The Effects of Perceptual Training on the Production of English vowel contrasts by Portuguese Learners

Anabela Rato and Andreia Rauber
University of Minho, Portugal; University of Tübingen, Germany
asrato@gmail.com; asrauber@gmail.com

ABSTRACT

This study investigated the short- and long-term effects of a high variability perceptual training on the production of three English vowel contrasts (/i/-/ɪ/; /ɛ/-/æ/; /u/-/ʊ/) by Portuguese learners. The perceptual training consisted of five sessions divided into two blocks that included both discrimination and identification tasks followed by immediate feedback with natural stimuli produced by multiple native talkers in different phonetic contexts. Vowel production data were collected three times, namely before training (pretest), immediately after (posttest), and two months later (delayed posttest) with a sentence-reading task. Both duration and the first two vowel formants were measured to calculate the Euclidian distance (Hz) and duration ratios (ms) between the vowels of the target contrasts. The acoustic analyses revealed that perceptual training had a significant effect on pronunciation accuracy of the target vowels, specifically in terms of vowel quality.

Keywords: Perceptual training, Transfer of training, L2 Vowel production, Vowel contrasts

1. INTRODUCTION

Several studies have demonstrated that second/foreign language (L2) speech learning might pose a challenge to adult learners, particularly in the perception and production of certain non-native contrasts [9]. The *Speech Learning Model* (SLM), proposed by Flege [6], predicts that difficulties in perceiving and, consequently, in producing non-native contrasts are due to the (dis)similarities of sounds of the L1 and the L2 . However, a substantial number of cross-language studies have shown that L2 speech learning is attainable for late learners, and their abilities to perceive and produce both segmental and suprasegmental non-native contrasts can improve, because the cognitive mechanisms used in the acquisition of the L1 sound system remain intact over the lifespan and are triggered by L2 speech experience [6]. Experimental studies that investigated the effects of perceptual training on non-native speech sound perception and production reported its success not only in the modification of adult learners' perceptual patterns, but also in the improvement of their pronunciation accuracy, thus confirming the plasticity of L2 learners' mature perceptual system (e.g., [1], [11], [17], [18]).

Non-native vowel contrasts have been widely described as a considerable part of the problems learners have in L2 speech learning [16]. To our knowledge, no study has yet examined L2 vowel production with European Portuguese learners. Furthermore, findings seem to be rather inconsistent regarding training effects on production accuracy, with studies reporting either positive effects (e.g., [1], [7], [8], [11]) or no effect (e.g., [2], [10], [17]). Moreover, few studies have investigated transfer of learning to production, in particular by means of quantitative analyses of L1-L2 acoustic similarity, i.e., acoustic measurements of duration and formant frequency.

Therefore, the present study[1] investigated the effectiveness of perceptual training in the learning of three English contrasts (/i /-/ ɪ /; /ɛ/-/æ/; /u/-/ʊ/) by a group of Portuguese learners of EFL (English as a Foreign Language). Specifically, it examined perceptual learning effects on production by means of acoustic analysis.

The American English vowel inventory includes four monophthongs (/i/, /ɪ/, /ɛ/, /æ/) and one diphthongized (/e/) in the front vowel space, whereas the European Portuguese (EP) vowel system consists of three front vowels /(/i/, /e/, /ɛ/) that differ in spectral quality and have intrinsic vowel duration differences [5]. In the high back space, EP has one high back vowel (/u/) and AmE has two vowels (/u/ and /ʊ/) that differ both in quality and length. The set of English vowel contrasts, /i /-/ ɪ /; /ɛ/-/æ/; /u/-/ʊ/, which differ both in terms of spectral quality and duration, was selected because they are reported to present production and perception difficulties for adult native EP speakers [6], [12]. English /ɪ/, /æ/ and /ʊ/ tend to be assimilated to the Portuguese vowel sounds /i/, /ɛ/ and /u/, respectively, and no distinction between the two vowels of each pair is made due to their acoustic and articulatory proximity [12].

