Computer-based auditory training improves second-language vowel production in spontaneous speech

Angelos Lengeris
Department of English Language and Linguistics, University of Kent, Canterbury, Kent, CT2 7NF, United Kingdom
a.lengeris@kent.ac.uk

Abstract: The current study examined the effectiveness of computer-based auditory training on Greek speakers' production of English vowels in read sentences and in spontaneous speech. Another group of Greek speakers served as controls. Improvement was evaluated pre- and post-training via an identification task performed by English listeners and by an acoustic analysis of vowel quality using a combined $F1/F2$ measure. Auditory training improved English vowel production in read sentences and in spontaneous speech for the trained group, with improvement being larger in read sentences. The results indicate that auditory training can have ecological validity since it enhances learners' production beyond the (read) sentence level.

[MG]
Date Received: April 29, 2018 **Date Accepted:** August 13, 2018

1. Introduction

Adult learners have difficulties when acquiring the sounds of a second language (L2). Such difficulties are attributed, among other variables, to the relationship between the learners' native (L1) and the L2 sound inventory, as acknowledged by a number of theoretical models proposed over the years. The Perception Assimilation Model (Best, 1995; Best and Tyler, 2007) for example, describes in detail a mechanism whereby L2 contrasts are assimilated into similar L1 categories, which in turn leads to L2 perception difficulties. This can be demonstrated by the difficulty Greek learners have with the English tense-lax /iː/-/ɪ/ distinction (e.g., *beat* vs *bit*), the reason being both English vowels being assimilated into Greek /i/ (Lengeris, 2009).

Despite such difficulties, L2 sound perception can be trained via computer-based instruction. The most successful training paradigm emphasizes the use of highly variable, naturally produced materials that contrast the target sounds in multiple phonetic environments (e.g., Logan *et al.*, 1991). During the so-called high-variability phonetic training, learners receive identification training with immediate feedback, using recordings of multiple minimal pairs from multiple speakers. High-variability training studies report significant improvement in the perception of English vowels by native speakers of Japanese (Lambacher *et al.*, 2005), French (Iverson *et al.*, 2012), Mandarin (Thomson, 2011), Spanish and German (Iverson and Evans, 2009), and Greek (Lengeris and Hazan, 2010).

High-variability phonetic training can also lead to production improvement without the trainees receiving any explicit pronunciation instruction (Bradlow *et al.*, 1997; Lambacher *et al.*, 2005; Lengeris and Hazan, 2010; Thomson, 2011) (for a recent review of the production training literature, see Sakai and Moorman, 2018). Such findings are encouraging for teachers and learners alike because computer-based training can supplement production teaching, especially in foreign language settings where authentic input is usually lacking. However, it still remains to be shown whether learners' improvement transfers to spontaneous speech. Bradlow *et al.* (1997), for instance, who were the first to show that auditory training on the English /r/-/l/ distinction can improve Japanese speakers' /r/ and /l/ production, examined isolated English words. Similarly for vowels, Lengeris and Hazan (2010) showed that auditory training improved Greek speakers' production of British English vowels in /bVt/ words.

Huensch and Tremblay (2015) and Huensch (2016) took important steps toward testing the effectiveness of high-variability phonetic training on learners' production of words spoken in sentence-level contexts. Huensch and Tremblay (2015) showed that training Korean speakers on isolated words and words in carrier sentences led to production improvement for English /ʃ/, /tʃ/, and /dʒ/ in both types of contexts.

Huensch (2016) extended Huensch and Tremblay's (2015) work by showing that the same type of training led to English /ʃ/, /tʃ/, and /ʤ/ production improvement in larger discourse contexts of continuous speech (read paragraphs) and that learning generalized to a new syllable structure (simple vs complex codas, e.g., *cash* vs *marsh*, respectively). Huensch (2016) pointed out that in her study, as opposed to previous training studies, learners were trained on both isolated words and words in carrier sentences, which provided additional variation to them and which may explain the creation of robust sound categories. While Huensch and Tremblay (2015) and Huensch (2016) demonstrated training-related improvement in continuous L2 speech, the ultimate goal of a learner remains, unarguably, to improve his/her production skills beyond the scripted speech level.

