Comparing the efficiency of vowel production training in immersion and non-immersion settings for Arabic learners of English

Wafaa Alshangiti [a, b] and Bronwen G. Evans [a]

Speech Hearing and Phonetic Sciences, UCL (University College London) [a]
English Language Institute, KAU (King Abdulaziz University) [b]
walshangiti@kau.edu.sa; bronwen.evans@ucl.ac.uk

ABSTRACT

Previous research in second language (L2) learning has shown that learners benefit greatly from focused training in their L2 (e.g., [5, 9]). The current study investigated whether or not the results of production training might differ according to learning environment, specifically whether or not learners in a non-immersion setting who have fewer opportunities to interact with native speakers might benefit differently from learners in an immersion setting.

Two groups of Arabic learners of English, one in London, and one in Saudi Arabia completed 5 sessions of vowel production training. A battery of pre- and post-tests tested for potential improvements in speech production and perception. Results indicated that learning environment affected learning outcomes; learners in the non-immersion environment improved in both perception and production, while learners in an immersion setting improved predominantly in production.

Keywords: L2 learning, production, perception.

1. INTRODUCTION

Auditory phonetic training has been proven to be highly successful in improving learning of difficult L2 phonemes. Most of these studies have used High Variability Phonetic Training (HVPT) where listeners listen to and identify phonemes produced in different contexts by multiple speakers, and receive corrective feedback on their responses (e.g., [5, 9]). Though some studies have trained learners on entirely new phonemic contrasts in a language that they do not use (e.g., [8]), many have focused on training L2 learners of English on English phoneme with L2 learners living in an English-speaking country (e.g., [9]).

More recently, HVPT has also been used to investigate the effects of intensive training on inexperienced learners living in their home country. For example, Iverson et al (2012) trained French learners of English with differing levels of English experience on English vowels; French speakers in France (inexperienced learners), and French speakers in London (experienced learners). Despite the fact that the French speakers in London had many more opportunities to interact with native English speakers, the results demonstrated that both training groups improved similarly. However, it is less clear whether production training operates similarly. For instance, learners who are trained in an immersion setting and who have more opportunities to consolidate learning through daily interaction with native speakers, may improve more than those trained in their home country. On the other hand, learners in an immersion setting may be exposed to richer array of stimuli than can be delivered by several sessions of training, and thus may receive little additional benefit from this type of focused training in comparison to those trained in their home country.

The present study compared the potential benefits of production training for production and perception of English vowels by Saudi Arabic learners of English in London, U.K. (immersion setting) and Jeddah, Saudi Arabia (SA; non-immersion setting). On average, the two groups of learners had similar initial ability with English, but differed in their experience. They completed the same production training and a battery of pre- and post-tests assessed improvements in production and perception.

2. METHODS

2.1. Participants

Twenty-five native Arabic participants (24 Saudi Arabian, 1 Egyptian) were tested; 16 in London, and 9 in Jeddah. All participants had the standard Arabic six-vowel system, were aged 19-43 years (median 28 yrs), and had begun to learn English aged 1-27 years (median 13yrs). London participants had 3 months-10 years (median 28 mths) experience of living in an English-speaking country. Those in SA were selected to have little experience of living in an English-speaking country; only one participant reported having lived in the U.K. for 1 year, 10 years previously. All participants completed the written grammar section of the Oxford placement test [5] to evaluate their English proficiency. This test was used to categorize learners as low (LP) or high proficiency (HP), giving the following split: London (10 HP, 6 LP) and SA (1 HP, 8 LP).

2.2 Apparatus and stimuli

All training, pre- and post-tests were conducted in a quiet room with stimuli played over headphones at a user-controlled comfortable level. Production training was delivered by an instructor (first author) using a custom-made computer program, CALVin (Computer Assisted Learning for Vowels Interface; [2]). The stimuli consisted of /h/-V-/d/ keywords plus two example words, and the isolated vowels. The keywords were arranged into groups of minimal pairs selected by dividing 14 British English vowels into five highly confusable clusters; High/front: /i ɪ e/ (e.g., *heed*, *hid*, *head*); Open: /æ ʌ ɒ/ (e.g., *had, hud, hod*); Central/low back: /ɜ ɑ ɔ/ (e.g., *heard, hard, hoard*); Back: /u aʊ əʊ/ (e.g., *who'd*, *how'd*, *hoed*); and Diphthongs: /eɪ aɪ/ (e.g., *hayed*, *hide*). The clusters were selected based on hierarchical cluster analyses on English vowel identification data from Arabic learners of English [3]. All stimuli were recorded by a monolingual male speaker of Standard Southern British English.

