The effect of HVP training in vowel perception on bilingual speech production

Authors:

Dr Jayanthiny Kangatharan, University of Winchester, Sparkford Road, SO22 4NR

jayanthinykangatharan@gmail.com

Dr Anastasia Giannakopoulou, University of Bedfordshire, Vicarage Strey, Luton, LU1 3JU

Anastasia.Giannakopoulou@beds.ac.uk

Professor Maria Uther, University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1 LY

M.Uther@wlv.ac.uk

Abstract

Prior investigations (Giannakopoulou et al., 2013) have indicated high variability phonetic training intervention can help L2 English adult learners change the perception of vowels such that they shift their attention to primary cues (spectral features) rather than secondary cues (e.g. duration) to correctly identify vowels in L2. This experiment explores if high-variability training impacts on L2 adult learners' production of L2 speech. Production samples from a prior experiment were used to conduct ratings of accuracy (Giannakopoulou, 2012). In the current experiment, the production samples were transcribed and rated for accuracy by twenty native English listeners. The intelligibility levels of L2 learners' speech samples as indexed by higher accuracy in transcription were observed as having been rated higher following training than prior to training. The implications of the results are considered with regard to theories on the connection between speech production and perception, and Flege's (1995) Speech Learning Model.

Keywords

Perceptual training, language acquisition, intelligibility, speech production, bilingualism

Introduction

In the domain of L2 spoken language acquisition non-native speakers can struggle with non-native phonetic contrasts. It has been, for instance, observed that monolingual Japanese speakers have problems with the discrimination of the English /r/-/l/ contrast (Goto, 1971). Similarly, monolingual Arabic speakers show difficulty discriminating the /b/-/p/ contrast (Flege and Port, 1981). There is evidence that high-variability phonetic training (HVPT) can be used to help non-native learners perceive non-native categories more accurately. For example, several studies showed that Japanese learners of English learned the difference between /r/ and /l/ phonetic categories successfully using HVPT. As a result, they were able to use this perceptual learning to understand novel speakers and novel tokens (Lively, Logan and Pisoni, 1993; Logan et al., 1991; Yamada, 1993).

HVPT is a unique form of computer-assisted pronunciation training (CAPT) (Thomson, 2011) where more than one voice is used to generate multiple target sounds of different words that are presented in multiple phonetic contexts, and L2 learners need to select the label that represents the sound they perceived. They then obtain information on response accuracy in the form of instant feedback. The HVPT has been used in prior research to enhance listeners' phonemic identification (Pisoni et al., 1994; Uther et al., 2007) and its use in teaching L2 learners with enhanced perception of L2 English vowels (e.g. Nishi and Kewley-Port, 2007a, 2008) and consonants (Bradlow et al., 1997) has been well documented.

HVPT can be contrasted to training that makes use of low variability input (LV), research on which first laid the groundwork for devising high-variability training approaches (Bradlow, 2008). In LV training approaches, trainees are presented with stimuli that are produced by only one talker (e.g. Strange and Dittman, 1984). One of the first studies to use a low-variability approach had Japanese speakers undergoing training on the English [r]–[l] contrast by raising learners' sensitivity to small differences between the speech sounds (Strange and Dittman, 1984). The training evaluated the extent to which the trained stimuli and training task could be learnt and it assessed the extent to which there can be generalizability to novel stimuli and novel task. It was found that although Japanese trainees' responses to synthetic stimuli could be modified with discrimination training, there was no generalization to words that were naturally produced (Strange and Dittman, 1984).

A later study that trained Japanese listeners on the same contrast using the low variability training approach aimed to find out whether presenting exaggerated exemplars that can be discriminated at the beginning of the training could lead to a generalization of exaggerated exemplars to less exaggerated exemplars. As a result, it was hypothesized that when the acoustic distance between training stimuli is gradually reduced, listeners could eventually discriminate between natural exemplars (McCandliss et al., 2002). The study found that the training group did indeed reveal enhanced native like identification and discrimination after training. However, no significant differences existed between the trained and untrained groups during testing on a new continuum of the contrast.

Previous studies that made use of both HV and LV input in second language phonetic training, showed that although both HV and LV groups consisting of Cantonese learners of English revealed improvement after being trained on English vowel contrasts for ten sessions, the HV group did outperform the LV group, with the HV group also generalizing more on identification tasks (Wong, 2012). Similarly, a training study that trained Dutch learners of Japanese on a consonant contrast for five sessions provided evidence for HV training resulting in better learning and generalization in an identification task compared to training with LV input (Sadakata and McQueen, 2013).

These findings on the differential effects of HV and LV training have been replicated by more recent research. For example, a study that trained adult Spanish learners on a French vowel contrast via a production training that included either HV or LV talker input, showed after three training sessions that though accuracy in production did enhance for both HV and LV conditions, only in the HV condition was there a generalization to stimuli presented by a novel speaker (Kartushina & Martin, 2019). It was also found that compared to the LV condition, the HV condition revealed more stable speech productions.

It has been suggested that high-variability phonetic training not only has a positive effect on the ability to distinguish non-native phonemic contrasts, but it can also positively affect speech production to render more towards native-like production (Bradlow et al., 1997). For example, L1 Japanese speakers' perception training that led to improved identification of English /r/-/l/ also resulted in the improved production of English /r/-/l/ following training because production samples were of higher intelligibility than prior to training

(Bradlow et al., 1997). Prior investigations have shown L2 speakers' accuracy in pronunciation was able to influence native English listeners' general level of intelligibility of L2 speech (Purcell and Suter, 1980; Varonis and Gass, 1982). It can therefore be said that perceptual learning is able to assist L2 speakers in gaining a non-native perceptual contrast. A transfer of perceptual learning by L2 speakers onto their L2 speech production can clearly occur. Thereby we can understand how L2 speakers' process of learning non-native phonetic contrasts can influence their regulation over how that contrast is generated (Bradlow et al., 1997).