The main goal of the current study was to examine the effectiveness of high-variability phonetic training on the production of L2 vowels in spontaneous speech. This differs from previous studies because no orthographic input was provided to learners when evaluating training effects. Planning which English vowel to use is more difficult than reading words/sentences off a screen but also more ecologically valid when testing learning. Native speakers of Greek who had learned English in a foreign language setting participated in the study. One group received five sessions of training on Southern British English /iː/, /ɪ/, /ɒ/, /ɔː/, /ɑː/, /æ/, and /ʌ/. Their English vowel production was tested before and after training. Another group served as controls, i.e., produced the same pre-/post-test speech materials but received no training to evaluate any learning that could come from test repetition. English /iː/, /ɪ/, /ɒ/, /ɔː/, /ɑː/, /æ/, and /ʌ/ are particularly problematic for Greek learners both to perceive and produce (e.g., Lengeris and Hazan, 2010). Greek has a typical five-vowel system /i, e, a, o, u/ and thus there are many instances where a single Greek vowel exists in the area occupied by two or three English vowels. Focusing on the seven target vowels of this study, this applies to Greek /i/ and English /iː/ and /ɪ/; Greek /o/ and English /ɒ/ and /ɔː/; and Greek /a/ and English /æ/, /ʌ/, and /ɑː/ (Lengeris, 2009). With respect to production, Greek learners of English, at least at the initial stages of learning, replace those English vowels with the closest Greek ones (Lengeris and Hazan, 2010); English /iː/ and /ɪ/ with Greek /i/ (e.g., *beat* and *bit* sound the same); English /ɒ/ and /ɔː/ with Greek /o/ (e.g., *cot* and *caught* sound the same); and English /æ/, /ʌ/, and /ɑː/ with Greek /a/ (e.g., *back*, *buck*, and *bark* sound the same).

One of the main methodological difficulties when testing L2 production in spontaneous speech is how to quickly and efficiently elicit naturalistic data that contain the target items (see, e.g., the discussion in Huensch, 2016). To this end, a variant of the diapix task (Baker and Hazan, 2011) was used. Diapix is a spot-the-difference task designed to elicit target words uttered in conversational, laboratory-quality speech. The task requires two participants sitting in different booths and communicating via headsets toward finding the differences between two pictures. In the current study, it was decided to record one participant at a time while describing the differences between two pictures. Greek speakers' production of English vowels was assessed by a forced-choice identification task performed by English listeners. They identified English /iː/, /ɪ/, /ɒ/, /ɔː/, /ɑː/, /æ/, and /ʌ/ in (a) /bVt/ words read off a screen by Greek learners in a carrier sentence "Say bVt again" and (b) CVC words in spontaneous speech. Perceptual evaluation was corroborated by an acoustic analysis of vowel quality ($F1$ and $F2$ formant frequencies).

2. Method

A typical procedure for training studies was followed, consisting of a pre-test phase, a training phase, and a post-test phase (Logan *et al.*, 1991). As mentioned before, the control group only participated in the pre/post-tests.

2.1 Participants

Twenty-eight female speakers of Greek, all university students at the Aristotle University of Thessaloniki were tested; the trained group had 15 speakers and the control group 13. Across groups, participants had a mean age of 20 yrs (range = 19–21 yrs). They had 9–12 yrs of formal English instruction (B2 and C1 level learners in the Common European Framework of Reference for Languages - CEFR) but had very little, if any, interaction with native English speakers and none had spent a period of more than 1 month in an English-speaking environment according to a questionnaire completed by all participants before testing. None of the participants reported any hearing or language impairments.

2.2 Pre-/post-test recordings

The pre-/post-test recordings were conducted in the Phonetics Laboratory of the School of English, Aristotle University of Thessaloniki using a high-quality cardioid condenser microphone (Rode NT1-A, Sydney, Australia). In the sentence condition, participants produced two repetitions of the target vowels in /bVt /words in a carrier sentence. In the spontaneous speech condition, participants were told that there were around 25 differences between the two pictures (see Fig. 1), that they had 10 min to find as many differences as possible and that the recording would stop when 10 min had passed. The experimenter sat outside the recording room and monitored the task over headphones. In the post-test, participants were presented with two similar pictures but with another set of differences to avoid any learning/familiarity effects from test

