Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Volume 14. Number 1 March 2023 Pp.3- 27
DOI: https://dx.doi.org/10.24093/awej/vol14no1.1

Investigating the Effects of Speaker Variability on Arabic children's Acquisition of English Vowels

Wafaa Alshangiti
English Language Institute, King Abdulaziz University, Jeddah, Saudi Arabia
Corresponding Author: walshangiti@kau.edu.sa

Bronwen G. Evans
Department of Speech, Hearing & Phonetic Sciences, University College London, London, United Kingdom

Mark Wibrow
Publisher Discovery Ltd, Bath, United Kingdom

Received:10/25/2022 Accepted:02/16/2023 Published: 03/24/2023

Abstract
This study investigated whether speaker variability in phonetic training benefits vowel learnability by Arabic learners of English. Perception training using High-Variability stimuli in laboratory studies has been shown to improve both the perception and production of Second Language sounds in adults and children and has become the dominant methodology for investigating issues in Second Language acquisition. Less consideration is given to production training, in which Second Language learners focus on the role of the articulators in producing second language sounds. This study aimed to assess the role of speaker variability by comparing the effect of using High-Variability and Low-Variability stimuli for production training in a classroom setting. Forty-six Arabic children aged 9-12 years were trained on 18 Standard Southern British English vowels in five training sessions over two weeks and were tested before and after training on their vowel production and category discrimination. The results indicate that Low-Variability stimuli may be more beneficial for children, however, High-Variability stimuli may alter some phonetic cues. Furthermore, the results suggest that production training may be used to improve the perception and production of Second Language sounds, but also to inform the design of Second Language pronunciation learning programmes and theories of Second Language acquisition.

Keywords: Arabic Children's acquisition of English, articulatory training, classroom setting of L2 learning, production training, vowel learning, speaker variability
Cite as: Alshangiti, W., Evans, B. G., & Wibrow, M. (2023). Investigating the Effects of Speaker Variability on Arabic children's Acquisition of English Vowels *Arab World English Journal, 14* (1):3-27. DOI: https://dx.doi.org/10.24093/awej/vol14no1.1

Introduction

Learning to perceive and produce the sounds of a second language (L2) is a significant challenge for L2 learners, due to the influence of the mappings between acoustic-phonetic properties of first language (L1) speech and abstract-categories (phonemes) formed during L1 acquisition. Models of L2 speech learning such as the Perceptual Assimilation Model of L2 speech Learning (PAM-L2, Best & Tyler 2007; Tyler, 2019) and the Speech Learning Model (SLM, Flege 1995, 2002; Flege & Bohn, 2021) characterise this influence according to the acoustic-phonetic similarity between L2 sounds and L1 phonemes. SLM classes L2 sounds as being identical, similar, or dissimilar to L1 phonemes: L2 sounds that are similar, but not identical, to an existing L1 phoneme are more challenging to learn (i.e., perceive and/or produce) accurately as they are perceived and produced as the L1 phonemes (e.g., English speakers producing the unaspirated Spanish consonants /p/, /t/, /d/ with the increased aspiration found in English, González Lopéz & Counselman, 2013; Gorba & Cebrian, 2023). PAM-L2 considers contrasts (rather than individual phonemes) and predicts L2 contrasts will be more challenging to learn if both phonemes in the contrast are similar to a single L1 phoneme, as they will be perceived and produced as the L1 phoneme (e.g., the difficulty that L1-Japanese speakers have with L2-English /r/-/l/ contrast due to the perceived similarity of both /r/ and /l/ to the Japanese /ɾ/, Bradlow, Pisoni, Akahane-Yamada & Tohkura., 1997).

The influence of the L1 phonological system on L2 speech learning can be overcome by increased exposure to L2 sounds, which helps a listener focus on more salient cues for L2 sounds or contrasts (Iverson, Hazan & Bannister, 2005; Yuan & Archibald, 2022), resulting in more native like perception and/or production, regardless of whether exposure comes from real-world communication (Flege, Takagi & Mann, 1996; Ingvalson, McClelland & Holt, 2011), formal instruction in the classroom (Camus, 2019), or from phonetic training in the laboratory (Logan, Lively& Pisoni, 1991; Kvasyuk, Putistina, & Savateeva, 2021). In particular, exposure from laboratory-based phonetic training has been shown to produce considerable improvements in perception and production (Thomson, 2018), and is increasingly used to investigate issues in L2 speech learning research, such as (i) the extent to which L1-attuned perceptual systems can adapt to L2 sounds (Hattori & Iverson, 2009), and (ii) how this adaption varies by age and/or experience (Ingvalson, Lansford, Federova, & Fernandez, 2017), (iii) how the organisation of L1 perceptual systems can influence L2 perception and production (Iverson & Evans, 2009), and (iv) the extent to which L2 perception and production are linked (Melnik-Leroy, Turnbull, & Peperkamp, 2021).

The current study aimed to train Arab children, who are learning English as a second language, to pronounce Standard British English vowels giving them articulatory instructions with the aid of computer assisted learning for vowel interface. Given that Training on L2 phonemes, improves learning, and that speaker variability has been mostly used in perceptual training, this study is incorporating speaker variability in production training which has been given less consideration in the literature compared to perceptual training. Moreover, to the authors' knowledge, a few studies, if any, have investigated the role of speaker variability in training Arab children to articulate English vowels using low vs. high speaker variability stimuli with a child-friendly computer assisted learning interface.

The main objective of the current study is to help Arab children learn English vowels production and perception by giving them some articulatory instructions on how to produce these vowels. Another objective of this study is to test whether high or low speaker variability might be

more beneficial for vowel learning with children. An additional objective is to build a teaching tool that can be used in a classroom setting to improve L2 pronunciation skill. The following questions were asked. To what extent does articulatory instructions help children learn English vowels? and what is the effect of speaker variability on learning English vowels by Arab children who are learning English?

Literature Review

Phonetic Training

Phonetic training requires participants to focus explicitly on either their perception of L2 phonemes (*perception training*) or their production of L2 phonemes (*production training*).

Perception Training

Perception training uses a forced-choice task where participants listen to words containing L2 phonemes (*/i:/ as in peel*) and identify the word from a closed set of alternatives (e.g., *peel*, *pill*, *pail*, *pile*). Feedback consists of a binary correct/incorrect visual (or audio-visual) response after every trial and an overall accuracy score presented at the end of each session. Laboratory-based perception training is effective in improving L2 perception in both adults (Iverson & Evans, 2007; Shinohara & Iverson, 2013b) and in children (Shinohara & Iverson, 2013a). In addition, perception-focused training has been found to improve L2 perception, and the improvement proceeds to improvement in phoneme production (e.g., Thomson, 2011; cf. Sakai & Moorman, 2018; Shinohara & Iverson, 2021). However, production-focused training, has been found to improve production (the trained domain), but the improvement in perception (the untrained domain) is not always consistent. Some studies found that production training improves productions and perception (e.g., Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015; Cibelli, 2022), while others found that production training only improves phoneme production, but not perception (e.g., Baese-Berk, 2019; Zhang, Cheng, Qin, & Zhang, 2021).

Production Training

In production training, participants listen to words containing L2 phonemes and repeat them out loud. Feedback focuses on accuracy in pronunciation rather than perception, and can take many forms, including (i) participants' self-correction from listening to recordings of their imitations and comparing them to the productions of native speakers (Hattori, 2010), (ii) examining visual feedback derived from the participants' productions such as spectrograms (Olson, 2014), formant plots (Kartushina & Martin, 2019) or ultrasound images (d'Apolito, et al., 2017), (iii) direct feedback from co-participants accuracy in identifying productions in a classroom setting (Linebaugh & Roche, 2015), and (iv) corrective feedback derived using automatic speech recognition (Neri et al., 2008; Evers & Chen, 2022, see also Cucchiarini, & Strik, 2018, for review).