Bradlow et al. (1999) were the first to demonstrate enhanced perception and production of /r/-/l/ minimal pairs. Here Japanese trainees' production was evaluated by among others presenting American English listeners with word stimuli in minimal-pair identification and open-set transcription tasks. Since then, subsequent perceptual training studies that used different languages were able to show significant improvements in production after perceptual training (Lambacher et al., 2002; Wang et al., 2003; Thomson, 2011; Huensch and Tremblay, 2015). For example, Wang et al. (2003) had American native speakers undergoing training to perceive Mandarin tones to find out if learnt contrasts could be perceptually transferred to production. Participants produced a set of Mandarin words prior to and following training that native Mandarin listeners assessed in an identification task. Results showed significant tone production enhancement following perceptual training, with improved production extending to novel stimuli (Wang et al. 2003).

Recently, connections between speech perception and speech production were also reported by studies that trained Cantonese learners on English vowel contrasts (Wong, 2014, 2015b) and that used contrasts set in wider settings of discourse (Huensch, 2016). Similarly, more recent research, in which Argentinian L2 learners of English underwent perceptual training in the acquisition of word-initial voiceless stops, revealed improved performance in the productions of /p/ and /t/ from pre-test to post-test (Alves and Luchini, 2017).

One study trained 22 Mandarin learners to recognise ten Canadian English vowels that was generated by 21 native speakers (Thomson, 2011). Target vowels were presented in monosyllabic frames by means of a HVPT technique to provide L2 learners with multiple speakers and contexts in which vowels were learnt. Learners indicated their response through the selection of one of the ten salient pictures that represented each vowel category. Learners were provided with both visual and auditory feedback. It was found that training significantly enhanced learners' vowel intelligibility in both trained and untrained contexts (Thomson, 2011). Similarly, a more recent study explored the impact of perceptual training on Korean L2 learners' perception and production of English syllable structures reported perceptual training to improve production and generalizability to new word stimuli and speakers in production and perception thereby pointing out that the production and perception systems in L2 are connected (Huensch and Tremblay, 2013).

Similar as in the case of Thomson's (2011) training study the perceptual training study by Giannakopoulou and colleagues (2013) intended to discover if L2 listeners' attention can

be turned to information, particularly cues important for performing accurate perceptual identification so that during their identification and discrimination between L2 phonetic segments, they can accurately weight the cues (Giannakopoulou et al., 2013). Specifically, L2 learners underwent training with natural and modified duration stimuli with a minimal pair list on the tense-lax /i:/-/I/ vowel contrast. Similar to prior research that yielded an enhanced perceptual effect in Japanese learners of L2 English with regard to the /r-l/ contrast (Nishi and Kewley-Port, 2007a), the training protocol in Giannakopoulou et al. (2013) consisted of Greek learners of L2 English being trained on a smaller number of word pairs and being exposed to a higher number of word pairs during testing to investigate if learning can be generalised. Specifically, the pre- and post-training tests used trained and untrained words articulated by an unfamiliar speaker. Only the Greek speakers took part in the training, with ten participants undergoing training on natural duration stimuli, while ten other participants underwent training on modified duration stimuli (Giannakopoulou et al., 2013). The comparison between Greek and English speakers on how they use duration and spectral information was the focus of the pre-training test session. By contrast, the post-training test session was conducted to find out if training significantly influenced cue weighting and vowel perception.

Pre-training test findings indicated English native speakers applied duration cues as secondary cues, and made use of spectral cues as primary cues. However, Greek learners of L2 English made use of duration cues as primary cues, which is indicated by their decreased performance for the tasks in which spectral information determined identification and discrimination performance (Giannakopoulou et al., 2013). Within the

condition in which training made use of natural vowel stimuli, post-training findings showed Greek L2 speakers of English to weight spectral cues as secondary, while weighting duration cues as primary. By contrast, within the condition, in which training made use of modified duration stimuli, performance was enhanced for the modified duration stimuli tasks as well as natural duration stimuli in the post-training test.

Overall, the result that high-variability perceptual training can assist in the re-learning of perceptual weighing of cues implies that HVPT seems to represent a stable technique to enhance results in perceptual identification, and in the acquisition of speech sound categories within a second language. It was also demonstrated that learning was generalised in that L2 learners did generalise to untrained words (Giannakopoulou et al., 2013), a trend also observed by Thomson (2011) as discussed earlier. The present study adds to the existing literature exploring whether native Greek speakers of L2 who underwent HVPT not only showed a difference in perception (Giannakopoulou, 2012; Giannakopoulou et al., 2013), but also whether this resulted in differences in production (specifically increased intelligibility) when producing the trained L2 vowels.

Thus, the aim of this experiment was to find out if perceptual training enhances production as evaluated through an orthographic transcription task that was completed by native English listeners. This would help assess the extent to which the Greek adult trainees' production changes through HVPT, specifically in enhancing overall word intelligibility. Based on prior investigation (e.g. Bradlow et al., 1999), the hypothesis was posited that speech as generated by the Greek adult trainees will be more intelligible after perceptual training than before. Based on previous research the current study included a word

condition (e.g. Bradlow et al., 1997, 1999), and also included a sentence condition to find out if the effect of HVPT on words does also extend to words produced within a sentence environment.

Method

Design

A within-subjects design was used for the orthographic transcription task. The independent variables are represented by time (pre- versus post-training), item type (word or sentence type), and speaker (speaker 1-8). The dependent variable includes intelligibility, the measurement of which occurred via the task of orthographic transcription.