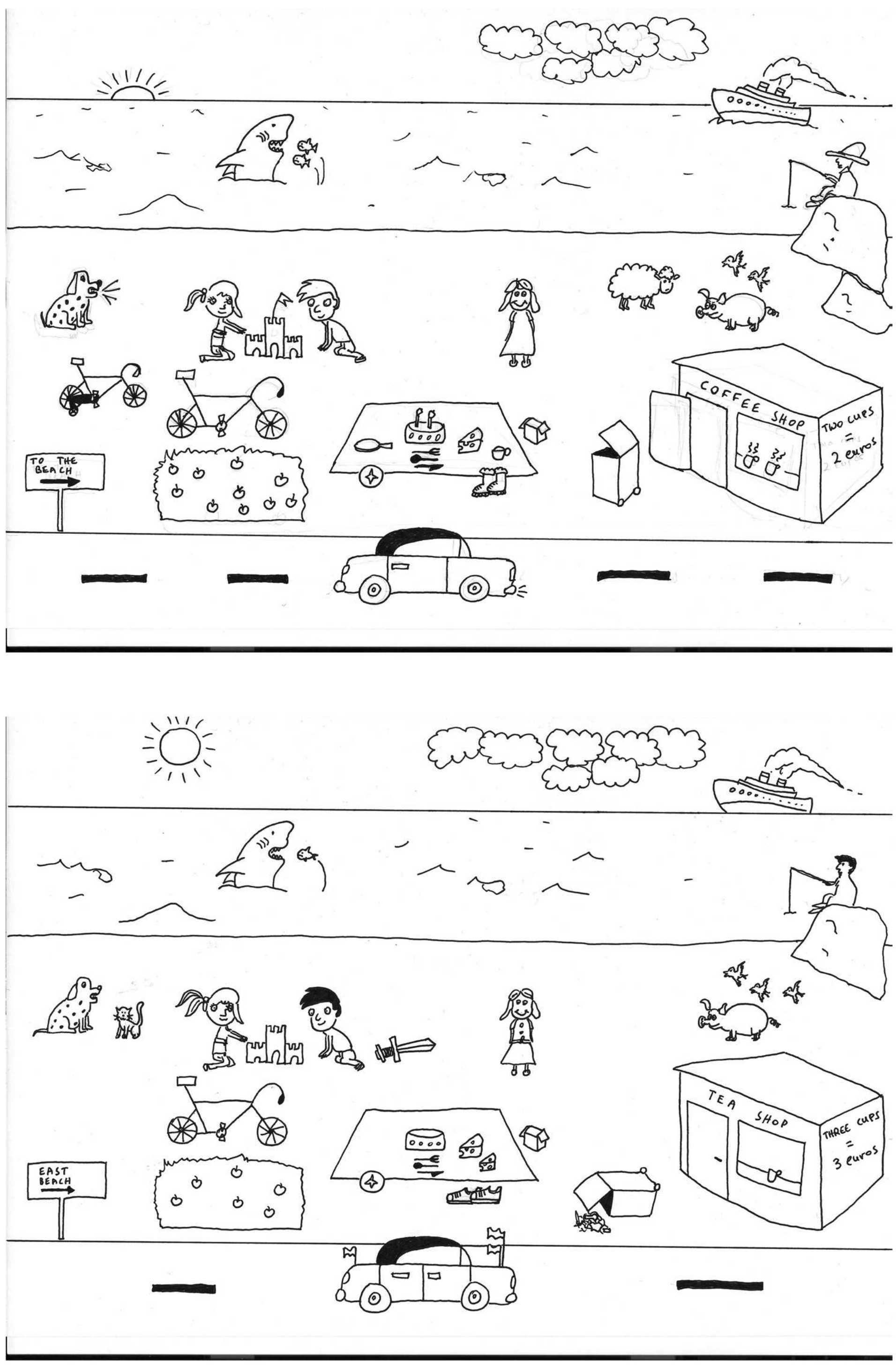

Fig. 1. Spot-the-difference task used to elicit target words in spontaneous speech in pre-test.

repetition. Each participant uttered a number of words containing the seven target vowels (e.g., *sheep*, *piece*, and *beach* for English /**i**:/, *ship*, *pig*, and *fish* for English /**ɪ**/). From the recorded words, two words produced by all 28 participants in pre-/post-test were selected for assessing English vowel production (*sheep* and *beach* for /**i**:/, *ship* and *pig* for /**ɪ**/, *cat* and *hat* for /**æ**/, *sun* and *cup* for /**ʌ**/, *shark* and *bark* for /**ɑ**:/, *dog* and *rock* for /**ɒ**/, and *sword* and *door* for /**ɔ**:/).

