vowels of Californian English (Southwest General American)*. Paper presented in the 1991 Convention of the Phonetic Society of Japan. Osaka, Japan.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

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*, 1684-1696.

Rauber, A. S. (2006). *Perception and production of English vowels by Brazilian EFL speakers*. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Rochet, B. (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.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131-145.

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production*. Unpublished doctoral dissertation. Simon Fraser University.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *34*, 539-552.

Yamada, R. A., Tohkura, Y., Bradlow, A. R., & Pisoni, D. B. (1996). Does training in speech perception modify speech production? *Proceedings of the fourth International Conference on Spoken Language Processing,* v. 2 (pp. 606-609). Wyndham Franklin Plaza Hotel, Pennsylvania. Philadelphia: Institute of Electrical and Electronics Engineers.