Participants

Eight native Greek learners of L2 English participated in the speech recording task. All participants were female, aged 20-30 years (mean age 27.4) and acquired their education in English as L2 that ranged between 8-9 years at school in Greece, and had spent not more than 2 weeks in an English-speaking environment (see Giannakopoulou et al., 2013 for further details on the non-native participants). Native speakers' samples were taken from two monolingual native English speakers (2 female) aged 20 (mean age 20) to compare native listeners' rating performance on both non-native and native samples. They were undergraduate students recruited from university and who received course credits for their participation.

A separate set of 20 listeners was recruited to rate the non-native and native speech samples, who were native speakers of English (16 female, 4 male) (mean age 19) from the Southeast England region. Each of them declared to have regular hearing and orthographic abilities. They were undergraduate students who were recruited from university. They received course credits for their participation.

Materials and Apparatus

Recording Materials with native and non-native Speakers

Materials in the current study consisted of minimal pair words and sentences. Words, for instance, were 'sit' and 'seat'. Sentences, for instance, were 'Please take a seat' and 'Sit down please' (see the Appendix for more information).

The materials of the current study were produced by a group of native English speakers and a group of native Greek learners of L2 English. Specifically, in the current study native English speakers' samples were taken from two monolingual native English speakers using a digital voice recorder (VN-712PC Olympus). The monolingual English speakers were recorded reading out aloud the list of words and sentences in a clear voice.

The speech stimuli produced by native Greek speakers as learners of L2 English were taken from a previous training study (see Giannakopoulou et al. 2013) who read out in a clear voice the same list of words and sentences before and after undergoing HVPT. They were recorded using a digital voice recorder (VN-712PC Olympus) with a microphone (Andoer Microphone Sidande Mic-01) attached.

Training material in Giannakopoulou et al. (2013)

The training stimuli were generated via 2 male and 2 female speakers in what is deemed typical Southern British English. They were presented in two styles: with a normal and a modified vowel duration. The training stimuli included nineteen minimal pairs of words.

Rating Materials with native Speakers

The minimal pair words and sentences formed the samples that were rated by the native English listeners in the present experiment. The display of stimuli occurred on a laptop with the e-prime software installed (Schneider et al. 2002a, b) via headphones (Sennheiser HD429) at a comfortable listening volume. Each participant's responses, which were submitted using the computer keyboard, were documented via the e-run software application.

Procedure

Training procedure in Giannakopoulou et al. (2013)

The HVP training included ten 30-minute sessions over a two-week duration. Participants were presented with minimal pairs, with each word of the pair being presented in an auditory and a visual mode. Every training session consisted of 304 trials that were delivered randomly. There was a four-time presentation of each minimal word pair. The visual stimuli stayed on the screen until the participant responded with a mouse click on the selection of words on the screen (see Giannakopoulou et al. 2013).

After each trial, participants received feedback and had the opportunity to listen to the auditory stimulus. Participants were also given the opportunity to correct themselves in 'Correction' trials that were presented when they gave incorrect responses. Feedback on correct and incorrect responses was delivered via happy and sad cartoon animations respectively. In addition, every correct answer led to the presentation of a coin, with the amounts of coins giving information on the number of correctly answered trials. There was a training of ten participants on the natural duration stimuli, while ten other participants underwent training on the modified duration stimuli (see Giannakopoulou et al. 2013).

Recording Procedure with native and non-native Speakers

The speech samples were produced by eight native Greek learners of L2 English who read out in a clear and loud voice the list of words and sentences (see Materials for a description of the stimuli contained in the list) prior to participating in the pre-test and HVPT sessions. The same procedure was repeated after participants had completed all perceptual training sessions and the post-test as reported in Giannakopoulou et al. (2013). Two recordings were produced by each participant (before and after perceptual training). Although in Giannakopoulou et al. (2013) more participants completed the perceptual training protocol, recordings for production data was only available for 8 participants in the current study. All participants completed a short demographic and linguistic questionnaire. This study was ethically approved by the Ethics committee at the Psychology Department at university. All participants filled in a consent form prior to participating and were debriefed following participation.

The native speech samples were produced by two monolingual English speakers who read out in a clear and loud voice the same list of words and sentences as the non-native Greek speakers. The native speech samples were collected to compare native listeners' rating performance on both non-native and native samples. The two monolingual English speakers completed a short demographic and linguistic questionnaire. They filled in consent form prior to participating and were debriefed following participation. This study was ethically approved by the Ethics committee at the Psychology Department at university.

Rating Procedure with native Speakers

In the rating task, twenty native English listeners completed a transcription and a confidence rating task. They carefully listened to every word stimulus and typed out using the keyboard within the space that was displayed on the screen what they heard. The orthographic transcription task represented a measurement of intelligibility of speakers' speech (Giolas and Epstein, 1963; Tikofsky and Tikofsky, 1964; Yorkston and Beukelman, 1981; Garcia and Cannito, 1996; Hustad, 2008). Participants then indicated on a Likert scale from 0-6 how confident they were in how accurate their transcription was (0=very confident; 6=not very confident at all). The presentation of a novel word stimulus occurred 500 milliseconds following listeners' rating of their confidence in their transcription. An arrow that appeared for 200 milliseconds indicated the presentation of the next word. The software e-prime was used to conduct the session. There was a random presentation of stimuli.

This study was ethically approved by the Ethics committee at the Psychology Department at university. All participants filled in a consent form prior to participating and were debriefed following participation.

Data Analysis

To analyze the transcription of the words and sentences (see supplementary materials for full list of words and sentences), a three-way within-subjects ANOVA (time, speaker, item) was made use of. The words and sentences were generated prior to and following the HVPT by L1 Greek learners of L2 English. Comparable to previous investigations (Bradlow and Bent, 2002; Bradlow and Alexander, 2007; Smiljanić and Bradlow, 2011; Munro and Derwing, 1999; Lane, 1963), no word candidates by native speakers were accepted in this transcription task where there was an incomplete match of the word in the stimulus utterance. When transcribed words were identified totally correctly, they were accepted. With this approach there was no ambiguity about whether near hits occurred because of typos or the presented word was not really identified. Thus, transcription data were analyzed on a binary basis, with words that are correct without ambiguity receiving the highest score of 1, while those that were unambiguously incorrect received the lowest score of 0. Responses were then averaged across words and sentences for each rater.