2.3 Training

The trained group completed five sessions of identification training with feedback for seven English vowels /**i**:/, /**ɪ**/, /**ɒ**/, /**ɔ**:/, /**ɑ**:/, /**æ**/, and /**ʌ**/ (same as the ones in the pre-/post-test). A different English speaker recorded the training stimuli for each training session (3 female, 2 male speakers). Training was administered in TP software (Rauber *et al.*, 2011); the software was installed on the trainees' laptops/desktops and they were asked to complete the training at home within a 10-day period without doing two training sessions on the same day. Stimuli were presented at a comfortable level set by each trainee. There were 196 stimuli per training session (7 vowels × 7 words × 4 repetitions). For example, there were four repetitions of *peel*, *beat*, *keel*, *team*, *deem*, *seat*, and *sheen* in the training stimuli for the purposes of teaching English /**i**:/. Each training session lasted about 30 min.

On each trial, the trainee heard an English word and chose one of seven bVt options as displayed on a computer screen. Before the experiment began, the trainees were told that they would hear consonant-vowel-consonant words with one of seven vowels found in *beat*, *bit*, *bat*, *but*, *Bart*, *bot*, and *bought*. If the target word was correctly identified, the trainee could proceed to the next trial. If the target word was misidentified, the correct answer was given and the trainee had to listen to the same stimulus again and choose the correct answer before continuing to the next trial testing another vowel.

2.4 Pre-/post-test assessment by native English listeners

Pre-/post-test recordings were presented to two Southern British English listeners (one male, one female) for identification using TP software. Each English listener performed 1568 judgments (28 speakers × 7 vowels × 2 speaking conditions × 2 tests × 2 repetitions) with vowels fully randomized by clicking on one of seven bVt options shown on a computer screen (*beat*, *bit*, *bat*, *but*, *Bart*, *bot*, and *bought*). The two English listeners were told to match the stimulus word (e.g., *sheep*) to a bVt word on the screen (e.g., *beat*). This was preferred over other options such as asking them to do so for stimuli in which the preceding and following consonant had been removed (e.g., listening to just the /**i**:/ portion of *sheep* and having to match it with *beat*) because listening to isolated vowels does not resemble naturalistic conditions.

2.5 Pre-/post-test acoustic analysis

Formant measurements for $F1$ and $F2$ frequencies at vowel midpoints were made using Praat's (Boersma and Weenink, 2018) formant-tracking algorithm (Burg) with default settings, followed by a manual check for errors. To quantify vowel distinctiveness, the overall Euclidean distance between the English vowels produced by Greek speakers before and after training was obtained by summing the individual Euclidean distances between adjacent vowel pairs /iː/-/ɪ/, /æ/-/ʌ/, /ɑː/-/ʌ/, and /ɒ/-/ɔː/; the larger the difference, the more differentiated the vowels produced by Greek speakers.

3. Results

3.1 Assessment by native English listeners

Figure 2 displays percent correct identification for English vowels produced by the trained group (upper panel) and the control group (lower panel) in sentences and in spontaneous speech in the pre-/post-test. Independent t tests with group as a between-subject factor showed that the two groups did not differ in pre-test in sentences (trained = 60.3% correct vs control = 61.3% correct) or in spontaneous speech (trained = 60.6% correct vs control = 61.8% correct), $p > 0.05$. An analysis of variance (ANOVA) of identification scores with test (pre-test, post-test) and condition (sentence, spontaneous) as the within-subject factors and group (trained, control) as the between-subject factor yielded a significant main effect of test [$F(1, 26) = 67.11$, $p < 0.001$, $\eta^2 = 0.72$], significant interactions between test and group [$F(1, 26) = 51.48$, $p < 0.001$, $\eta^2 = 0.66$], and test and condition [$F(1, 26) = 4.36$, $p < 0.001$, $\eta^2 = 0.14$], as well as a significant interaction between test, group and condition [$F(1, 26) = 4.36$, $p < 0.001$, $\eta^2 = 0.14$]. The test × group × condition interaction was explored by two ANOVAs