Production training has been shown to improve the production of L2 sounds (Lopéz & Counselman, 2013; Taimi et al., 2014). For example, Taimi et al., (2014) trained 7–10-year-old L1-Finnish girls on the L2-Swedish /y:/ – /ʉ:/ contrast using listen-and-repeat production training and found that after short training sessions of only two days, participants' productions of the L2-Swedish vowel /ʉ:/ was closer to L1-Swedish productions in the F2 dimension. However, the effects of production training on L2-perception are limited (e.g., Kartushina et al., 2015), and in a study comparing three groups of perception training, production training and a hybrid perception-and-production training focussing on the SSBE vowels /ɪ/ and /iː/, Wong (2013) found that the

production training group did not improve their perception after training. More positive results, however, were reported by Linebaugh and Roche (2015), who used production training with L1-Arabic speakers on the L2-English contrasts /æ/ – /ʌ/, /g/ – /ʤ/ and /ɜ/ – /ɔ/, including visual and verbal instruction regarding the position of vocal tract articulators. After only one training session, participants' accuracy in the perception of the /æ/ – /ʌ/ and /g/ – /ʤ/ contrasts had improved.

Variability in Phonetic Training Materials

A key issue in phonetic training is developing training materials that are optimal for training (Carlet & Cebrian, 2019). In particular, being able to generalize learning to new speakers and new words has been shown to depend on the variability in the training materials, with most perception training research showing an advantage for 'High Variability' (HV) training materials (Lively, Logan & Pisoni, 1993; Kartushina & Martin, 2019), where variability can be introduced by a variety of means: using signal processing to manipulate specific acoustic cues of a phonetic category (Kondaurova & Francis, 2010; Iverson et al., 2005; Cheng, Zhang, Fan & Zhang, 2019), using different phonetic contexts (Strange et al., 2007), presenting multiple versions of the same token produced by multiple speakers (Giannakopoulou et al., 2017), a combination of using multiple phonetic contexts with multiple speakers (Sadakata & McQueen, 2013), or by manipulating specific acoustic cues in tokens that are produced by multiple speakers in multiple contexts (Giannakopoulou, Uther & Ylinen, 2013). Although HV training using multiple speakers has become the dominant approach in laboratory-based approaches to L2 learning (Thomson, 2018), differences in operationalisation of 'high variability' make it difficult to draw conclusions about the advantages of HV across different studies, or to determine what aspect of variability confers an advantage (Kartushina & Martin, 2019; Zhang, Qin, & Zhang, 2021). Some studies found that speaker variability does not affect learners' performance (e.g., Zhang et al., 2021; Wiener, Chan & Ito, 2020). Furthermore, HV training materials may not yield greater benefits for some phonetic categories, transfer to different tasks (Thomson, 2011) or be effective for all age groups (Hwang & Lee, 2015; Giannakopoulou et al., 2017). For example, Giannakopoulou, Brown, Clayards & Wonnacott (2017), investigated whether adults and children would benefit in the same way from HV or Low-Variability (LV) training materials and found that children's perception only improved with LV materials.

Very few studies have focused on the role of variability in training materials using production training. For example, Kartushina and Martin (2019) contrasted HV (five speakers) and LV (one speaker) in L2 production training for the L2-French /e/-/ɛ/ contrast using an imitation task. Participants received visual feedback in the form of a formant plot comparing the first two formants of the participants' vowels compared to those of the target speaker. They found that only LV production training resulted in more accurate vowel productions. These results suggest that children may not benefit from HV phonetic training.

The Current Work

As outlined above, increased exposure to L2 sounds using perception training with HV training materials in the laboratory has been given considerable attention in L2 research, primarily with adults. Less attention has been given to production training and the role variability in training materials may play with this type of training. Furthermore, a few production-training studies have been carried out with children. Finally, most phonetic training has taken place in the laboratory and usually focuses on one or two phonetic contrasts considered difficult for particular L2 learners.

The current work describes the Computer Assisted Learning of Vowels interface (CALVin), software designed to be used for production training in a classroom with children. A group-training programme in the classroom was chosen over traditional laboratory training, to partly overcome the artifice of laboratory training, but also to try and retain the children's engagement with the training over multiple training sessions. Furthermore, it would be valuable to demonstrate if production training based on L2 research can translate into classroom environments, and whether data collected from classroom environments can be of value to L2 researchers (Linebaugh & Roche, 2015), as the use of such training in the classroom is not widespread (Barriuso & Hayes-Harb, 2018), and only a few studies (Wang & Munro, 2004; Ueda & Hashimoto, 2019) have attempted to bridge the gap between laboratory research findings and the techniques used in teaching programs.

The production training in the current study operationalised variability using multiple speakers, using four speakers for High-Variability (HV) and one speaker for Low-Variability (LV). The training focuses on the learning of 18 SSBE vowels as most production training studies have trained only a small number of contrasts (Taimi, Jähi, Alku, & Peltola, 2014; Kartushina et al., 2019), and previous work has shown that perception training with a full set of vowels produces better learning outcomes (Nishi & Kewley-Port, 2007).

Participants in the training study were native Arabic speakers from Saudi Arabia, who speak the Hijazi dialect found mainly in the western region of Saudi Arabia. This dialect has 8 monophthongs: /i:/, /i/, /a:/, /a/, /u:/, /u/, /e:/, /o:/, and two diphthongs /aj/, /aw/ (Jarrah, 1993). Standard southern British English (SSBE), has twenty vowels: /iː/, /ɪ/, /e/, /ɜː/, /æ/, /ɑː/, /ɒ/, /ɔː/, /ʌ/, /uː/, /ʊ/, /eɪ/, /aɪ/, /aʊ/, /əʊ/, /ʊə/, /ə/, /eə/, /ɔɪ/, and /ɪə/ (Wells, 1982), which are more confusable for Arabic learners of SSBE than consonants (Evans and Alshangiti, 2018), and therefore form the focus of this study. The current study aims to investigate the effect of speaker variability in production training on vowel perception and production. Given the difference in the vowel inventory between Arabic and English, it can be assumed that some vowels would fall into new vowel categories which may make learning the L2 vowels easy according to SLM. On this account, learners may assimilate two different vowels into two different categories, which leads to an accurate perception of L2 phonemes. We hypothesised that children might benefit from lower speaker variability and that training would help children learn to produce L2 vowels more accurately, but maybe their vowel perception would not improve.

Methods

The current study used articulatory training with the aid of a child-friendly computer assisted learning interface to investigate the effect of speaker variability on SSBE vowels acquisition by Arab children. The data was collected using a quantitative approach presented by pre- and post-test to measure any possible improvement after the training.

Participants

Forty-six native Arabic-speaking children aged 9-12 years old (median=11, mean age=10.7, SD=0.9) were recruited for the training study and randomly assigned to one of 2 training conditions, LV (one speaker) or HV (four speakers). Participants were recruited from a public girls' school in Jeddah, Saudi Arabia, and had little exposure to English input. They learn English at school from non-native English teachers, and they started learning the English alphabets and basic reading when they were eight years old. Therefore, orthography and identification tasks were

avoided, so they were asked to repeat words, presented with ,child friendly visual aids to help them produce certain vowels, and categorise vowels in an oddity task that does not involve orthography. None of the participants reported any speech, hearing, or language impairments. All participants and their parents gave informed consent in writing, and the study was carried out with the permission of the ethics committee of King Abdulaziz University.

Research Instruments

Training materials

The training materials consisted of SSBE words (see Appendix A) and isolated SSBE vowels. The words were derived from 18 SSBE vowels (/iː/, /ɪ/, /e/, /ɜː/, /æ/, /ɑː/, /ɒ/, /ɔː/, /ʌ/, /uː/, /ʊ/, /eɪ/, /aɪ/, /aʊ/, /əʊ/, /eə/, /ɔɪ/, /ɪə/) embedded in monosyllabic words, selected to represent objects that would be familiar to the children, and informally judged to be imagable (Ellis and Beaton, 1993). In the training software words and vowels were grouped according to vowel clusters shown to be highly confusable for Arabic learners of English (Evans and Alshangiti, 2018): *High/front*: /iː/, /ɪ/, /e/; *Open* /æ/, /ʌ/, /ɒ/; *Central/low-back;* /ɪə/, /eə/, /ɜː/; /ɑː/, /ɔː/; *Back* /ʊ/, /uː/, /aʊ/, /əʊ/; *Diphthongs*: /eɪ/, /aɪ/, /ɔɪ/. Words were arbitrarily assigned as 'keywords' and 'example words'.