Results

Transcription

The boxplots below compare native listeners' transcription of the two native English speakers and the eight non-native speakers of English. The plots show that at both word level (see Figure 1) and sentence level (see Figure 2), native English speakers on the whole

were perceived as more intelligible than the non-native speakers of English, with apparent more pronounced differences between native and non-native speakers at word level.

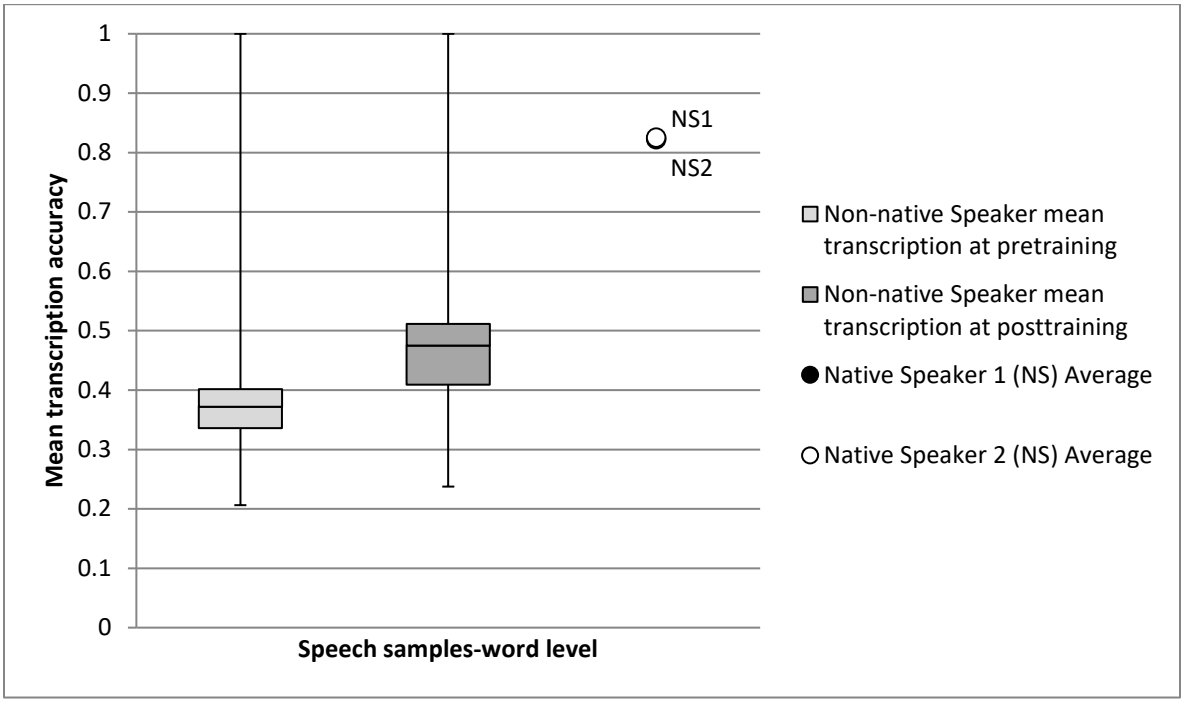

Figure 1: Average marker of the two native speaker' speech samples (NS1 and NS2) and boxplots of the non-native speakers' speech samples as transcribed by native listeners at word level.

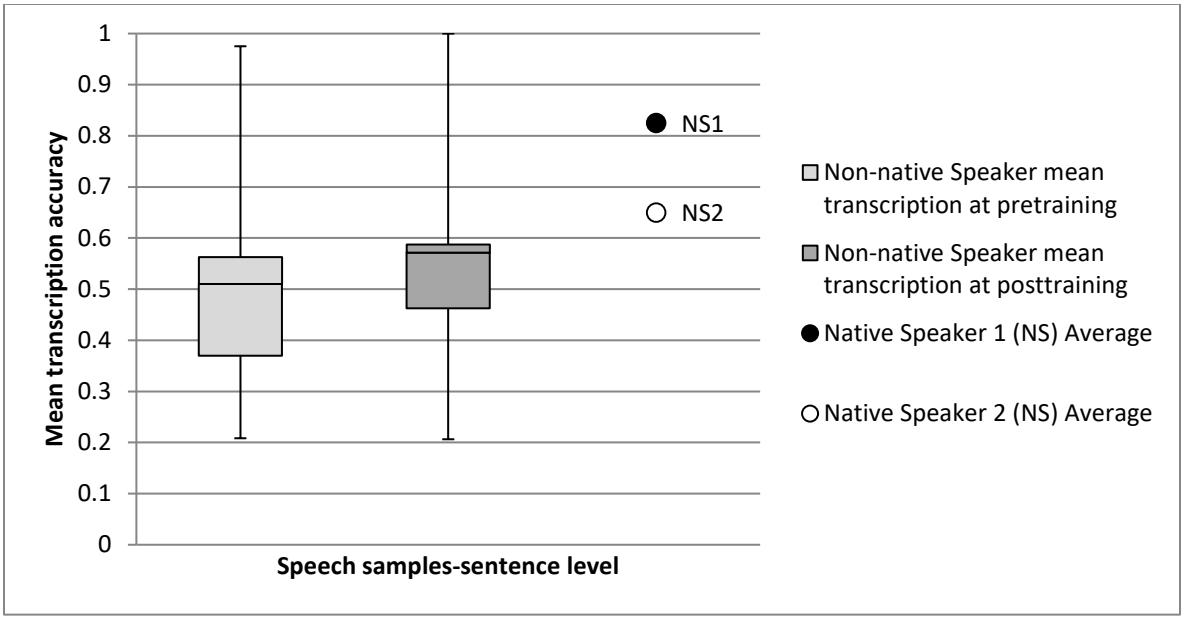

Figure 2: Average marker of the two native speaker' speech samples (NS1 and NS2) and boxplots of the non-native speakers' speech samples as transcribed by native listeners at sentence level.