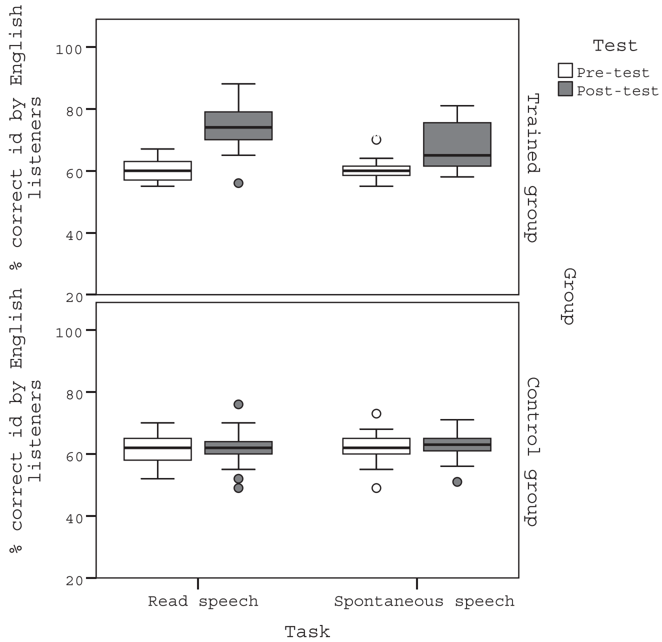

Fig. 2. Boxplots of English listeners' identification accuracy for English vowels produced by the trained (upper panel) and the control (lower panel) group of Greek speakers in the pre-/post-test. Whiskers extend to at most 1.5 times the interquartile range of the box. Circles mark outliers.

run for each group separately with test (pre-test, post-test) and condition (sentence, spontaneous) as within-subject factors. For the trained group, there was a significant main effect of test [$F(1, 14) = 78.7$, $p < 0.001$, $\eta^2 = 0.85$], and a significant interaction between test and condition [$F(1, 14) = 6.4$, $p < 0.05$, $\eta^2 = 0.32$]. Paired sample t tests showed that this was due to improvements being larger for English vowels produced in sentences (from 60.3% correct in pre-test to 74.1% correct in post-test, i.e., an improvement of 13.8%) than improvements for English vowels produced in spontaneous speech (from 60.6% correct in pre-test to 67.7% correct in post-test, i.e., an improvement of 7.1%), $t(14) = 2.54$, $p < 0.05$.

3.2 Acoustic analysis

Figure 3 plots in the $F1 \times F2$ vowel space the seven English vowels spoken in sentences and in spontaneous speech in pre- and post-test. It can be seen that, across speaking conditions, English vowels produced in the pre-test [Figs. 3(a) and 3(b)] were arranged around three clusters; /iː/-/ɪ/, /ɑː/-/æ/-/ʌ/, and /ɒ/-/ɔː/ with very little, if any, differentiation of the members of each cluster from one another. This suggests that Greek speakers replaced the seven target vowels with the closest native ones; they used Greek /i/ for English /iː/ and /ɪ/; Greek /a/ for English /ɑː/, /æ/, and /ʌ/; and Greek /o/ for English /ɒ/ and /ɔː/. In the post-test [Figs. 3(c) and 3(d)], there is considerable differentiation between vowels. When comparing vowels produced in sentences vs spontaneous speech, consistent with English listeners' perceptual assessment discussed in Sec. 3.1, differentiation seems to be larger for vowels spoken in the former condition.