The pre- and post-test stimuli for vowel identification and category discrimination were the same as those used in [10]. They comprised recordings of English /b/-V-/t/ words [English vowels that created non-words (e.g., /ʊ/) were not included], and were recorded by 10 speakers of British English (5 male, and 5 female); none of these words or speakers were used in the training, ensuring that all pre- and post-tests measured generalization to new stimuli. The stimuli for speech recognition in noise were recordings of IEEE sentences (72 lists of 10 sentence). Each sentence contains 5 key words that were identified by the listener, e.g., "Glue the sheet to the dark blue background". The sentence lists were recorded by a male SSBE speaker and were taken from existing recordings made at University College London. All the recordings were made in sound treated room. The speech was mixed with white noise; the noise level was fixed to 71dBA, and the level of the speech was varied adaptively.

2.3. Procedure

2.3.1. Training

Production training consisted of five 40 minute sessions using CALVin completed over 1-2 weeks with no more than one session per day. CALVin was used to play the 14 keywords (one for each vowel) and each with 2 example words, and displayed an animation showing the positions of the tongue, jaw and lips for each stimulus. Each training session started and finished with a 10-minute phase on all 5 vowel clusters (high/front, open, central/low back, back and diphthongs, with the order reversed in the last 10-minute phase). The middle 20 minutes consisted of training on a specific cluster. Prior to the first session, participants completed a 10-minute familiarisation session with the software and were shown the relationship between the different positions of their tongue, jaw and lips and resulting vowel sound, using a hand mirror to observe the changes in articulator position.

For each vowel, subjects heard a keyword (e.g., *heed*) and then the vowel (e.g., /i/) in isolation. They then viewed an animation of its articulation (in midsagittal section), and were guided through a function that described the principal articulatory positions. They were asked to produce the isolated vowel first, then the keyword and example words (e.g., *heat*, *feet*). Then they were asked to record themselves producing the isolated vowel, keyword, and the example words, play them back and compare them with the native speaker. They also received feedback from the instructor. All training was completed in English.

2.3.2. Pre/post-tests

There were four pre- and post-tests, (i) vowel identification, (ii) category discrimination, (iii) speech recognition in noise, and (iv) English vowel production. The vowel identification test consisted of 84 trials of a closed-set identification task consisting of /b/-V-/t/ words, randomly selected on each trial. The category discrimination task involved 66 trials, each consisting of three English /b/-V-/t/ words spoken by three different speakers. The three words contrasted different vowel pairs where two words were the same (i.e., the same vowel) and participants had to identify the one that was different. The vowel pairings were /ɪ/-/ɛ/, /ɒ/-/ʌ/, /eɪ/-/aɪ/, /aʊ/-/əʊ/, /ɑ:/-/ɔ:/, /ɜ/-/ɑ/, /u/-/əʊ/, /i/-/ɛ/, /u/-/aʊ/, /ɜ/-/ɔ:/, /i/-/ɪ/, and were selected based on previous experiments on vowel perception in Arabic learners of English. The most confusable vowel pairs were selected in descending order until each of the 14 stimulus vowels appeared at least once.

For the speech recognition in noise test, participants listened to IEEE sentences in noise. They were asked to verbally repeat what they had heard and the number of correctly identified keywords was recorded manually. An adaptive Levitt-procedure ([4]) varied the signal-to-noise ratio to find the Speech Reception Threshold (SRT). Participants identified two blocks of 20 sentences at the pre- and post-test. Each sentence was presented only once. Finally, participants recorded 3 repetitions of each of the /b/-V-/t/ words that they identified in the vowel identification task. The recordings of all participants in the pre- and post-test were analysed acoustically and identified by native English speakers.

3. RESULTS

3.1. Vowel identification

Figure 1: Boxplots showing overall performance (average proportion correct) on the vowel identification task across the two training groups.