Audio and Video Stimuli

All audio and video stimuli were recorded in sound attenuated recording booths at University College London. The audio and video recordings of the keyword and example words were recorded simultaneously. Four native SSBE speakers (2 male, 2 female) recorded each word three times and the best recording (i.e., free of vocal artefacts) was selected for use in training. The words were presented to the speakers via a computer in a random order to avoid list intonation. The audio stimuli were recorded using 16-bit resolution at 44.1 kHz, bandpass filtered using PRAAT (Boersma & Weenink, 2016) between 60 Hz – 20 kHz with 10 Hz smoothing, downsampled to 22,050 Hz, and equalized for amplitude at 70 dB. The resulting audio was also used to replace the lower-quality audio in the video recordings: the replacement audio and video were manually aligned using Lightworks (https://lwks.com/) and the video was cropped for display in CALVin using FFMPEG (https://www.ffmpeg.org/).

The isolated vowels (audio-only) were recorded by one of the SSBE male speakers at the same time as the word list, and post-processed in the same way as the audio for the keyword and example words. In addition, all speakers recorded a short extract from *A Bear Called Paddington* (Bond, 1958). For the HV condition, sentences from each speakers' recording were spliced together to form a single 'multi-speaker' extract used to familiarise participants with the speakers' voices.

Image Stimuli

The images for the keywords, example words, and the mid-sagittal section animation of the isolated vowels, were created by the third author using Inkscape (Harrington and Engelen, 2004; images for the keywords and example words are available from https://github.com/mwibrow/CALVin-images).

Pre-test and Post-Test Stimuli

Recordings of English *hVd*-words, and *bVd*-words (see Appendix B and C) covering containing the 18 vowels used in the training materials were recorded by 4 SSBE speakers (two males, two females). None of these speakers had recorded the training materials, so the pre-test and post-test measured generalization to new stimuli and speakers. The speakers recorded each word three times, and the best recording (i.e., free of vocal artefacts, etc.) was selected.

Research Procedure

Training

Participants were randomly assigned to groups made up of four to five children for each session. This setting was particularly designed to give pronunciation feedback in classroom setting rather than restricting the training to laboratory settings. Participants in both training groups (HV vs. LV) completed five training sessions, a maximum of one session per day, with all training sessions were completed over two weeks. Participants in the HV training condition were trained with four speakers: one per session for the first four sessions and then a mixture of all four speakers in the final session. Participants in the LV training condition were trained using a single speaker for all training sessions.

All training sessions were facilitated by an instructor (the first author), a native Arabic speaker who is fluent in English. All sessions used 'CALVin', training software based on (Alshangiti, 2015) rewritten and adapted for use with children (source code available from https://github.com/mwibrow/CALVin) and was presented via a laptop controlled by the instructor.

Before training, participants were familiarised with the speaker(s) by listening to the Paddington Bear story (single speaker version for LV training; multi-speaker version for HV training) while looking at pictures of the story on slides. Then the instructor explained how opening the jaw, and moving the tongue and lips affect the way different vowels are produced. This part of the training aimed to enable participants to become aware of how their articulators move and how this changes the vowel they produce.

In each training session, participants sat around a table, facing the laptop, which was connected to a high-quality speaker and a microphone. The instructor selected the appropriate speaker, vowel group, and keyword, to ensure that no vowel group or keyword was repeated. Within each vowel group, the instructor selected a keyword and presented it to the participants (Figure one, top-left), followed by the isolated vowel for that keyword, explaining (in Arabic) how to produce the vowel using the vocal tract animation in the software (Figure one, top-right). Then, for each example word (Figure one, bottom left), participants watched video recordings of the speaker producing an example word (Figure one, bottom-right). Participants took turns to record their production of the keywords, isolated vowels, and example words, so they could compare their recordings to those of the native speakers as this "self-perception" (i.e., listening to one's own production) has been argued to help in learning L2 sounds (Baker & Trofimovich, 2006).

These steps were repeated for the other vowels in the vowel group. Each session ended with a review of the vowels covered, led by the instructor, and lasted approximately 45 minutes. Each training session then proceeded in the same way. The training procedure was the same for all training groups, the only difference was the number of speakers, in the HV group (four speakers) and in the LV group (one speaker).

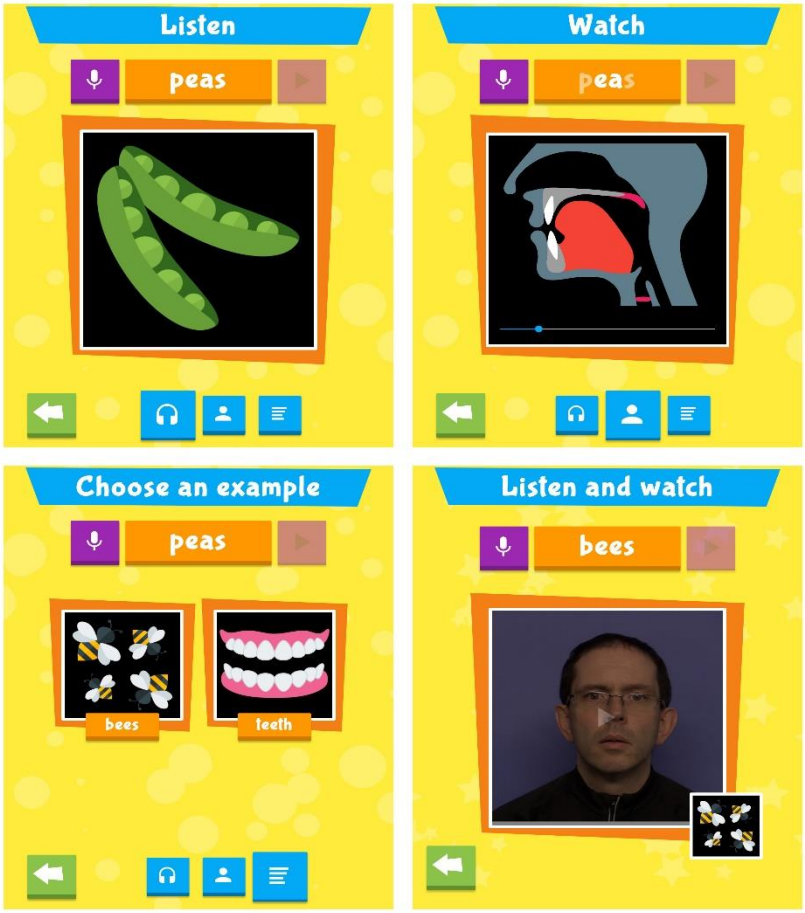

Figure 1. Selected screenshots from CALVin with keywords, the animated vocal tract, example words and native-speaker videos.

Pre- and Post-Tests

Participants completed all pre-test and post-test tasks individually in a quiet room. Participants were familiarised with the tasks by completing a short practice session before the tests, and there was a short break between each task.

Category Discrimination Task

Category discrimination was measured using an oddity task. On each trial, participants heard 3 *bVd*-words where two words were the same and one was different. Participants were asked to judge which one was different, giving their response by clicking on a box labelled 'A', 'B' or 'C', where the first word was 'A', the second as 'B', and the third word as 'C'. They received no feedback and were not able to replay the stimuli. There were 15 pairs of vowels e.g., /ɒ/–/ɔ/ each played six times, three times with /ɒ/, and three times with /ɔ/ as the odd stimulus, with the odd stimulus played first, second or third, giving a total of 90 trials. The task took 30 minutes, with a break after every 30 trials. The vowel pairs were /e/–/ɪ/, /ʊ/- /uː/, /ɒ/-/ɔ/, /ʌ/-/ɔ/, /ɑ/-/ɜː/, /ʌ/-/ɑ/, /ɜː/-/e/, /əʊ/-/aʊ/, /eɪ/-/e/, /eɪ/-/aɪ/, /aɪ/-/ɪ/, /æ/- /ʌ/, /iː/-/ɪ/, /ɪə/-/eə/, /ɔɪ/-/ɔː/.

