Learning to speak in a second language: Does multiple talker production training benefit production of English vowels in Arabic children?

Wafaa Alshangiti[1], Bronwen G. Evans[2] and Mark Wibrow[3]
[1]*English Language Institute, King Abdulaziz University, Jeddah, Saudi Arabia,* [2]*University College London, London, UK,* [3]*Cloudfind, Bath, UK.*
[1] *walshangiti@kau.edu.sa,* [2] *bronwen.evans@ucl.ac.uk,* [3] *m.wibrow@gmail.com*

ABSTRACT

High-variability phonetic training (HVPT) has been shown to be highly effective in improving second-language (L2) perception in adults, and to also benefit production. In contrast, recent studies have suggested that children may benefit more from low-variability phonetic training (LVPT), in particular for production. The present study compares HVPT and LVPT articulatory training for production and perception of Standard Southern British English (SSBE) vowels in children in a non-immersion context. Forty-six monolingual Arabic children aged 8-12 years were randomly assigned to single- (LVPT) or multi-talker (HVPT) training. Both groups completed five articulatory training sessions on 18 vowels and a battery of perception and production tests evaluated improvement. The results showed that the LVPT group performed better not only in production, but also in category discrimination. The results support previous studies that have suggested that LVPT training might be more successful with children.

Keywords: Phonetic training, Articulatory training, Vowel production by Arabic children

1. INTRODUCTION

The majority of training studies have used high-variable phonetic training (HVPT) with adult participants, presenting training materials recorded by multiple speakers and/or multiple phonetic contexts, and typically focussing on the effects of perceptual training for perception (e.g., category identification or category discrimination). Although the use of low-variable phonetic training (LVPT) has been shown to provide some improvements in perception, only listeners trained using HVPT perform better when presented with new talkers [13]. Subsequently, HVPT has been the dominant approach in the field, and has been found to be effective in improving the perception of difficult non-native contrasts [11] with some studies finding that this also transfers to production tasks (e.g. [2,19]).

However, more recent work with both adults and children has failed to find a high variability advantage. For example, Greek children and adults on the perception of English /i/-/ɪ/ contrast for ten sessions in a computer-based word learning game, where they heard imageable words produced by either a single (LVPT) or multiple talkers (HVPT) [6]. Their task was to decide which picture best represented the word they heard. The pictures were minimal pairs, e.g., if they heard *sheep*, they chose between a picture of a *sheep* and a *ship*. They could re-play the stimuli, and had immediate feedback. They completed a battery of pre- and post-tests including category discrimination and word-learning. These showed that both adults and children improved during training, but both improved more with LVPT. Although adults showed a numeric advantage for HVPT on a 3-interval oddity task, this was not reliable or only near-reliable, statistically. In contrast, children showed the reverse effect: they improved significantly more in LVPT.

In a similar study, Spanish adults and children were trained on the English /i/-/ɪ/ contrast, also using a computer-based word learning game with feedback [4]. Participants were assigned to either a single- (LVPT) or multiple talker (HVPT) training condition and completed 5 training sessions. To assess potential improvement, participants completed a category discrimination task (words and non-words, new talkers) and a word repetition task (production) before and after training. All subjects improved across training sessions, but LVPT-children improved more than HVPT-children. However, only children and not adults, improved in word-based category discrimination, and only those in the HVPT condition improved in non-word discrimination. In contrast, LVPT but not HVPT-children improved in their production of the /i/-/ɪ/ contrast. One possible interpretation of these results is that LVPT might be more beneficial for the acquisition of new articulatory targets but that variability may be crucial for the generalization of perceptual learning.

To further investigate the role of variability in L2 learning, the current study took a different approach. We built a child-friendly, computer program, CALVin (Computer Assisted Learning for Vowels interface) which we used to train native, monolingual Saudi Arabic children in the production rather than perception of SSBE vowels. Pre-/post-tests

investigated whether or not there was any improvement in production, and whether this also led to improvements in perception.

The evidence for production-perception transfer in production training studies with adults is mixed. For example, a study in which Japanese learners were trained with English /r/-/l/ production over 10, one-to-one sessions using a multi-faceted approach that used explicit feedback from the instructor, and feedback with synthesised versions of their own productions, found that whilst production became more native-like, perception of English /r/-/l/ did not improve [8]. Likewise, a study comparing the effects of HVPT and production-based training for adult Arabic learners of English also found that training appeared to be domain-specific: those given HVPT improved in vowel identification but not vowel production, whilst those given production training showed only small improvements in performance on perceptual tasks, but much greater improvement in production [1]. In contrast, adult US English speakers trained in either the production or perception of a Spanish 3-way intervocalic contrast, improved primarily in their identification of the contrast [10].