The data showed higher accuracy in the transcription of samples produced at word level (F (1, 19) =11.081; $p < .05$; $\eta^2p = .368$; see Figure 3) and sentence level (F (1, 19) =12.679; $p < .05$; $\eta^2p = .400$; see Figure 4) following HVPT compared to baseline accuracy. This suggests that there is improved intelligibility after high-variability phonetic training.

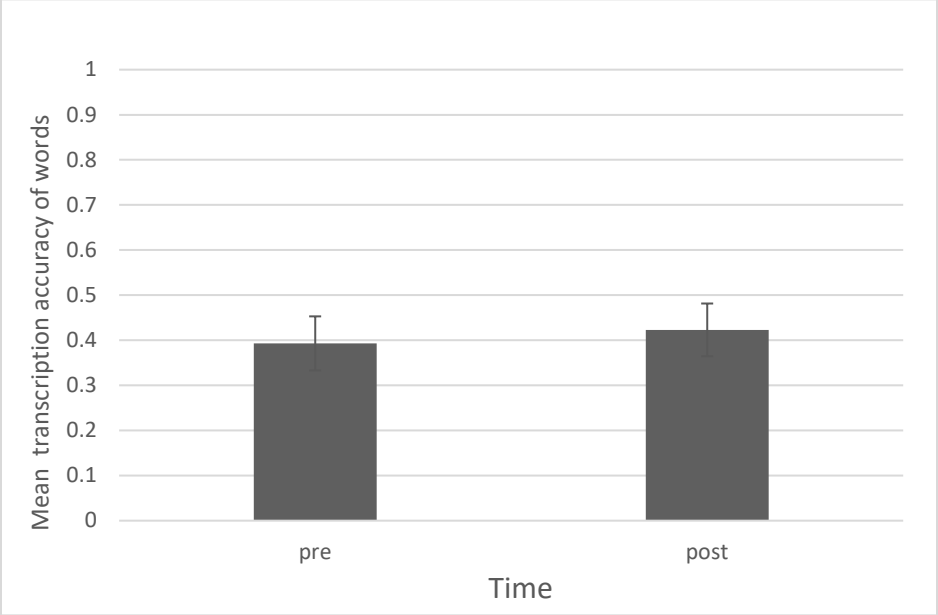

Figure 3: Native speakers' transcription accuracy of Greek speakers' speech samples at word level. Error bars show +/- 1 standard error from the mean.

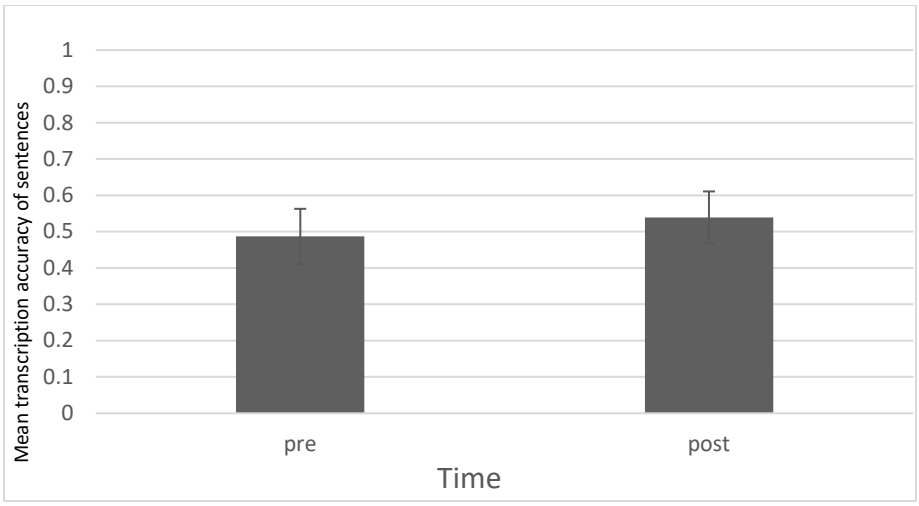

Figure 4: Native speakers' transcription accuracy of Greek speakers' speech samples at sentence level. Error bars show +/- 1 standard error from the mean.

There was an interaction between time and speaker at word level (F (7, 133) = 4.379; p < .05; η^2p = .609; see Figure 5) and sentence level (F (7, 133) = 2.589; p < .05; η^2p = .120; see Figure 6), which indicates intelligibility improved for some speakers from pretest to posttest.