Word Imitation Task.

Participants listened to recordings of 18 *hVd*-words (produced by one SSBE female speaker), and after each word they waited until they heard a tone, and then repeated the word out loud (the tone was added to avoid children relying on using the phonological loop (Baddeley et al., 1984).

To assess production accuracy, five monolingual native-SSBE speakers (22-46 years old, median 30 years old) were recruited from the University College London Psychology subject pool to carry out a listening task using recordings of the children's *hVd*-word productions as stimuli. To avoid listener fatigue, 10% of the stimuli were played to all listeners and the remaining 90% were split evenly between the listeners. For each stimulus, listeners identified the word from a closed set, and then rate how 'native-like' the stimulus was on a Likert scale from one (poor) to seven (native-like). Each listener identified and rated 460 stimuli and was only able to hear each stimulus once.

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 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 conditions.

Perception task: Category Discrimination

As displayed in Figure two, the training groups performed differently from pre-test to post-test in category discrimination accuracy. To test for potential effects of training, a linear mixed effects model was fit with the score as an outcome variable, with fixed effects of test (pre vs post), group (HV vs LV) and their two-way interaction. The model was fit with a maximal random effects structure, which includes the random slope of the participant. The model indicated that the main effect of the test was significant, $\chi^2(1) = 9.122$, $p<.05$. The planned contrast showed that the performance was better at the post-test, $b= 0.045$, $SE=0.0143$, $z=2.161$, $p<.01$, indicating a change in the category discrimination accuracy after training. The effect of the training group was significant, $\chi^2(1) = 6.5$, $p<.05$. The model also showed that the interaction between the training group and the test is reaching significance, $\chi^2(1) = 3.60$, $p=.05$. The contrast between factors confirmed that the LV group performed slightly better at the post-test than the HV group, $b=-0.1998$, $SE=0.0924$, $z=2.161$, $p<.05$.

To investigate vowel improvement in percentage after training, confusion matrices were built for each group at the pre- and post- test (see tables one to four in Appendix D). As shown in the matrices, for the HV group, there were some improvements in vowels ranging between 3%-7%, but for the vowel /i:/ as in *beat*, there was an 11% improvement in accuracy after training. For the LV group, there was a noticeable (more than 10%) improvement for six vowels after training. There was a 15% improvement for /eə/ as in *bared*, 16% improvement for /æ/, as in *bat*, 14% improvement for /ɜ:/ as in *bert*, 18% improvement for /əʊ/ as in *boat*, 11% improvement for /ɒ/ as in *bot*, and 12% improvement for /ɔɪ/ as in *buoyed*. These percentages confirmed that participants in the LV improved their vowel learning better than those in the HV group. To see if the difference in these six vowels (/æ/, /eə/, /ɜ:/, /əʊ/, /ɒ/, /ɔɪ/) are significant, a linear mixed effect was built with

the vowel pairs that include these six vowels(/æ/- /ʌ/,/ɪə/-/eə/, /eə/-/ɜː/,/ɑː/-/ɜː/, /ʌ/-/ɒ/, /ɑ/-/ɜː/,/ɔɪ/-/əː/əʊ/-/aʊ/). The model was built with test, group and their interaction as fixed factors and random slope of vowel pair and participant as random factors. The effect of test was significant, χ^2 (1) = 8.67, p<.05, at the post test, b=0.0715, SE=0.019, t=3.58, p<.001. There was no significant effect of the training group, but there was a significant interaction between test and group, the LV group performed slightly better at the post-test than the HV group, b= -0.057, SE=0.029, t=-1.978, p<.05.

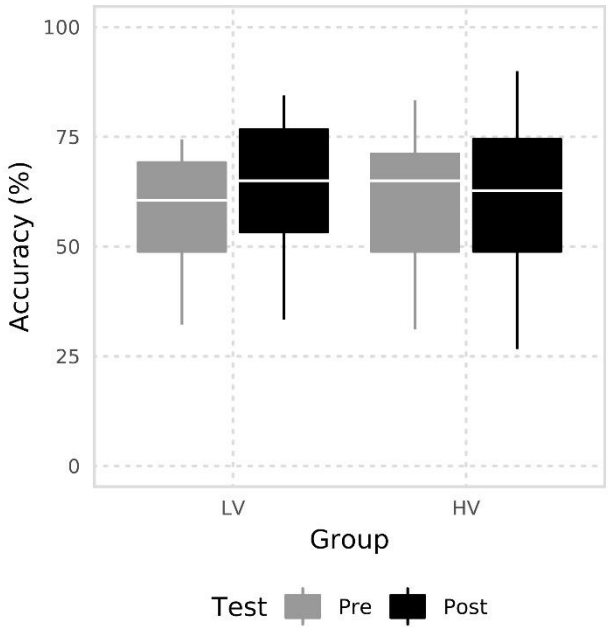

Figure 1. Boxplot to show category discrimination accuracy percent correct at the pre-test (white boxes) and post-test (grey boxes) split by training group.

Production: Word imitation task

Acoustic Analysis

The acoustic measurements were made in PRAAT (Boersma & Weenink, 2016). The stimuli were located manually, and then the first formant (F1) and the second formant (F2) were extracted using hand corrected LPC analyses. Formant frequencies were measured from the midpoint of the vowel (as mentioned above). All F1 and F2 raw values were checked for any value 2 standard deviations outside the range, and these measurements were hand corrected, as necessary. All duration measurements were taken from the beginning of the F2 transitions to the end of the F2 transitions.

Spectral Analysis

To accommodate age related differences in vocal tract size between the children and SSBE speakers, each speaker's formant data were normalised according to Lobanov (1971), using the equivalent formulation described in Flynn and Foulkes (2011), where the normalized formant values were calculated as the z-scores for each formant for each speaker.

Figure three displays the average F1 and F2 vowel measurements produced by the participants at the pre and post-tests. The vowel plot shows some subtle changes from pre-test to post-test; /ɔ/ and /ʊ/ shift closer to the SSBE reference vowels, the /u:/ vowel is more fronted after training especially for the LV group, and the central vowel /ɜ:/ appears to be more centralised at the post-test. A linear mixed model was fit with test (pre, post) and training group (LV, HV) as fixed factors. The random factors were crossed intercepts for participant and stimulus with a random slope for test. There were no significant differences in either F1 or F2, p>.05, confirming that there were no reliable changes in the F1 and F2 values from pre to post-test in either training group.

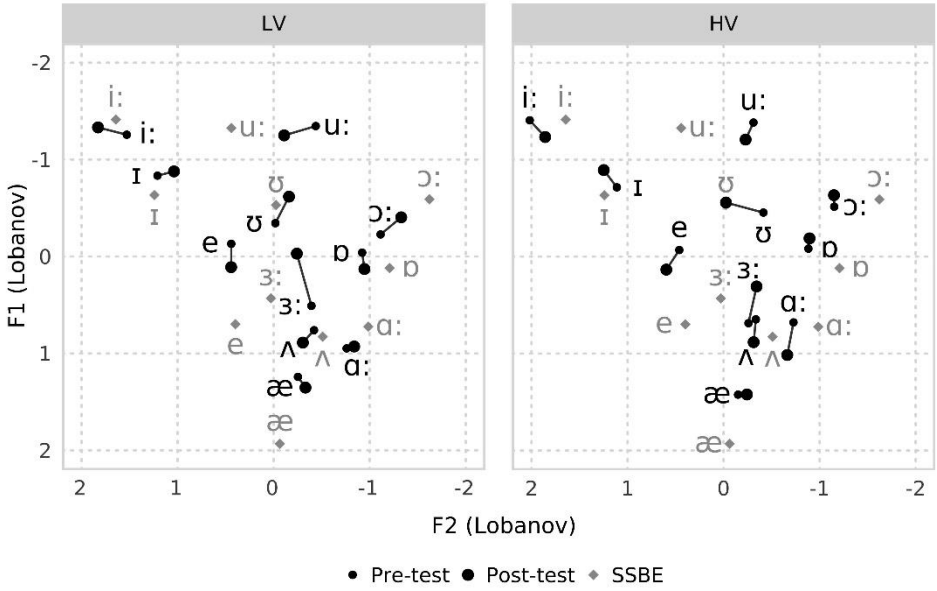

Figure 2. F1/F2 plot of monophthongs produced in *hVd*-words at pre-test (smaller, grey filled circles) and post-test (larger filled black circles). SSBE speakers used in training diamonds are included as reference points. Data from children in the HV condition is displayed in the right-hand panel, and data from children in the LV condition is displayed in the left-hand panel.