2. METHOD

2.1. Participants

Thirty-four Portuguese undergraduate students (18F and 16M; mean age=23 years, SD=6.8) participated. Twenty-two of the participants were assigned to an experimental group and 12 to a control group. The groups' performance on a perception pretest was equivalent. Both groups participated in the perceptual training program to control for the effect of task repetition on degree of improvement. The trainees underwent a high variability phonetic training (HVPT) on vowel contrasts and the controls received training on consonants. Both groups of adult EFL learners had started to learn English at the age of 10 years (mean=9.7 years, SD=0.9) and had been exposed to English mostly through formal instruction for a mean of 8.2 years (SD=1.7). The two groups of participants suffered attrition from the posttest to the delayed posttest, with four dropouts.

Seven native speakers (NS) of American English (mean age=39.7 years, SD=11.3) participated as a baseline group. The overall performance of the NSs ranged from 96.6% to 100% accuracy in all perceptual tasks.

All participants reported having no hearing impairments or speech disorders.

2.2. Materials

2.2.1. Test materials

The productions of the American English vowels by the 34 participants were recorded three times by means of a sentence-reading-aloud task: before (pretest), immediately after (posttest), and two months after training (delayed posttest). The American English corpus consisted of 63 monosyllabic CVC words. The six target vowels and a distractor vowel (/ʌ/) were embedded in nine phonetic contexts and flanked by voiceless stops and fricatives to minimize duration variability. The target words were preceded by a picture depicting a high frequent CVC word (prime) that rhymed with the (target) word to be read in the carrier sentence: "Say (CVC word) now".

Each participant read the 63 sentences three times, but only one production for each context was considered for analysis (7 vowels x 9 contexts), yielding 63 vowel tokens per participant, and totaling 6174 vowels that were acoustically analyzed.

2.2.2. Training materials

Two similar five-session perceptual training programs with identification and discrimination tasks were designed for both groups. The training tasks included both immediate and cumulative feedback. The training program of the experimental group was divided into two blocks according to degree of task complexity. The first three sessions consisted of AX categorial discrimination tasks and two-alternative-forced-choice (2AFC) identification tasks, each focusing on one of the three target vowel contrasts. The two final training sessions included oddity discrimination tasks and 7AFC labeling tasks with the whole set of target vowels and a distractor (/ʌ/). In the three AX categorial discrimination tasks participants had to indicate whether two stimuli in randomized token pairs belonged to the same category or not, and in the two oddity tasks learners heard three different stimuli produced by three different talkers in each trial and identified the position of the token that represented a different vowel category (i.e., the odd token). In each triad, either the odd stimulus was presented in one of three positions (change trial) or the tokens were all drawn from one category (catch trial). The AFC identification tasks consisted of labeling a given vowel segment from two or more response options. The training stimuli comprised the target vowels inserted in CVC words in the four contexts /bVt/, /tVk/, /sVt/, /hVd/ produced by 12 American English NSs.

The training program of the controls was focused on two sets of English consonants that are challenging to Portuguese learners, namely the dental fricatives (/θ/ and /ð/) and the nasals (/m/, /n/, /ŋ/). The five sessions, divided into two blocks, included AFC identification and discrimination tasks, and natural spoken stimuli were produced by seven native American English talkers.

2.3. Procedures

2.3.1. Production test

The recordings of the Portuguese participants' productions were conducted individually in a sound-attenuated booth with an Edirol R-09HR digital recorder at a 44-HZ sampling rate, with 16-bit accuracy, and a unidirectional Edirol CS-15 microphone.

The sentence-reading task was set up in a custom-designed computer program using CSharp, which automatically randomized the presentation of the stimuli. Prior to testing, participants were familiarized with the task, and were asked to read

the sentences at a normal speech rate, with a falling intonation, and pause between each set of 63 sentences.

2.3.2. Perceptual training

Both groups of participants undertook five 45-minute sessions that followed the same sequence: (1) articulatory-visual description of the target segments; (2) instructions for each task; (3) discrimination task; and (4) identification task. Although brief articulatory information was provided, production of the target vowels was not encouraged.

The training tasks were administered in TP software [14] running simultaneously in several computers in a quiet computer room. Participants trained individually, and heard the stimuli (up to three times each) at a comfortable listening level over headphones.