These observations were tested by an ANOVA of the overall Euclidean distance between vowels with test (pre-test, post-test) and condition (sentence, spontaneous) as factors. The ANOVA yielded significant main effects of test [$F(1, 14) = 12\,356.1$, $p < 0.001$, $\eta^2 = 0.98$] and condition [$F(1, 14) = 508.4$, $p < 0.001$, $\eta^2 = 0.96$], and a significant interaction between test and condition [$F(1, 14) = 79.6$, $p < 0.001$, $\eta^2 = 0.84$]. Paired sample t tests exploring the interaction showed that the change in Euclidean distance from pre- to post-test was larger in sentences (from 165.43 to 617.34 Hz, i.e., a change of 451.8 Hz) than in spontaneous speech (from 141.42 to 534.4 Hz, i.e., a change 392.4 Hz), $t(14) = 8.91$, $p < 0.001$.

4. Discussion

This study examined the effects of high-variability phonetic training on the production of English vowels in sentences and in spontaneous speech. The participants were native speakers of Greek who had learned English in a foreign language setting. Improvement was evaluated pre- and post-training via an identification task performed

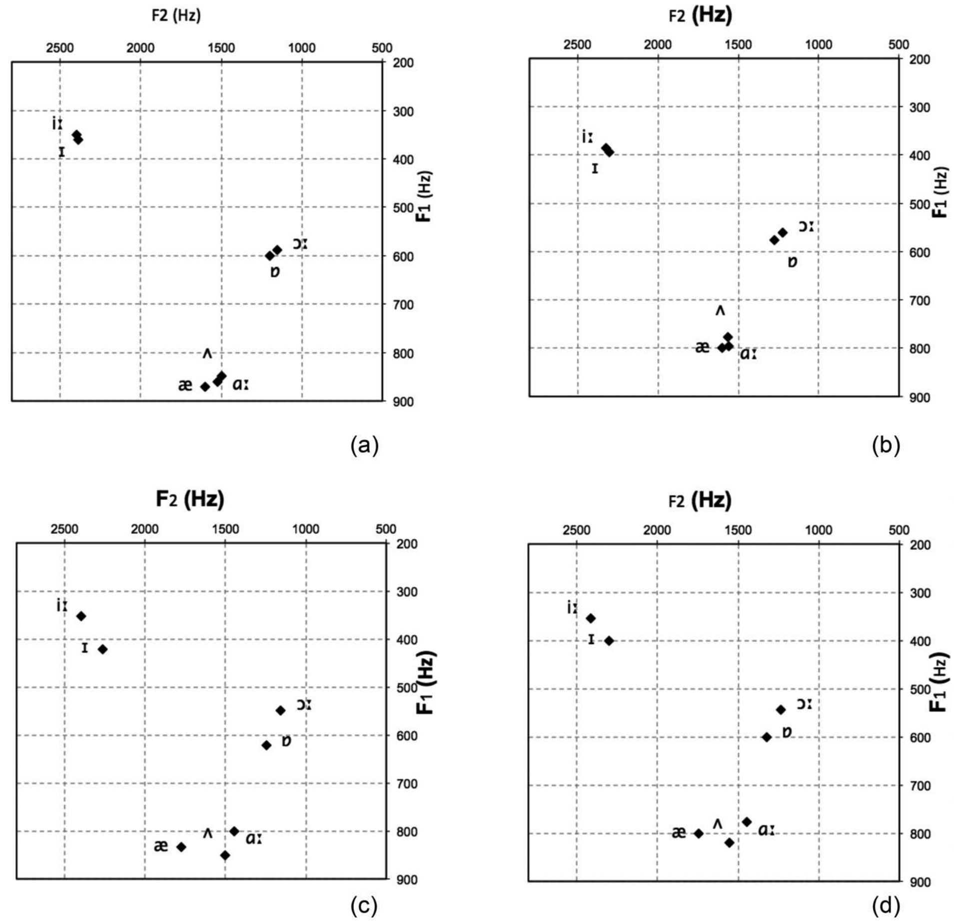

Fig. 3. English vowels produced by the trained group in pre-test sentence (a), pre-test spontaneous (b), post-test sentence (c), and post-test spontaneous (d) materials plotted in the $F1/F2$ vowel space.

by English listeners and by an acoustic analysis of vowel quality using a combined $F1/F2$ measure.