Duration

Figure four displays the duration of monophthongs at the pre and post-tests. As shown in the figure, participants produced longer vowels in the post-test. A linear mixed model showed significant effects of test χ^2 (1) = 32.771, p<.05, indicating that there was a change in the overall duration of the vowels from pre-test to post-test. There was also a significant effect of group, χ^2 (1) = 4.185, p<.05 and a significant interaction between test and group, χ^2 (1) = 8.615, p<.05.

Post-hoc analyses comparing the training effects from pre-test to post-test demonstrated that vowel duration was significantly longer at the post-test for the two training groups, p<.001, such that the duration range changed to better match native speakers after training. Participants in the HV group produced longer vowels at the post-test than those in the LV group, b=16.69, SE=6.007, z=2.54, p<.001. However, the difference between groups was small.

To investigate which vowels changed in duration after training, the vowels were divided into two groups: G1(ɪ, e, ʌ, ɒ, ʊ) and G2 (æ, ɑ:, ɜ:, i:, ɔ:, u:), and were analysed separately in comparison to the vowel duration produced by SSBE speakers. For both vowel groups, a separate

linear mixed model was fit with the test, group, and their interactions as fixed factors, and random intercept of vowel and participant as random factors.

For the G1(ɪ, e, ʌ, ɒ, ʊ), The effect of the test was significant, χ^2 (1) = 20.83, p<.05, but the effect of the group was not. However, the interaction between test and group was significant, χ^2 (1) = 12.4, p<.001, in which participants in the HV group produced longer vowels at the post test, b=29.939, SE=8.502, t=3.52, p<.00. This may indicate that the variability aspect helped learners to change their vowel duration to better match that of the native speakers.

For the G2 (æ, ɑː, ɜː, iː, ɔː, uː), There was a significant effect of test, χ^2 (1) = 145.6, p<.001, at the post test, b= 55.408, SE=7.18, t=7.8, p<.001. There was no significant effect of the training group or the interaction between test and group, which indicates that for this vowel group, learners changed their vowel duration after training regardless of their training group.

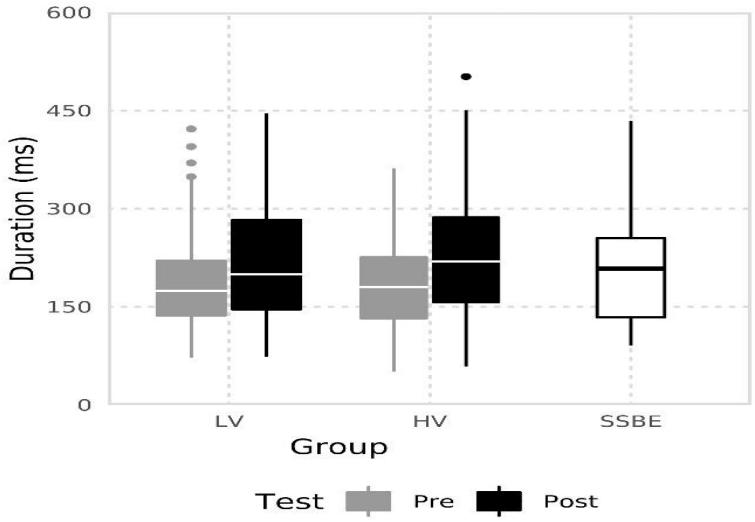

Figure 3. Boxplot to show the duration of monophthongs at the pre-test (grey boxes) and post-test (black boxes) for children in the LV and HV training groups. SSBE speakers are included for reference.

Vowel Intelligibility and Goodness Ratings

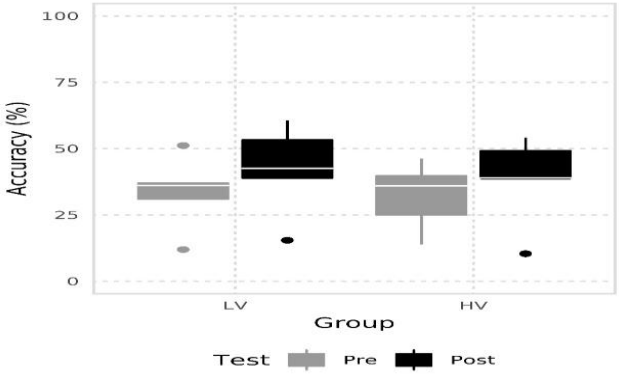

Figure 4. Boxplot to show intelligibility of Arabic children's vowel production accuracy at the pre- white boxes and post-test black boxes.

Due to a large number of stimuli, listeners identified and rated a common 10% of the recordings. To test the level of inter-rater-agreement (i.e., whether the raters used the scale in the same or similar way), a Pearson's correlation was calculated between the 5 raters to see if they were using the scale in the same way. There was no correlation between the raters scores, r=.09, p>.05, indicating a poor consistency in ratings amongst the raters. Therefore, for the following models, the 'rater' was added as a random factor.

Figure five displays vowel intelligibility accuracy at the pre-test and post-test for both training groups. Both groups performed similarly but there appears to be a subtle advantage for LV group in the post-test. A linear mixed model fit by maximum likelihood was built for the identification data based on the correct/incorrect binomial responses, with Test and Group and their interactions as fixed factors and the random intercept of participants and vowel as random factors. The model indicated that there was a significant effect of test χ^2 (1) = 10.97, p<.05, which indicates that participants improved their vowel production from pre-test to post-test. The planned contrasts indicated that participants were more intelligible at the post-test, b= 0.447, SE=0.032, z=3.55, p<.05. The model also showed a significant effect of the training group, χ^2 (1) = 7.65, p<.05, and the orthogonal contrast showed that LV group were slightly more intelligible than the HV group, b= 0.374, SE=0.141, z=2.645, p<.05. However, there was no significant interaction between test and group, p>.05. This model showed an overall accuracy for all the vowels.

To investigate which vowels have improved after training, confusion matrices were created (Appendix D) to show the percent correct of each word. As these matrices show, there was a difference in vowel improvements after training. For the HV group, there were some improvements for most of the vowels averaging from 4%-7% improvements, but some vowels had more than 10% improvements. There was 12% improvement for the vowel /eɪ /as in *hayed*, 25% improvement for /ɜː/ as in *heard*, 16% for /əʊ/ as in *hoed*, and 26% improvement for /ʊ/ as in *hood*. For the LV group, there was a general improvement in most of the vowels ranging from 4% to 10% improvement, but for some vowels the improvement was more than 10%. There was a 22% improvement for /eə/, as in *haired*, 20% improvement for /eɪ/, as in *hayed*, 25% improvement for /iː/, as in *heed*, 15% improvement for /ɔː/, as in *hoard*, and 15% improvement for /əʊ/, as in *hoed*, 14% for /ʊ/, as in hood, 13% for /ʌ/, as in *hud*, and 20% improvement for /uː/, as in *who'd*. Given that participants in both groups improved their vowel production for these 8 vowels (/iː/, /ɜː/, /ɔː/, /ʌ/, /uː/, /ʊ/, /əʊ/, /eə/), a mixed-effect model was built for the accuracy of this vowel group with test, group and their interaction as fixed factors and random intercept of vowel and participant as random factors. There was a significant effect of group, χ^2 (1) = 7.07, p<.05, where the LV group performance was significantly different from that of the HV group, b=0.085, SE=0.033, t=2.52, p<.05. There was also a significant effect of test, χ^2 (2) = 15.6, p<.05, at the post test, b= -0.106, SE=0.034, t=-3.12, p<.05, but there was no significant interaction between group and test.