2.4. Analysis

Vowel production data were acoustically analyzed by measuring duration and the first two formants with Praat 5.3.39 [3]. The waveform and the wideband spectrogram of the production data were visualized and an annotation text was added for each audio file to segment and label the target vowels. To manually segment the vowels, boundaries were set at zero crossings where the first positive and the last negative periodic pulses with considerable amplitude could be seen, thus marking the start and endpoints of each target vowel. After segmentation, the first two formants were automatically measured at the central 40% of the target vowels. Vowel formant data were normalized with NORM by selecting the Lobanov method [15]. After normalizing the formant values, the median F1 and F2 values of each vowel were calculated, as well as the Euclidean distance (ED), i.e., the space in Hertz (Hz) between the vowels of the target pairs in the F1/F2 space. This procedure allowed the comparison between the EDs of the vowel contrasts produced by the Portuguese and the American English speakers. The duration ratios of the three target vowel pairs were calculated by dividing the mean duration values (ms) of the longer vowels (/i/, /æ/, /u/) by the mean duration of the shorter vowels (/ɪ/, /ɛ/, /ʊ/), following the same procedure as Wang [17]. Taking into account that the higher the values of the ratios, the greater the differences between the vowels of a pair, this measure was expected to reveal duration differences between the vowels of the three contrasts.

To compare the median values of sets of vowels and verify whether they differed statistically according to testing time, parametric inferential tests of significance were used. The extent to which the Portuguese learners' vowel productions differed from the American English reference values was estimated by comparing the differences between the EDs of the participants and the AE NSs.

3. RESULTS

Before training, the formant values for the vowels /ɛ/ and /æ/ produced by the trainee and control groups overlapped, and the two high vowel contrasts overlapped partially. No between-group differences were found in the production of the two high vowel pairs, but the ED of /u/-/ʊ/ was significantly larger (t=2.09(32), p<.05) when produced by trainees than by controls. In terms of duration ratios, no differences were found between groups.

When comparing participants' EDs and duration ratios to native AmE reference values, we observed that the EDs of two of the target vowel contrasts produced by Portuguese speakers were significantly smaller than those of AmE speakers (F(df)=25.68 (2,40), p<.001, for /i/-/ɪ/; F(df)=19.42 (2,40), p<.001, for /u/- /ʊ/). The ED between /æ/ and /ɛ/, though not significantly different, was also smaller when produced by both groups of L2 learners (54.71 Hz and 59.38 Hz) than by AmE NSs (109.19 Hz). Likewise, the duration ratios were also significantly smaller for vowel contrasts /i/-/ɪ/ and /ɛ/-/æ/ in comparison to NSs (F(df)=220 (2,40), p<.001 /i/-/ɪ/; F(df)=46.01 (2,40), p<.001, respectively), which indicated that a native-like durational distinction between vowels in each pair was not made. Note that, because participant assignment to the trainee and control groups was based on pretest perception results, not their production performance, the pretest measures of ED and duration were not controlled between participants, which explains the significant intergroup difference in the articulation of /u/-/ʊ/ at pretest. However, in the pre-training vowel spaces of both groups vowels were similarly distributed in terms of height and frontness/backness (see Fig. 1). Moreover, both ED and duration values indicate that almost no spectral or durational distinction was made between vowels of each target contrast by the 34 EFL learners. In sum, the non-native vowels /ɪ/, /æ/ and /ʊ/ were produced as /i/, /ɛ/ and /u/, respectively, and the vowels /æ/ and /ɛ/ were produced with median F1 and F2 values closer to AmE /ɛ/.

Immediately after training, the EDs of the target vowel contrasts (/æ/-/ɛ/, /i/-/ɪ/, and /u/-/ʊ/) produced by the experimental group increased significantly; thus, vowels /i/-/ɪ/ were no longer overlapped, /u/-/ʊ/

overlapped only slightly, and /æ/-/ɛ/ partially overlapped (see Fig. 2).

Figure 1: Vowel space of the trainees (black) and the controls (grey) at pretest.