Based on this work, we predicted that our children would benefit most from single-talker training, but that any improvement in production might not transfer to perception.

2. METHODS

2.1. Participants

Forty-six native Saudi Arabic children aged 9-12 years old (all female) were randomly assigned to LV (single-talker) or HV (multiple-talker) training condition. Participants were recruited from public schools in Jeddah, were all in their final 2 years of primary school, and had had little prior exposure to English. They had begun learning the English alphabet and some words aged 9yrs. None of the participants reported any history of speech, hearing or language impairments. In addition, 5 SSBE speakers aged 22-46 yrs (median 30 yrs), rated participants' production for intelligibility.

2.2. Apparatus and stimuli

For the pre-/post-tests, stimuli were played over headphones at a comfortable level, and a laptop was used to play stimuli and collect responses. Articulatory training was delivered by an instructor (1st author) using CALVin. Over the training sessions, the children were trained on 18 vowels covering the majority of the SSBE vowel space; /iː, ɪ e, ɜː, æ, ɑː, ɒ, ɔː, ʌ, ɜː, uː, ʊ, eɪ, aɪ, aʊ, əʊ, eə, ɔɪ, ɪə /. They heard vowels produced in isolation and in CVC example words, e.g., /əʊ/ for *coat*.

Stimuli were recorded by 4 native adult SSBE speakers (2 male, 2 female). In addition, AV recordings were made for the example and key words. These were later embedded in CALVin so that children could see lip and jaw movement. Lastly, a young, male child was recorded producing instructions. Stimuli were played using a high-quality external speaker connected to a laptop.

The pre- and post-test stimuli for category discrimination (recordings of the /b/-V-/d/ and /b/-V-/t/words), and word imitation (/h/-V-/d/ words) were recorded by 4 British English speakers (2 male and 2 female); none of these words or speakers were used in the training, ensuring that all pre- and post-tests measured generalization to new stimuli.

2.3. Procedure

Participants in both training groups completed 5 training sessions, each focussing on a different set of vowels, selected to be highly confusable based on the results of previous studies with Arabic learners [3]. They completed only 1 training session per day, and all training sessions were completed over two weeks. Participants in the HV group were trained with a different SSBE speaker for each of the first 4 sessions, and the fifth session included all 4 of these speakers. The LV group were trained using only 1 SSBE speaker. Children were trained together in groups of 4/5, which enabled them to learn in a similar environment to the one that they were familiar with in the classroom.

In the first session, children were introduced to CALVin. Each training session then proceeded in the same way. At the beginning of each session, the instructor explained how jaw opening, tongue movement and the lips affect the way different vowels are produced, and encouraged children to practise producing different sounds. The children were then invited to choose one of the 5 vowel groups (the instructor ensured no group was repeated and that the order in which vowel groups were completed was counterbalanced across groups), and within that, 1 vowel, represented by an image and a keyword. When clicking on the keyword, they heard the keyword, recorded themselves producing it and listened back to their recordings. The instructor gave instructions, in Arabic, about how to move the jaw, lips and tongue to produce similar vowels. Participants were asked to produce a vowel similar to the one that was shown in the animation, then try it individually. After feedback from the instructor, each child recorded their production of the isolated vowel, before comparing it to the SSBE speaker in CALVin. They then did the

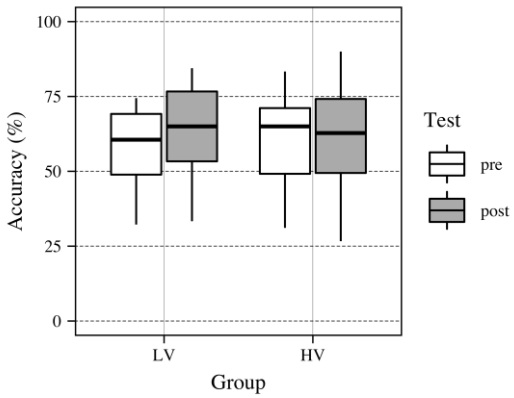
same thing for the example word, but this time they watched the AV recording before recording their own version and listening again to compare it with the native speaker. They then repeated these steps for the other 2-3 vowels in the group. Each session ended with a review of the vowels covered, led by the instructor, and lasted approximately 45 mins.

3. RESULTS

Independent samples t-tests on the pre-test category discrimination (% correct) and vowel intelligibility (proportion correct) scores, showed that all children, regardless of age and training condition (HV, LV) performed similarly, confirming that there was no significant difference between the groups at pre-test, p >.05. All further analyses therefore investigate potential differences as a result of training condition.