Goodness Ratings

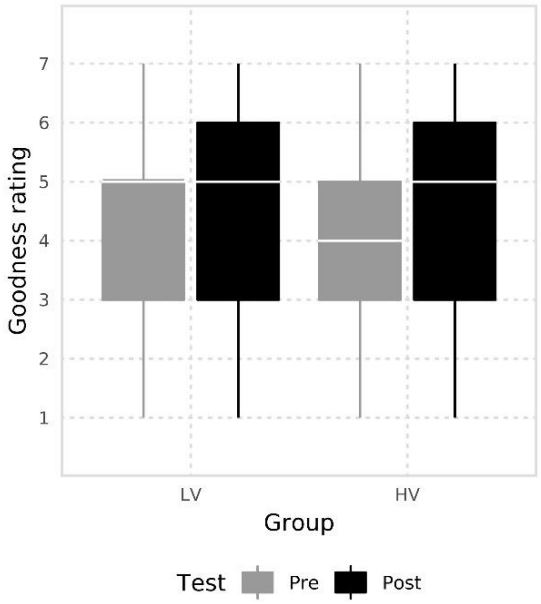

Figure 5. Boxplots of accent ratings for the SSBE listeners. The stimuli were the h-V-d words at the pre-test (light grey) and post-test (black) produced by Arabic children.

Figure six shows the vowel rating produced by Arabic children, the figure shows that there is a slight difference between pre-test and post-test, but no big difference between training groups. A linear mixed effects model was built for the rating scores. The best-fitting model indicated that there was a significant effect of the test, $\chi 2$ (1) = 9.522, $p < .05$, which suggests a difference in performance between the pre-test and the post-test. The contrast showed that the rating at the post-test was slightly higher, b= -0.308, SE=0.095, p<.05. There was no significant effect of group, and there was no significant interaction between group and test, which indicates that participants improved their accent after training regardless of their training group.

In short, the results showed that articulatory effected vowel learning. For the category discrimination and vowel intelligibility, there was an advantage for the LV condition, which indicates that the LV was possibly more beneficial for children who learning a second language. The results from the acoustic analysis showed a small change of the formant values after training, however, participants' vowel duration seemed to improve to get closer to the vowel duration values produced by SSBE speakers. This might indicate that children used their L1 cue, duration, to acquire L2 vowels.

Discussion

The current study investigated the effect of speaker variability in production training on SSBE vowel learning. Arabic children were trained to produce 18 SSBE vowels by receiving articulatory instructions using a computer assisted vowel learning interface (CALVin). The training was either with multiple HV, or a single talker, LV. Their production was assessed by an

objective (acoustic) and a subjective measure (native speakers' ratings), while their perception was measured by a category discrimination task.

To answer the first research question of whether articulatory instructions helped in vowel learning, the training helped the participants to improve some vowels. Inspection of the acoustic data showed that participants made small changes to certain monophthongs, /ɜː, ɑː, ʊ/ and the closing diphthongs /eə, ɪə/ in the F2 dimension, in that these vowels were produced with more native-like realizations. Although this change was not significant overall, it might indicate that children had started to acquire vowel targets that do not exist in their L1. This is possible because none of these vowels exists in Arabic, they found these easier to acquire than those where there is a competing, nearby Arabic vowel, such as /e/. This is in line with the predictions of theories of L2 learning such as the SLM (Flege, 1995) which proposes that the greater the distance between the L2 and L1 categories 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 so small, particularly in terms of changes in acoustic measurements? One possibility is that the task we used to measure production, an imitation task, over-estimated performance at the pre-test. Although we designed our imitation task so that children could not rely on using the phonological loop when repeating the words, it is possible that they were still able to use what they heard to support their own production. However, Cleland, Scobbie, Nakai & Wrench (2015) used a similar imitation and subsequent identification task to assess the production of unfamiliar vowel contrasts at the pre-tests; children were not able to accurately imitate at the pre-test only 5% of sounds identified correctly, but production improved to 32% accuracy at the post-test. Another possibility is that the number of training sessions limited improvement. Participants only completed 5 training sessions, and although previous studies have found improvement both in intelligibility with this number of sessions in children (Evans & Martin-Alvarez, 2016), it is possible that children, in the current study, of a similar age but with less English experience may require more sessions in order to show any greater improvement. The number of vowels trained may also have affected learning. Previous studies have shown that for adults, training with a full set of vowels is more effective than learning with a subset (Nishi & Kewley-Port, 2007). Consequently, we decided to train children with the full vowel inventory, rather than focusing on a smaller number of challenging contrasts. However, some studies that have trained with a small number of contrasts (e.g., Cleland et al., 2015; Evans & Martin-Alvarez, 2016) have shown much larger improvements and so it is possible that if we had trained children on a subset of vowels, children would have shown greater improvement. 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 large: 54 words, alongside AV recordings produced by between one and four SSBE speakers depending on training condition. It may be that, coupled with the relatively small number of training sessions, meant that children did not receive enough training to show a significant change in vowel formant values.

Although the acoustic analysis of vowel production showed no significant change, children were more intelligible after training. This shows that despite the advantage of using objective measures for production accuracy (Delvaux, Huet, Piccaluga & Harmegnies, 2013), providing information about native speakers' perception of vowel quality is also important. Especially since the objective measures showed a slight change and the subjective measures confirmed that change to be significant. The findings for production training are in line with previous studies that showed that children can benefit from production training (cf. Cleland et al., 2015; Taimi et al., 2014).

Though these studies have used different measures (using ultrasound as a visual feedback and listen and repeat method), and therefore any differences in results might have to do with the methods used to assess participants' production.

The second research question was whether speaker variability affects vowel learning. Our results indicated an advantage of the LV group (cf. Evans & Martin-Alvarez, 2016), while both HV and LV groups were able to generalize their learning to the imitation of unfamiliar speakers and words, participants in the LV were more intelligible. This is in contrast to production training studies with adults where only HV training has been shown to generalize to the imitation of a new speaker (Kartushina & Martin, 2019). A possible explanation for LV advantage is that learners might find it easier to remember how a particular speaker produces a given sound and can 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 variations within their own language, due to an increased processing cost (Bent & Atagi, 2015, 2017). Adapting to multiple speakers, even those with a similar accent, may therefore introduce an added processing cost which means that children have fewer cognitive resources for learning a new articulatory target (cf. Antoniou & Wong, 2015). This may have been still harder for our children who were tested in a non-immersion setting, where they do not regularly hear native English speakers.

That being said, the HV training showed some advantage on the phonetic cue level, where children in the HV group improved their vowel duration to better match that of the SSBE speakers more than the LV group. This might indicate that HV training helps children change some cues that are salient for discriminating L2 phonemes, as has been shown with adults (e.g., Iverson et al., 2005). Therefore, we can argue that adaptation to different speakers could have an initial processing cost, but with more training sessions, children may benefit further from exposure to multiple speakers. For practical reasons, it was only possible to conduct a relatively small number of training sessions. One possibility then, is that if children completed more training sessions, all children would have improved more, but those in the HV training condition might have improved in their production as much as, or perhaps more than those in the LV condition.

Unlike our hypothesis, Production Training appeared to improve vowel perception: both HV and LV groups improved in their performance on a category discrimination task. Participants in the LV group improved their perception in a number of vowels (/ɜː/,/ɑ/, /əʊ/, /eə/, /ɔɪ/) more than those in the HV group (tables one to four, in Appendix D). These vowels improved more than the rest of vowels, perhaps because they are not assimilated to any L1 vowel category. Given that this group of vowels do not exist in participants' L1 vowel inventory, which is according to the perceptual assimilation model leads to better phonemic perception (Best, 1995). Participants in the LV group might have used the articulatory information they received during training to learn to perceive this group of vowels accurately.