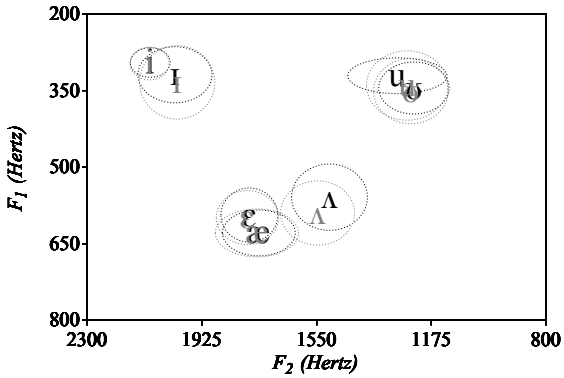
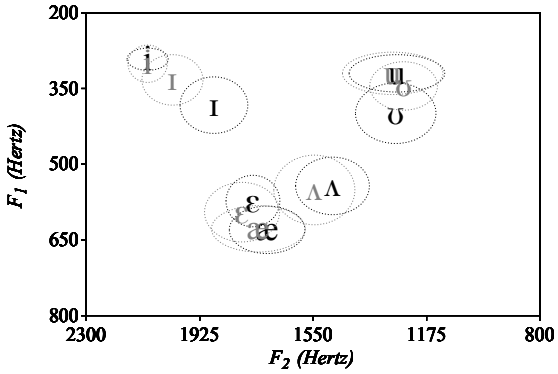


Figure 2: Vowel space of the trainees (black) and the controls (grey) at posttest.



The controls also produced the /u/-/ʊ/ contrast with a higher ED in relation to pretest, but increase was not as large (21.38 Hz) as for the trainees (49.76 Hz). The vowels of the /i/-/ɪ/ pair were closer at posttest (90.49 Hz) than at pretest (100.84 Hz), but the difference was not significant. Both groups differed significantly in the production of /i/-/ɪ/ (t=4.29(32), p<.001) and /u/-/ʊ/ (t=2.42(32), p<.05).

Two months after training, the two groups continued to differ significantly in the pronunciation of the same vowel pairs. The EDs of /i/-/ɪ/ (t=3.95(28), p<.001) and /u/-/ʊ/ (t=3.45(28), p<.01) were significantly larger in the productions by the trainee group than by the control group. Despite not being a significant difference, the ED of /æ/-/ɛ/ was somewhat larger in the trainees' productions.

The results of the delayed posttest showed that the experimental participants achieved a near-native like production in terms of vowel quality given that the only vowel contrast ED that still differed significantly from NSs was the /i/-/ɪ/ pair (F(df)=22.09 (2,36), p<.001). However, the distance between the vowels of each pair was much higher in the vowel trainees than in the control group (260 Hz and 114 Hz, respectively).

In terms of duration ratios and in comparison to NSs, EFL learners did not distinguish vowels by duration. Overall, the duration ratios decreased immediately after training, and in the particular case of the high vowel pair /i/-/ɪ/ duration ratio was significantly lower (t=2.99(21), p<.01) in the productions of the vowel-trainees. At pretest, the durational differences between vowels of the pairs /æ/-/ɛ/, /i/-/ɪ/, and /u/-/ʊ/ were significant, but immediately after training and two months later the durational distinction between the vowels of the high vowel contrasts was not significant, which indicates that EFL learners produced vowels of each pair with similar durational values.

4. DISCUSSION

The present study examined the short and long-term effects of a high variability perceptual training on the production of three English contrasts.

The results showed that after training, the trainees achieved a near-native-like acoustic distance (ED) between the vowels of the target contrasts. However, no positive effect of training was found on vowel duration. To some extent, this corroborates Bohn's [4] desensitization hypothesis, because at pretest it seemed that spectral differences were insufficient to distinguish vowel contrasts, i.e., learners seemed not to be sensitized to rely on quality; thus, duration differences were used to differentiate the English vowel contrasts. However, after training spectral differences seemed to override vowel length. As training progressed, awareness of spectral vowel dissimilarities went hand in hand with a decrease of durational differences. This finding seems to indicate that training had a reverse effect on vowel length ratios, i.e., by redirecting learners' attention to spectral differences, durational distinctions became less evident. At the onset of training, L2 learners seemed to rely more on duration than at its offset. In conclusion, perceptual training had a significant effect on pronunciation accuracy of the target vowels particularly in terms of vowel quality. Perceptual training seemed to have been effective in raising awareness to L1-L2 vowel dissimilarity and, consequently, new phonemic categories seemed to have been established for the non-native sounds.