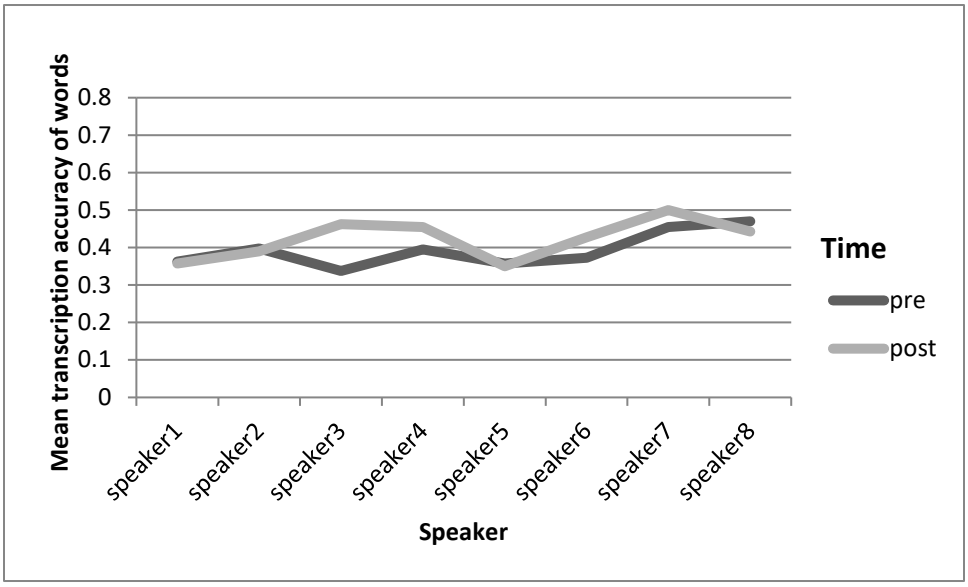

Figure 5: Native speakers' transcription accuracy of Greek speakers' speech samples in interaction with time (pre; post) at word level.

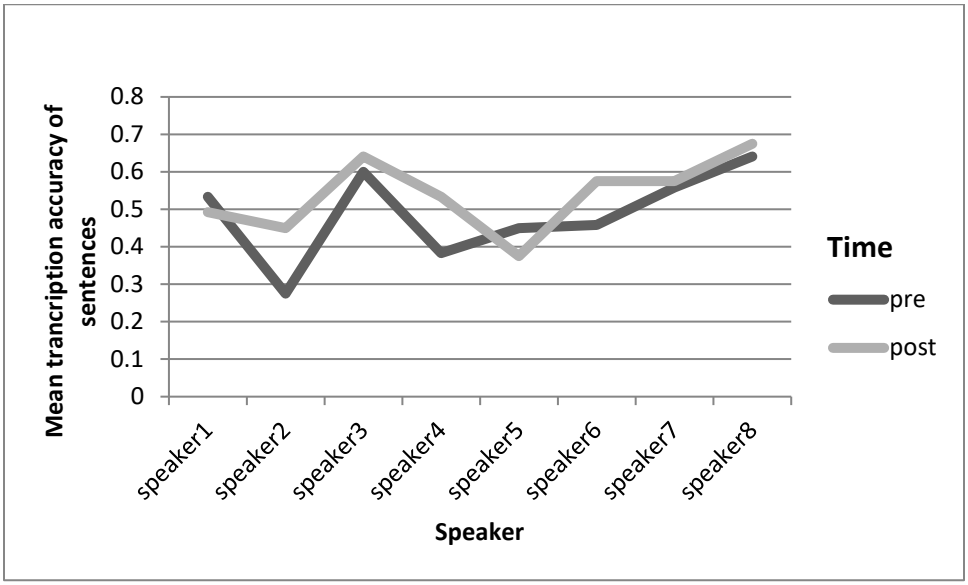

Figure 6: Native speakers' transcription accuracy of Greek speakers' speech samples in interaction with time (pre; post) at sentence level.

Discussion

The goal of the study was to find out if improvements in vowel quality identification and discrimination as shown by Greek adult learners of L2 English in Giannakopoulou et al. (2013) after perceptual training does indeed help the same L2 learners produce more intelligible speech in L2. Therefore the speech samples that had been produced by Greek adult learners of L2 English at word and sentence level prior to and following high-variability perceptual training were rated by twenty native speakers of English for intelligibility. Based on previous research (e.g. Bradlow et al., 1997, 1999), the present study used a word condition as well as sentence condition to find out if the effect of HVPT on words does extend to words embedded in a sentence context.

As expected, the finding that post-training accuracy levels improved after training compared to baseline accuracy confirms that vowel intelligibility can be improved through high-variability training at word and sentence level for L2 learners' not only perception but also is arguably transferred to production of L2 speech. This result appears to confirm the benefit of using stimuli from a multiplicity of speakers and phonetic contexts, and of targeted training. In particular, high-variability phonetic training, in which specific cues are absent, and one attends to important ones, could be considered to potentially assist L2 learners in perceptually reorganising cues, so that cues, which initially were regarded as secondary cues, are turned into primary ones (Lively, Logan & Pisoni, 1993; Logan et al., 1991; Yamada, 1993). This could convert the perceptual advantage of improving identification and discrimination of non-native speech contrasts into practical use during L2 speech production (Bradlow et al., 1997, 1999).

By suggesting that high-variability perceptual identification training can positively impact on L2 learners' perception and production of L2 speech, this study could be regarded to support prior research that showed perceptual improvement that followed high-variability training perceptual improvement could be moved to L2 learners' production of L2 speech (e.g. Lambacher et al., 2005; Lengeris, 2008; Huensch, 2016; Wong, 2014; 2015b). This is also in line with studies that have demonstrated the same effect on L2 learners' perception and production of L2 speech using perceptual training methods (Logan & Pruitt, 1995; Ylinen et al., 2010). This outcome is in line with the conclusions of a recent meta-analysis that reviewed literature of the past 25 years to understand the question whether perception

training can enhance the production of phonemes in L2 (Sakai & Moorman, 2018). Specifically, it was shown that if perception training was strictly regulated, it resulted in medium-sized enhancements in speech perception (d = 0.92, SD = 0.96) and small enhancements in speech production (d = 0.54, SD = 0.45).