The results showed that identification scores for English vowels were higher after training both in sentences and in spontaneous speech. This extends previous findings regarding the effects of training on L2 vowel production in read materials (Lambacher *et al.*, 2005; Iverson and Evans 2009; Lengeris and Hazan, 2010). The acoustic analysis showed that this improvement was, in part, due to Greek speakers making larger distinctions in vowel quality in the post-test than they did in the pre-test (where they simply replaced the seven target vowels with the closest Greek vowels available to them, i.e., they used Greek /i/ for English /**i**ː/ and /**ɪ**/; Greek /a/ for English /**ɑ**ː/, /**æ**/, and /**ʌ**/; and Greek /o/ for English /**ɒ**/ and /**ɔ**ː/). It is possible that this improvement was also due to improvements in the use of temporal cues. For instance, after training, Greek speakers might have used vowel lengthening more consistently than before training, which would contribute to better identification of the vowels /**i**ː/, /**ɑ**ː/, and /**ɔ**ː/ by native listeners. In other words, identification scores may reflect improvement in both vowel quality and length, whereas acoustic analyses reflect training-related changes in vowel quality only.

Production improvement was larger for English vowels produced in sentences than for vowels produced in spontaneous speech, which was reflected both in identification scores and in the acoustic analysis. This is not surprising considering how demanding the spot-the-difference task is compared to a sentence reading task. It seems that, while Greek speakers significantly improved in both tasks, it was easier for them to apply the knowledge they had recently acquired when reading sentences than when speaking spontaneously.

Although production deviations can cause a number of problems to L2 learners such as speaking anxiety (Baran-Łucarz, 2011) and negative evaluation and discrimination (Munro, 2003), production teaching is a highly neglected area in TESOL and TEFL. Even when production has a clear place in the curriculum (which is rarely the case), some teachers believe that improvement is not possible while others may lack the confidence and/or the ability to teach production (e.g. Breitkreutz *et al.*, 2002). The results of this study address both issues as they show that a few hundred pre-recorded words delivered to L2 learners via a computer can have an impact on their spontaneous speech production. High-variability phonetic training is a quick, effective, and easy to implement approach (using freely available software like TP) that can benefit not only ESL/EFL students but also other groups of L2 learners interested in improving their production. Future work could investigate the effects of training on learners' production of vowels in conversational speech.

References and links

Baker, R. C., and Hazan, V. (**2011**). "DiapixUK: A task for the elicitation of spontaneous speech dialogs," Behav. Res. Meth. **43**, 761–770.

Baran-Łucarz, M. (**2011**). "The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation," Stud. Sec. Lang. Learn. Teach. **1**, 491–514.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Flege*, edited by M. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer" [Computer program], version 6.0.39, retrieved from http://www.praat.org/ (Last viewed April 3, 2018).

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Breitkreutz, J., Derwing, T. M., and Rossiter, M. J. (**2002**). "Pronunciation teaching practices in Canada," TESL Can. J. **19**, 51–61.

Huensch, A. (**2016**). "Perceptual phonetic training improves production in larger discourse contexts," J. Sec. Lang. Pronunc. **2**, 183–207.

Huensch, A., and Tremblay, A. (**2015**). "Effects of perceptual phonetic training on the perception and production of second language syllable structure," J. Phon. **52**, 105–120.

Iverson, P., and Evans, B. (**2009**). "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German Speakers," J. Acoust. Soc. Am. **126**, 866–877.

Iverson, P., Pinet, M., and Evans, B. G. (**2012**). "Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels," Appl. Psycholing. **33**(1), 145–160.

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

Lengeris, A. (**2009**). "Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese," Phonetica **66**, 169–187.

Lengeris, A., and Hazan, V. (**2010**). "The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek," J. Acoust. Soc. Am. **128**, 3757–3768.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Munro, M. J. (**2003**). "A primer on accent discrimination in the Canadian context," TESL Can. J. **20**, 38–51.

Rauber, A., Rato, A., Kluge, D., and Santos, G. (**2011**). TP- S (Version 1.0) [Application software]. Retrieved from http://www.worken.com.br/tp_regfree.php (Last viewed May 21, 2017).

Sakai, M., and Moorman, C. (**2018**). "Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research," Appl. Psycholing. **39**, 187–224.

Thomson, R. I. (**2011**). "Computer assisted pronunciation Training: Targeting second language vowels: Perception improves pronunciation," CALICO J. **28**, 744–765.