In conclusion, these results support the claim that high-variability perceptual training may be effective in promoting the accurate production of non-native vowels in the foreign language classroom within a short period of time.

5. REFERENCES

[1] Aliaga-Garcia, C. 2013. *Effects of identification and articulatory training methods on L2 vowel production*. Paper presented at New Sounds 2013, 17-19 May, Montreal, Canada.

[2] Aliaga-Garcia, C., & Mora, J. C. (2007). Assessing the effects of phonetic training on L2 sound perception and production. In A. Rauber, M. Watkins, B. Baptista (eds.), *Proc. 5th Int. Symposium on the Acquisition of Second Language Speech*. Florianópolis, 25-28.

[3] Boersma, P., & Weenink, D. 2013. *Praat: doing phonetics by computer* (Version 5.3.39) [Software].

[4] Bohn, O.-S. 1995. Cross-language Speech Perception in Adults: First Language Transfer Doesn't Tell It All. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* Timonium. MD: York Press, 275-300.

[5] Escudero, P., Boersma, P., Rauber, A., Bion, R. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *J. Acoust. Soc. Am.* 126(3), 1379-1393.

[6] Flege, J. 1995. Second Language Speech Learning: Theory, Findings and Problems. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press, 233-277.

[7] Lacabex, E. G., Lecumberri, M. 2010. Investigating training effects in the production of English weak forms by Spanish learners. In K. Dziubalska-Kołaczyk, M. Wrembel, M. Kul (Eds), *Proc. 6th Int. Symposium on the Acquisition of Second Language Speech* Poznan, 137-143.

[8] Lengeris, A. 2008. The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. *Proc. 2nd ISCA Workshop on Experimental Linguistics* Athens, 133-136.

[9] Moyer, A. 2013. *Foreign Accent: The Phenomenon of Non-Native Speech*. Cambridge: CUP.

[10] Nobre-Oliveira, D. 2007. *The Effect of Perceptual Training on the Learning of English Vowels by Brazilian Portuguese Speakers* (Doctoral dissertation). Florianópolis: Univ. Federal de Santa Catarina.

[11] Pereira, Y., Hazan, V. 2013. *Impact of different training modes on the perception and production of English vowels by L2 learners*. Paper presented at New Sounds 2013, 17-19 May, Montreal, Canada.

[12] Rato, A., Rauber, A., Soares, L., Lucas, L. 2014. Challenges in the perception and production of English front vowels by native speakers of European Portuguese. *Diacrítica - série ciências da linguagem* 28(1), 137-155.

[13] Rauber, A. S. (2010). *Acoustic characteristics of Brazilian English vowels: perception and production results*. Saarbrücken: Lambert Academic Publishing.

[14] Rauber, A., Rato, A., Kluge, D., & Santos, G. (2012). TP-S (Version 1.0). [Software]. Brazil: Worken.

[15] Tyler, K., & Thomas, E. R. (2010). Vowels: Vowel Manipulation, Normalization, and Plotting in R. R package, (Version 1.1.) [Software]. Retrieved from http://ncslaap.lib.ncsu.edu/tools/norm/

[16] Strange, W. 2007. Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In: Bohn, O.-S., Munro, M. (eds), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege.* Amsterdam/Philadelphia: John Benjamins, 35-55.

[17] Wang, X. 2008. *Perceptual Training for Learning English Vowels – Perception, Production, and Long-Term Retention*. Saarbrücken: VDM Verlag.

[18] Wang, Y., Jongman, A., Sereno, J. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *J. Acoust. Soc. Am.* 113(2), 1033-1043.

[1] This paper is part of a larger study that also investigated the training effects on the identification of English vowels and generalization to new stimuli produced by novel talkers.