A previous study with adults using a similar training paradigm (Alshangiti, 2015) found that production training led to improvements in production but not in perception (cf. Hattori, 2010; Baese-Berk, 2019). Based on these studies, one might hypothesize that training is domain-specific and therefore any improvements in production as a result of training, would not lead to improved performance on a category discrimination task. However, in the current study, all children improved in their performance on this task after training. Additionally, those in the LV group appeared to improve more than those in the HV group. However, while improvement was very small for both groups: 5.3% for the LV and 0.8% for the HV group, this result offers a modest

suggestion about a link between the two speech domains, and that production training may improve perception.

Our results from the perception task are in line with results reported by (Kartushina et al., 2015) who trained adult French speakers with no experience of Danish, in their production of two Danish vowel contrasts, and found a subtle change in their perception. Their participants also received five training sessions of similar duration to ours. However, because they were trained on a smaller number of contrasts, this meant that they received approximately one hour of training per vowel. Even so, adults only showed a very small amount of improvement in perception, 4.56%, consistent with other training studies (e.g., Akahane-Yamada et al., 1998; Bradlow et al., 1999) and similar to what we find with a much smaller amount of training per vowel. Kartushina et al. (2015) argue that their participants may have shown reduced training benefits in perception because their category discrimination task contrasted speakers from male and female voices, but in training participants had only been exposed to vowels produced by a speaker of their own gender. They argue that to have been able to succeed at the discrimination task, participants needed to have established abstract, speaker-independent representations for the new vowel contrasts, and that the single-speaker training is not sufficient for learners to be able to do this. Our results suggest that at least for children this may not be the case; LV training may instead mean that they have the cognitive resources needed to start to establish new representations or learn to map the incoming signal to their existing underlying representations (cf. Iverson & Evans, 2009).

Conclusion

This study investigated the effect of speaker variability in training Arab children on producing SSBE vowels. The current work shows that children can benefit from production training. After training, children improved in their production and perception of SSBE vowels, suggesting a modest link between the two speech domains. However, these improvements were small, probably due to the number of vowels covered and the limited number of training sessions. Likewise, it is unclear whether children might benefit from variability in training. Children in the LV condition improved more in terms of intelligibility, and category discrimination, while children in the HV showed some subtle cue shifting after training. While the current study presents some benefits or production training with children using LV training, further research could investigate the benefit of training material variability with fewer vowel contrasts in different learning environments (e.g., contrasting immersion vs. non-immersion settings) and what the implications of this are for L2 learning.

Acknowledgements

This research was supported by King Abdulaziz University grant G-474-270-38. We would like to thank all participants and their families, and 211 primary school in Jeddah for their help.

About the authors

Wafaa Alshangiti is an Assistant Professor at the English Language Institute, King Abdulaziz University. She teaches English and second language acquisition courses. Her research is focussed on second language speech perception and production. https://orcid.org/0000-0002-2808-5857

Bronwen Evans is an Associate Professor in the Department of Speech, Hearing & Phonetics, University College London, where she teaches courses and supervises research students in the

areas of Phonetics and Sociophonetics. Her research combines theory and methods from phonetics and behavioural psychology to investigate adaptation and learning in a second language or dialect. https://orcid.org/0000-0002-8495-2687

Mark Wibrow is a Senior Artificial Intelligence Engineer with Publisher Discovery Ltd, UK. He has a bachelor's degree in Computational Linguistics, a master's degree in Computer Science, and a PhD in Speech, Hearing and Phonetic Sciences from University College London. https://orcid.org/0000-0001-5993-8894

References,

Akahane-Yamada, R., Strange, W., Downs-Pruitt, J., & Masuda, Y. (1998). Modification of L2 vowel production by perception training as evaluated by acoustic analysis and native speakers. *Journal of the Acoustical Society of America*, *103*(5), 3089-3089.

Alshangiti, W. M. M. (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 Doctoral dissertation). University College London, United Kingdom.

Antoniou, M., & Wong, P. C. (2015). Poor phonetic perceivers are affected by cognitive load when resolving speaker variability. *The Journal of the Acoustical Society of America*, *138*(2), 571-574. https://doi.org/10.1121/1.4923362

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233-252.

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*, 981-1005.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231-250.

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *CATESOL Journal*, *30*(1), 177-194.

Bent, T., & Atagi, E. (2015). Children's perception of nonnative-accented sentences in noise and quiet. *The Journal of the Acoustical Society of America*, *138*(6), 3985-3993.

Bent, T., & Atagi, E. (2017). Perception of nonnative-accented sentences by 5-to 8-year-olds and adults: The role of phonological processing skills. *Language and Speech*, *60*(1), 110-122.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, 171-206.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. Lang Exp. *Second Lang. Speech Learn*, *1334*, 1-47.

Best, C. T., MacKain, K. S., & Strange, W. (1982). A cross-language study of categorical perception for semi-vowel and liquid glide contrasts. *The Journal of the Acoustical Society of America*, *71*(S1), S76-S76.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0. 15 (2016).

Bond, M., & Fry, S. (1958). *A bear called Paddington*. London: Collins.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Camus, P. (2019). The effects of explicit pronunciation instruction on the production of second language Spanish voiceless stops: a classroom study. *Instructed Second Language Acquisition*, *3*(1), 81-103.

Carlet, A., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. *A Sound Approach to Language Matters–In Honor of Ocke-Schwen Bohn. Dept. of English, School of Communication & Culture, Aarhus University*, 91-119.

Cibelli, E. (2022). Articulatory and perceptual cues to non-native phoneme perception: Cross-modal training for early learners. *Second Language Research*, *38*(1), 117-147.

Cleland, J., Scobbie, J. M., Nakai, S., & Wrench, A. A. (2015, August). Helping children learn non-native articulations: The implications for ultrasound-based clinical intervention. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), Glasgow, 10-14 August 2015*. International Phonetic Association.

Cucchiarini, C. & Strik, H. (2018). Second Language Learners' Spoken Discourse: Practice and Corrective Feedback Through Automatic Speech Recognition. In I. Management Association (Ed.), *Smart Technologies: Breakthroughs in Research and Practice* (pp. 367-389). IGI Global. https://doi.org/10.4018/978-1-5225-2589-9.ch016.

d'Apolito, S., Sisinni, B., Grimaldi, M., & Fivela, B. G. (2017). Perceptual and ultrasound articulatory training effects on English L2 vowels production by Italian learners. *World Academy of Science, Engineering and Technology International Journal of Cognitive and Language Science*, *11*, 2447-2453.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2013, August). Production training in second language acquisition: a comparison between objective measures and subjective judgments. In *INTERSPEECH, Belgium* (Vol. 2375, p. 2375-2379).

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning*, *43*(4), 559-617.

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.

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. *New Sounds, Aarhus University, Denmark*.

Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, *35*(8), 1869-1889.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, 3-83.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/and/l. *The Journal of the Acoustical Society of America*, *99*(2), 1161-1173.

Flynn, N., & Foulkes, P. (2011, August). Comparing Vowel Formant Normalization Methods. *International Congress for Phonetic Sciences*, 683-686.

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. *Peer Journal*, *5*, e3209. https://doi.org/10.7717/peerj.3209

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), 201-218.

Gorba, C., & Cebrian, J. (2023). The acquisition of L2 voiced stops by English learners of Spanish and Spanish learners of English. *Speech Communication*, *146*, 93-108.

Harrington, B., & Engelen, J. (2004). Inkscape. *Software available at http://www. inkscape. org*.

Hattori, K. (2010). *Perception and production of English/r/-/l/by adult Japanese speakers* (Unpublished Doctoral dissertation). University College London, United Kingdom.

Hattori, K., & Iverson, P. (2009). English/r/-/l/category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*(1), 469-479.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105-120.

Hwang, H., & Lee, H. Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. In *International Congress for Phonetic Sciences*. https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2015/Papers/ICPHS0466.pdf

Ingvalson, E. M., Lansford, K. L., Federova, V., & Fernandez, G. (2017). Listeners' attitudes toward accented speakers uniquely predicts accented speech perception. *The Journal of the Acoustical Society of America*, *141*(3), EL234-EL238.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of phonetics*, *39*(4), 571-584.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842-2854.