In Giannakopoulou et al.'s study (2013) the HVPT was of perceptual nature and Greek adult learners did not receive explicit training in speech production compared to prior investigations (e.g. Kartushina et al., 2015; Kartushina & Martin, 2019) that trained L2 learners on L2 vowels with production training to investigate the link between production and perception. The finding of improved intelligibility in L2 speech production in the current study therefore can be considered to have occurred due to acquired knowledge from the perceptual domain to the production domain, especially as pre-training test results were at chance level for the L2 learner group as reported in Giannakopoulou et al. (2013). The finding of perceptual learning being transferred onto speech production could be explained by the possibility that during high-variability training L2 speakers might have undergone a perceptual adjustment to the gestural features of the English /i:/ and /I/ distinction (Fowler, 1986), which is not present in their L1 (Evans & Martín-Alvarez, 2016; Lively, Logan & Pisoni, 1993; Ylinen et al., 2010). Specifically, based on the information they received from the auditory mode, it can be suggested that the Greek L2 adult learners did change their L2 /i:/ and /I/ phonetic categories (Fowler, 1986), as a result of perceptual training (Grenon, Kubota, & Sheppard, 2019).

According to this direct-realist approach, L2 speakers' /i:/ and /I/ phonetic categories might have been more accurately defined after the high-variability training, which then must have directed their L2 speech production at post-test (Fowler, 1986). This view suggests that the L2 speakers' modifications to their phonetic categories can be seen in both speech perception and production. Support for this view comes from a recent study (Evans & Martín-Alvarez, 2016) that trained Spanish speakers on the English /i/-/ɪ/ contrast, and that suggested that perceptual training could be sufficient to at least partly enhance the production of phonemes.

Compared to the direct-realist approach, the motor theory assumes that speech production and speech perception share a collective representation in form of a phonetic module that mediates both processes of speech perception and speech production (Liberman and Mattingly, 1989). In contrast to the direct-realist approach, the motor theory therefore posits that speech perception and speech production are more closely connected. By this view, the observed enhancement in L2 speech production by adult L2 learners could be theorised to be the result of adult L2 learners modifying their articulatory commands during the perceptual training. According to this view, the subsequent alterations to the adult L2 learners' internal phonetic representation that is shared by the domains of perception and production could be considered to explain the observed effect of perceptual learning onto enhanced speech production (Liberman & Mattingly, 1989). Accordingly, the motor system is considered to be used for the process of speech perception (Galantucci, Fowler & Goldstein, 2009). This would also be in line with findings from imitation and shadowing (that is direct repetition without requesting imitation) experiments that indicated a dialogue

between speech perception and speech production (Goldinger, 1998; Goldinger et al., 2000). Accordingly, phonetic properties that one perceives can influence phonetic properties that one produces on a quite brief time-scale.

Although the present data do not conclusively support the direct realist approach or the motor theory, the result that adult L2 learners' perceptual learning improved their intelligibility in L2 speech production without any direct training in speech production, indicates that a mental representation exists that is shared by the perceptual domain and the production domain, and that integrates the processes of speech perception and production. In that regard, the present data are in line with the direct-realist approach and the motor theory.

Though in the present study individual differences between L2 speakers in producing English as L2 that is perceived as intelligible were evident, the observation of increased intelligibility in L2 learners' speech as result of high-variability training could be considered to support the Speech Learning Model (SLM) by Flege (1995). This model argues that the amount of experience L2 learners have with L2 can positively influence L2 learners' pronunciation accuracy. In Giannakopoulou (2012)'s training study L2 learners were exposed to increased quantity and quality of phonetic experience in L2 English in form of stimuli from a multiplicity of speakers and phonetic contexts during HVPT, which in the present study led to an improved performance in intelligibility at word and sentence level by Greek L2 adult learners at post-training test. The observed higher intelligibility in L2 learners' speech after training can therefore be considered to provide support for SLM.

Moreover, according to SLM, L2 speakers can improve in L2 speech production through perceptual training when the auditory-acoustic phonetic space utilised for both speech perception and production is reorganised. The present data with regard to intelligibility support the SLM by showing that in adult second-language learning shifts in perception could be transferred to changes in production.

The original study by Giannakopoulou et al (2013) included in total 20 participants, of which ten participated in a perceptual training condition that used natural duration stimuli, and ten took part in a perceptual training condition that used modified duration stimuli. It should be noted that, for the goal of the present experiment, the ten adult participants who had participated in the natural duration condition were asked to record samples before and after HVPT training as the natural duration condition exposed them to what is a 'natural' sounding input they had during perceptual training. However, due to incomplete data recordings, there were only eight non-native speakers whose samples were evaluated for intelligibility in the present experiment. The small amount can be argued to limit the generalization of the current finding.

Moreover, one needs to point out that the present experiment did not include a control group of native Greek learners of L2 English who did not undergo HVPT or who underwent a different type of training. Therefore in contrast to prior research, (e.g. Bradlow et al., 1999), the current study did not compare the performance of the non-native speakers' speech production to the speech production of a control group whose L2 speech production

was assessed by native English listeners. Such a comparison would have been useful to fully confirm the effectiveness of HVPT for the intelligibility of non-native speakers' speech production. It could therefore be argued that the improvements in intelligibility as shown by the native Greek learners of L2 English in the current experiment could have come from testing at pre-test and post-test level and not from HVPT training.

The outcomes of the current experiment also lead one to ask if the approach of high variability provides a larger advantage when it comes to creating a long-term improvement in speech production compared to different techniques such as low-variability training interventions (e.g. Strange and Dittman, 1984). The main advantage of the high-variability procedure regarding perceptual learning is that it leads to learning that is highly generalized. This approach can be considered as specifically valuable when it comes to enhancing speech production since speech production involves knowing how phonetic categories can differ in various segmental and prosodic contexts. Future research could aim to understand how low-variability training might assist non-native learners' acquisition of non-native phonetic categories, and how its effects might differ from high-variability approaches when it comes to producing intelligible speech.