3.1. Category discrimination

Figure 1: Boxplots showing category discrimination accuracy at the pre- and post-test for low and high variability training.



A linear mixed-effects logistic regression model was built with the correct/incorrect binomial responses as the dependent variable, training group (HV, LV) and time (pre, post) as fixed factors and participants as crossed random effect. The main effect of test (pre-post) was significant, χ^2 (1) = 9.122**,** p<.05, which suggests that there was a change in category discrimination accuracy from pre- to post-test. There was no significant effect of training group, χ^2 (1) = 0.938, p>.05. However, the model showed a significant interaction between training group and test, χ^2 (1) = 4.667, p<.05, and the contrast between factors showed that the LV group performed better at the post test than the HV group, b=-0.1998, SE=0.0924, z=2.161, p<.05.

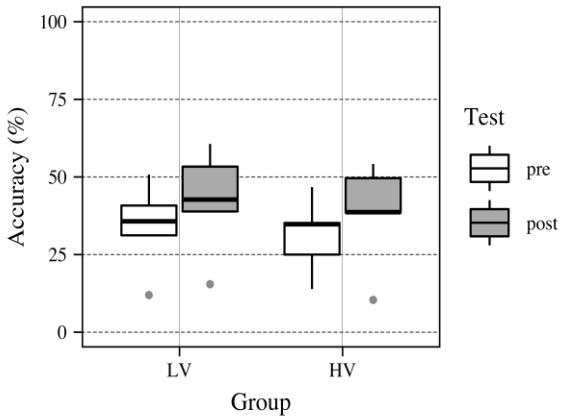
3.2. Imitation task

3.2.1 Acoustic analysis

Linear mixed models were built separately for F1 and F2, with test (pre-post) and group (LV-HV) as fixed factors, and participants and vowels as random factors. The models showed that there was no significant change in the F1 and F2 values from pre to post-test in both training groups. Using data collected from [9] to model a 'prototypical speaker' of SSBE, the participant's productions were classified according to the closest vowel in the SSBE model (as measured using the Mahalanobis distance). The classification accuracy showed a small improvement for LV speakers from pre- to post-test although this just failed to reach significance (p = 0.0528).

3.2.2 *Vowel Intelligibility*

Figure 2: Boxplots showing vowel intelligibility accuracy (i.e., SSBE speakers' identification of Arabic children's production) at the pre and post-tests.



A linear-mixed effects logistic regression model fit by maximum likelihood, was built for identification data based on the correct/incorrect binomial responses. The best fitting-model indicated that there was a significant effect of time $\chi 2$ (1) = 16.762, p < .001, indicating that participants improved in their vowel production from pre- to post test. The planned contrasts indicated that participants were more intelligible at the post-test, b= 0.447, SE=0.125, z=3.55, p<.001. The model also showed a significant effect of training group, $\chi 2$ (1) = 7.65, p < .01, with the contrast showing that the LV group performed slightly better than the HV group, b= 0.374, SE=0.141, z=2.645, p<.01.

4. DISCUSSION

The current study investigated whether Arabic children improved in their production of SSBE vowels after training, whether any improvements in production generalized to perception, and additionally, whether or not children benefitted differently from HV or LV training. The results

demonstrated that although improvements were small, all children improved in their perception, but that children in the LV condition improved most. Similarly, all children improved in their production. Again, these changes were small, but those in the LV condition improved more than those in the HV condition.

The findings for production are in line with previous studies [4] that have found LV training to be more beneficial for children in learning new articulatory targets. One possibility is that this is because in LV training, learners find it easier to remember how a particular speaker produces a given vowel and are able to use this as the basis for their own production. This might be particularly true for children, who find it harder than adults to adapt to multiple talkers [16]. Further, the children in the current study were tested in a non-immersion setting, where they do not regularly hear native English speakers, and this may have made it still harder to adapt to talker variability.

That being said, all children improved somewhat in production, including those in the HV training condition, and the difference between the groups at post-test was not large. For practical reasons, it was only possible to conduct a relatively small number of training sessions. One possibility then, is that were children to have completed more training sessions, all children would have improved more, but that those in the HV training condition might have improved in their production as much as, or perhaps more than those in the LV condition. That is, in the HV condition, learning may have been slower initially, and 5 training sessions may not have been enough to benefit from talker variability, but with further training, children in the HV condition may have learned as much as or more than those in the LV condition.