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.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Iverson, P. et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

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*(1), 145-160.

Jarrah, M. A. (1993). *The Phonology of Madina Hijazi Arabic: A Non-linear Analysis*. (Unpublished Doctoral dissertation). University of Essex, United Kingdom.

Kartushina, N., & Martin, C. D. (2019). Speaker and acoustic variability in learning to produce nonnative sounds: evidence from articulatory training. *Language Learning*, *69*(1), 71-105.

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.

Kondaurova, M. V., & Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods. *Journal of phonetics*, *38*(4), 569-587.

Kvasyuk, E. N., Putistina, O. V., & Savateeva, O. V. (2021). The use of multimedia language laboratory in teaching English phonetics at the university. In *SHS Web of Conferences* (Vol. 113, 00053). EDP Sciences. https://doi.org/10.1051/shsconf/202111300053

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning*, *9*(1), A1-A17.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and speaker variability in learning new perceptual categories. *The Journal of the acoustical society of America*, *94*(3), 1242-1255.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606-608.

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), 874-886.

López, V. G., & Counselman, D. (2013). L2 acquisition and category formation of Spanish voiceless stops by monolingual English novice learners. In *Proceedings of the 16th Hispanic Linguistics Symposium*, 118-127.

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, *38*(3), 581–605. https://doi.org/10.1177/0267658320988061.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*(5), 393-408.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 20(6),1496-1509. https://doi.org/10.1044/1092-4388(2007/103).

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology*, *18*(3), 173-192.

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, *134*(2), 1324-1335.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224.

Shinohara, Y., & Iverson, P. (2013a). Computer-based English/r/-/l/perceptual training for Japanese children. In *Proceedings of Meetings on Acoustics ICA2013*, *19*, (1). Acoustical Society of America.

Shinohara, Y., & Iverson, P. (2013b). Perceptual training effects on production of English/r/-/l/by Japanese speakers. In J. Przedlacka, J. Maidment., & M. Ashby (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference* (pp. 83-86).

Shinohara, Y., & Iverson, P. (2021). The effect of age on English/r/-/l/perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*, 101108.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, *122*(2), 1111-1129. https://doi.org/10.1121/1.2749716

Taimi, L., Jähi, K., Alku, P., & Peltola, M. S. (2014). Children learning a non-native vowel-The effect of a two-day production training. *Journal of Language Teaching and Research*, *5*(6), 1229-1235, doi:10.4304/jltr.5.6.1229-1235.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744-765.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208-231.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. *A sound approach to language matters–In honor of Ocke-Schwen Bohn*, 607-630.

Ueda, R., & Hashimoto, K. I. (2019). Perceptual Training in a Classroom Setting: Phonemic Category Formation by Japanese EFL Learners. *Pronunciation in Second Language Learning and Teaching Proceedings*, *10*(1), 213-249.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.

Wells, J. C. (1982). Accents of English: The British Isles (Vol. 2). Cambridge University.

Wiener, S., Chan, M. K., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions?. *The Modern Language Journal*, *104*(1), 152-168.

Wilson, I., Gick, B., O'Brien, M. G., Shea, C., & Archibald, J. (2006). Ultrasound technology and second language acquisition research. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 148-152). Somerville, MA: Cascadilla Proceedings Project.

Wong, J. W. S. (2013). The effects of perceptual and/or productive training on the perception and production of English vowels/ɪ/and/iː/by Cantonese ESL learners. *Interspeech*, 2113-2117.

Yuan, Q., & Archibald, J. (2022). Modified Input Training and Cue Reweighting in Second Language Vowel Perception. *Frontiers in Educational Research*, *5*(6),65-75, DOI: 10.25236/FER.2022.050613.

Zhang, X., Cheng, B., Qin, D., & Zhang, Y. (2021). Is speaker variability a critical component of effective phonetic training for nonnative speech?. *Journal of Phonetics*, *87*, 101071. https://doi.org/10.1016/j.wocn.2021.10107

Appendices

Appendix A
Word list

ball, bears, bees, bike, book, boys, cake, card, cat, clock, coat, coin, cot, cow, cup, cut, food, foot, gate, gears, hat, house, kite, knife, knot, leg, lid, loaf, mat, men, mouse, nurse, nuts, park, paws, pears, piers, pin, ring, road, room, seed, shark, shirt, skirt, spade, squares, suit, sword, tears, teeth, ten, toys, wood.

Appendix B
hVd-words

heed, hid, head, heard, had, hard, hod, hoard, who'd, hood, hud, hayed, hide, how'd, hoed, haired, hoyed, hear.

Appendix C
bVd-words

The bVd-words include some additional words in /bVt/ and /pVt/ contexts as some vowels do not form real words in the /bVd/ context: beat, bit, bet, bait, bite, bart, bat, bot, but, bird, bought, bout, boat, bared, beard, buoyed, bet, booed, poot, put, port, pout, beard.

Appendix D

Table 1. *A confusion matrix for the HV group at the pre-test showing the percent correct of the category discrimination task, the matrix shows the expected and the actual response.*

Expected/ Actual response	bait	bared	bart	bat	beard	beat	bert	bet	bird	bit	bite	board	boat	bot	bout	buoyed	but	poot	put
bait	70							12			18								
bared		45			20				35										
bart			61				11							13			15		
bat				47													53		
beard		41			59														
beat						68				32									
bert			39				61												
bet	14							73		13									
bird		56							44										
bit						10		12		70	9								
bite	21									14	64								
board												70				30			
boat													68		32				
bot			27											52			21		
bout													26		74				
buoyed												39				61			
but			23	21										14			42		
poot																		61	39
put																		23	77

Table 2. *A confusion matrix for the HV group at the post-test showing the percent correct of the category discrimination task, the matrix shows the expected and the actual response.*

Expected/ Actual response	bait	bared	bart	bat	beard	beat	bert	bet	bird	bit	bite	board	boat	bot	bout	buoyed	but	poot	put
bait	71							14			14								
bared		51			20				30										
bart			66				11							12			12		
bat				53													47		
beard		38			62														
beat						79				21									
bert			36				64												
bet	16							65		19									
bird		68							32										
bit						8		16		67	10								
bite	22									11	67								
board												71				29			
boat													71		29				
bot			20											55			24		
bout													38		62				
buoyed												32				68			
but			23	22										13			42		
poot																		68	32
put																		35	65

Table *3. A confusion matrix for the LV group at the pre-test showing the percent correct of the category discrimination task, the matrix shows the expected and the actual response*

Expected/ Actual response	bait	bared	bart	bat	beard	beat	bert	bet	bird	bit	bite	board	boat	bot	bout	buoyed	but	poot	put
bait	74							14			12								
bared		34			26				40										
bart			57				16							11			16		
bat				44													56		
beard		36			64														
beat						78				22									
bert			49				51												
bet	18							65		17									
bird		64							36										
bit						9		15		65	11								
bite	17									17	66								
board												67				33			
boat													63		38				
bot			24											49			27		
bout													28		72				
buoyed												44				56			

ISSN: 2229-9327

but			24	24										13			39		
poot																		64	36
put																		43	57

Table 4. *A confusion matrix for the LV group at the post-test showing the percent correct of the category discrimination task, the matrix shows the expected and the actual response.*

Expected/ Actual response	bait	bared	bart	bat	beard	beat	bert	bet	bird	bit	bite	board	boat	bot	bout	buoyed	but	poot	put
bait	69							19			13								
bared		49			16				35										
bart			60				15							13			12		
bat				60													40		
beard		31			69														
beat						74				26									
bert			35				65												
bet	9							69		22									
bird		61							39										
bit						9		12		69	10								
bite	17									10	72								
board												71				29			
boat													81		19				
bot			16											60			24		
bout													29		71				
buoyed												32				68			
but			21	25										11			44		
poot																		60	40
put																		39	61