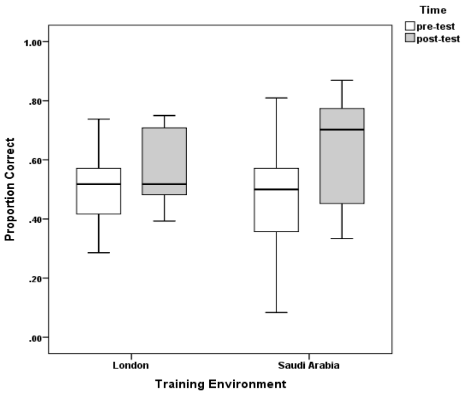


Fig.1 shows the pre- and post-test vowel identification accuracy for learners in London and SA. A logistic mixed effects model demonstrated that there was a main effect of time: $\chi^2(1)$=35.65, p<0.001. The planned contrasts confirmed that there was an improvement in identification from pre- to post-test, b=-0.276589, SE=0.046319, z=-5.971, p<.001. There was no significant effect of training environment, $\chi^2(1)$=0.3268, p>.05 However, there was a significant two-way interaction between training environment and time, $\chi^2(1)$=15.556, p<0.001; the group that was trained in SA improved significantly more than the equivalent group in London, b=-0.16874, SE= 0.042784, z=-3.944, p<.001.

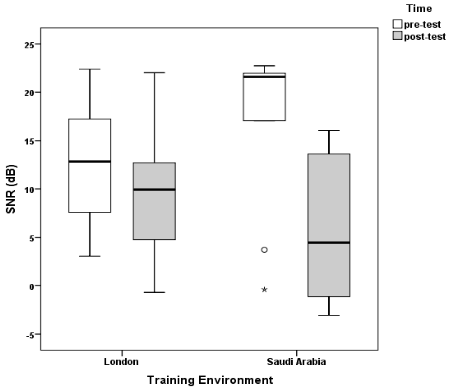
3.2. Category Discrimination

A linear mixed model was built for category discrimination data. The best fit-model included time (pre-post) as a fixed factor, and participant and word pair as random factors. There was no significant effect of the factors, indicating that there was no overall significant change in category discrimination performance from pre- to post-test. However, there was a significant interaction between training environment and proficiency level, $\chi^2(1)$=6.866, p<.05. The planned contrasts showed that the HP learners who were trained in London improved more than those trained in SA, b=-3.518, SE=1.663, pMCMC<.05. However, there was only one HP participant in SA group so it is difficult to know how generalisable this finding is.

3.3. Speech recognition in noise

All participants improved after training; a logistic mixed effects model indicated that there was a significant effect of time, $\chi^2(1)$=6.661, p<.05 and the planned contrasts showed a significant change from pre- to post-test, b=2.634, SE=1.0205, pMCMC<.05. The main effect of proficiency was also significant, $\chi^2(1)$=5.267, p<.05. The planned contrasts indicated that the LP participants performed better at the post-test than HP ones, b=-3.8173, SE=1.663, pMCMC<.05. There was no main effect of training environment but there was a significant two-way interaction with proficiency, $\chi^2(1)$=4.475, p<.05. The planned contrasts showed that the LP participants who were trained in SA performed better at the post-test compared to the equivalent proficiency group who were trained in London, b=-3.518, SE=1.66, pMCMC<.05. However, there was no significant interaction between HP proficiency and training environment, possibly because the HP group in SA only contains one participant.

Figure 2: Boxplots showing speech reception threshold (dB SPL) for L2 listeners across training environment at the pre- and post-tests.



3.4. English vowel production

3.4.1. Acoustic analysis

Monophthongs were divided into three groups: M1 (*beat, bit, bet, bert*), M2 (*bat, but, bart*), and M3 (*boot, bought, bot*). Changes in F1 and F2 were analysed using linear mixed effects models. We summarize the main findings below.

For M1 vowels, there was no main effect of time for F1, suggesting that there was no significant change in this dimension from pre- to post-test. This was surprising as learners in both groups appeared to alter F1 for /ɪ/-/e/, such that after training they produced /ɪ/ with lower F1 values and /e/ with higher

F1 values, so that these vowels were more similar to native F1 values for this contrast.

For M2 vowels, there was a significant effect of training environment for F2, $\chi^2(1)=6.770$, p<.05. London & SA participants produced these vowels using different F2 values, though the difference was small; b=-0.088, SE=0.029, pMCMC<.05.

For M3 vowels, there was a significant effect of proficiency for F1, $\chi^2(1)=4.4301$, p<.05. The planned contrasts indicated a significant difference in F1 values for HP compared to LP participants, b=-0.0717, SE=0.034, pMCMC<.05. HP participants tended to produce *bot* and *bought* with lower F1 values than the LP participants, though these effects were small.