Indeed, inspection of the acoustic data showed that all children, regardless of training conditions, were able to generalise their learning in production to new stimuli in the imitation task. This showed that for certain monophthongs, /ɜː, ɑː, ʊ/ and the closing diphthongs /eə, ɪə/, participants in both groups changed their F2 values such that these vowels were produced with more native-like realizations. Although this change was not big enough to show any significant improvement overall, it might indicate that children had started to acquire vowel targets that do not exist in their L1. One possibility is that because none of these vowels exist in Arabic, they found these easier to acquire than those where there is a competing, nearby Arabic vowel. This is in line with the predictions of theories of L2 learning such as the SLM [5] which proposes that the greater the distance between the L2 and L1 category is, the more likely it is that the phonetic differences between the sounds will be detected, and a new phonetic category will eventually be established.

Why was the amount of improvement small? As previously mentioned, children only completed 5 training sessions, and although previous studies have found improvement with this number (e.g., [4]), it is possible that our children of a similar age but with less English experience required more sessions in order to show any greater improvement. Another possibility is that the number of vowels trained affected learning. Previous studies [16] have shown that when it comes to perception for adults, training with a full set of vowels is more effective than learning with a subset. Consequently, we decided to train children with the full vowel inventory, rather than focussing on a smaller number of challenging contrasts. However, it is possible that had we trained children on a subset of vowels, spending more time on more difficult contrasts (e.g., [4]), children would have improved more in their production overall.

Another reason for the small changes in vowel production might have to do with the richness of input. Although children were trained with all 18 vowels, the number of different stimuli was relatively small: 36 imageable keywords, alongside AV recordings, produced by between 1 and 4 SSBE speakers (depending on training condition). It may be that this, coupled with the relatively small number of training sessions, meant that children did not receive enough input to show greater improvement in production.

Lastly, in contrast to previous findings with adults [e.g., 1, 8], all children also improved in perception, but those in the LV group appeared to improve more than those in the HV group. This suggests that phonetic training may not be domain-specific and also that, unlike adults, children may benefit from single-talker perceptual training (cf. [6]), perhaps because as previously discussed, they find it harder to adapt to talker variability and so are less able to benefit from this in training.

5. CONCLUSION

The current study found that LV phonetic training appears to be more beneficial for children's vowel production and perception than HV training, unlike for adults. However, improvements in both production and perception were relatively small. It remains for further research to investigate whether increasing the number of training sessions and the richness of stimuli would aid learning of novel L2 contrasts in children.

6. REFERENCES

[1] Alshangiti, W. 2015. *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.* Unpublished PhD thesis, University College London, UK.

[2] Bradlow AR, Akahane-Yamada RA, Pisoni DB, Tohkura Y. 1999. Training Japanese listeners to identify English /r/and /l/: long-term retention of learning in perception and production. *Perception & Psychophysics 61:977–985*

[3] Evans, B. G., Alshangiti, W. 2018. The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics*, *68*, 15-31.

[4] Evans, B.G., Martin-Alvarez, L. 2016. Age-related differences in second-language learning? A comparison of high and low variability perceptual training for the acquisition of English /i/-/ɪ/ by Spanish adults and children. *Paper presented at New Sounds, Aarhus University, Denmark*.

[5] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

[6] Giannakopoulou, A., Brown, H., Clayards, M., Wonnacott, E. 2017. High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209.

[7] Giannakopoulou, A., Uther, M., Ylinen, S. 2013. Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy* 29 (2) pp. 201 – 218

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

[9] Hawkins, S., & Midgley, J. 2005. Formant frequencies of RP monophthongs in for age groups of speakers. *Journal of the International Phonetic Association*, 35(6), 183-189.

[10] Herd, W., Jongman, A., Sereno, J. 2013. Perceptual and production training of intervocalic/d, ɾ, r/in American English learners of Spanish. *The Journal of the Acoustical Society of America*, *133*(6), 4247-4255.

[11] Iverson, P., Evans, B. G. 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877.

[12] Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., Golestani, N. 2015. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832.

[13] Lively SE, Pisoni DB, Yamada RA, Tohkura Y, Yamada T. 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.

[14] Llompart, M., Reinisch, E. 2018. Imitation in a Second Language Relies on Phonological Categories but Does Not Reflect the Productive Usage of Difficult Sound Contrasts. *Language and Speech*, 0023830918803978.

[15] Logan, J. S., Lively, S. E., Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/: a first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-86.

[16] Magnuson, J. S., Nusbaum, H. C. 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance*, *33*(2), 391.

[17] Nishi, K., Kewley-Port, D. 2008. Nonnative Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order of Training. *Journal of Speech Language and Hearing Research*, *51*(6), 1480.

[18] Shinohara, Y., & Iverson, P. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242–251.