3.4.2. Vowel intelligibility

A logistic mixed-effects model demonstrated that there was a significant main effect of time, $\chi^2(1)=8.615$, *p*<.05, and a significant main effect of proficiency $\chi^2(1)=4.035$, p<.05. As displayed in Fig. 3, the planned contrasts confirmed that there was a significant improvement in intelligibility from pre- to post-test, such that all participants were more intelligible after training, b=-0.355, SE=0.1209, z=-2.935, p<.05, though the effect seemed larger for SA participants. Investigation of confusion matrices indicated that this was predominantly due to improvements in intelligibility of *bit, bet* and *bought*. HP participants were also more intelligible than LP participants overall, b=0.2208, SE=0.1099, z=2.009, p<.05.

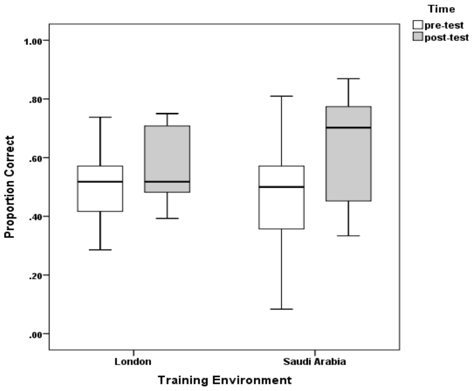


Figure 3: Boxplots showing the identification (proportion correct) for vowels produced by L2 speakers, split by training environment.

4. DISCUSSION

The present study investigated whether training environment (immersion vs. non-immersion) affects the outcome of production training for production and perception. The results demonstrated that both groups improved in production; an acoustic analysis showed that participants changed their production of *bit* and *bet*, and these vowels, along with *bought* were also more intelligible after training. However, the group that was trained in a non-immersion setting (i.e., in SA) appeared to benefit more from training overall, improving more in vowel identification and speech recognition in noise as well as production.

Previous work has suggested that training is domain-specific, that is, that production training improves production but not perception, and perception training improves perception but not production ([7]). However, these results support the notion that not only natural exposure to speech improves performance in perception, but that some aspect of directing learners' attention to phonetic differences in production is beneficial for speech perception as well as production.

Why did SA improve more than London learners in perception? Perhaps it is the case that because the SA group did not have regular interactions with English speakers, they used the production training as a more holistic tool for acquiring English than did the London group. Another possibility is that because participants in SA were mostly recruited from a language institute, they may have been keener to learn and improve their English perception and production. In contrast, participants in London were mostly recruited from Brunel University in London, were not studying English and instead, spent a lot of time working independently in laboratory-based research. It is possible that at least in terms of improving their production and perception for spoken English, this group of participants were not as motivated (see [6] for a review).

Although all SA participants improved in vowel identification performance, they did not reliably improve in their category discrimination. One explanation is that participants are better at distinguishing certain categories based on their existing representations, and perform well with these in identification tasks as a result of training, but do not change their underlying representations, i.e., no change in performance in the category discrimination task. This provides additional evidence for the hypothesis that training does not lead to low-level changes in category representations but instead, enables learners to better match their existing representations with the sounds that they hear and have learned to produce in the L2 [9].

In sum, the results suggest that production training can yield improvements in perception as well as production, but that this may be dependent on the learning environment itself, as well as learners' motivation for learning.

5. REFERENCES

[1] Allan, D. (1992). *Oxford Placement Tests 1.* Oxford University Press, Oxford, UK.

[2] Alshangiti, W., Evans, B. G. (2014). Investigating the domain-specificity of phonetic training for second-language learning: Comparing the effects of production and perception training on the acquisition of English vowels by Arabic learners of English. In the *Proceedings of the International Seminar for Speech Production*, Cologne, Germany, May 2014.

[3] Alshangiti, W. *Speech production and perception in adult Arabic learners of English: A comparative study of the role of production and perception training in the acquisition of British English vowels.* Doctoral thesis submitted to University College London, UK.

[4] Baker, R. J., Rosen, S. (2001). Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking. *British Journal of Audiology*, *35*(1), 43-52.

[5] 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. *J. Acoust. Soc. Am*., 101, 2299-2310.

[6] Dornyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal* 78 (3).

[7] Hattori, K. (2009). *Perception and Production of English /r/-/l/ by adult Japanese speakers.* Doctoral thesis submitted to University College London, UK.

[8] Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *J. Acoust. Soc. Am*., 116(4), 2384-2394.

[9] Iverson, P., Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am*., 126(2), 866–77.

[10] Iverson, P., Pinet, M., Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(01), 145-160.