Future research could therefore make efforts to understand to which extent variability in the input actually represents an advantage for adult second-language learners in acquiring non-native speech sounds (e.g. Sinkevičiūte et al., 2019). One source of concern here is that since the existing literature in the field of psychology is affected by publication bias, where non-significant results are not reported or might not be included in journals

(Rosenthal, 1979), it is unlikely that phonetic training studies that will be conducted and might not find any benefits of variability on generalization will be published. It also has to be noted that compared to what was included in past phonetic training studies (e.g. Evans & Martin-Alvarez, 2016) in terms of sample size, to find significant differences between HV and LV training, future phonetic training studies would need to use considerably large sample sizes (Dong et al., 2019).

Future research could therefore make use of a higher number of L2 learners who could participate in the high-variability perceptual identification study (e.g. Dong et al., 2019), and a higher number of participants per condition (Clopper and Pisoni, 2004; Perrachione et al., 2011), include a larger set of vowels in training rather than selected vowel contrasts (Hwang and Lee, 2015; Nishi & Kewley-Port, 2008), and make use of a longer training period (Strange and Dittmann, 1984), and longer training sessions (Aliaga-Garcia, 2020). Future research could use samples in the perceptual rating study that were recorded in a natural setting with experimental control over confounding variables.

What future investigations also could do is carrying out accent ratings to evaluate the strength of L2 learners' accents before and after undergoing high-variability training. According to SLM, it is the high use of L1 that creates the stronger accent in L2 speech in L2 learners than in low-L1-use L2 learners (Flege et al., 1997). This can occur due to both L1 and L2 phonetic elements being co-activated when L2 learners produce L2. Similar to previous research (Bradlow et al., 1997), it would be of relevance to explore if improved perception and improved production did occur in parallel in each adult L2 learner. This

would allow one to see whether the processes of speech perception and production occurred at the same rate in each adult L2 learner. Thus, future research could compare the extent of improvement in each adult L2 learner's speech perception and production and find out if those with the most enhanced performance in perceptual learning do also improve the most in speech production.

One could speculate and find similar to prior research (Bradlow et al., 1997) that the processes of speech production and perception might not proceed at the same rate in each adult L2 learner. This would be in support of previous evidence that showed that though speech production and perception are intimately connected, learning within the area of production does not necessarily follow learning in the perceptual domain (Goto, 1971; Sheldon and Strange, 1982; Yamada et al., 1994), thus suggesting that the processes of learning underlying in these areas seem different in individual adult L2 learners.

In conclusion, following on the previous study in which high-variability training was found to positively influence Greek L2 adult learners' vowel quality identification and discrimination, the present study appears to provide evidence in support of the speech learning model (SLM) by Flege, and for the usefulness of the high-variability training in improvement of L2 production.

Specifically, this study appears to suggest that by providing increased quantity and quality of phonetic experience in the form of stimuli from a multiplicity of speakers and phonetic contexts, the high-variability approach can lead to enhancing L2 learners' production of L2 speech in terms of intelligibility. However, more research is needed with a larger sample

size and the inclusion of a control group to fully substantiate the effectiveness of HVPT for enhancing non-native speakers' extent of intelligibility in their speech production in English as a second language.

Word count: 5925 words

Declaration of interest statement

The authors declare that they have no conflict of interest.

References

Aliaga-Garcia, C. (2020). "High-variability Phonetic Training and training Set Size considerations: Are L2 learners able to learn more than five L2 vowel categories at a time?", 35-36. Editor: Linda Taschenberger. (2020, March 29). Book of Abstracts: 2nd Workshop on Speech Perception and Production across the Lifespan (SPPL2020). Zenodo. http://doi.org/10.5281/zenodo.3732383.

Alves, U. and Luchini, P. (2017). "Effects of perceptual training on the identification and production of word-initial voiceless stops by Argentinian learners of English." *A Journal of English Language, 70* (3): 15-32. http://dx.doi.org/10.5007/2175-8026.2017v70n3p15.

Bradlow, A. (2008). L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition*, 1394-1424. Philadelphia: John Benjamins.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., & Tokhura, Y. (1997). "Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production." *Journal of the Acoustical Society of America, 101* (4): 2299-2310. https://doi.org/10.1121/1.418276.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. https://doi.org/10.3758/BF03206911.

Bradlow, A. R., & Alexander, J. (2007). "Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners." *Journal of Acoustical Society of America, 121* (4): 2339–49. https://doi.org/10.1121/1.1487837 112, 272–284.

Bradlow, A. R., & Bent, T. (2002). "The clear speech effect for non-native listeners." *Journal of Acoustical Society of America, 112* (1): 272–284. https://doi.org/10.1121/1.1487837.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of Talker Variability on Perceptual Learning of Dialects. *Language and Speech*, *47*, 207–238. https://doi.org/10.1177/00238309040470030101.

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, 7, e7191.

Evans, B. G., & Martín-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. *Proceedings of New Sounds: 8th International Conference on Second Language Speech*. Aarhus, Denmark.

Garcia J, & Cannito M. (1996). "Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria." *Journal of Speech and Hearing Research* 39 (4): 750–760. https://doi.org/10.1044/jshr.3904.750.

Flege, J. E. (1995). "Second-language speech learning: Theory, findings, and problems." In S*peech perception and linguistic experience: Theoretical and methodological issues, edited by* W. Strange, 229-273. Timonium, MD: York Press.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). "Amount of native-language (L1) use affects the pronunciation of an L2." *Journal of Phonetics* 25 (2): 169-186. https://doi.org/10.1006/jpho.1996.0040.

Galantucci, B., Fowler, C. & Goldstein, L. (2009). Perceptuomotor compatibility effects. *Attention, Perception, & Psychophysics*, *71*, 1138-1149. https:// https://*doi*.org/10.3758/APP.71.5.1138.

Giannakopoulou, A, (2012). "*Plasticity in second language (L2) learning: perception of L2 phonemes by native Greek speakers of English*." PhD dissertation, Brunel University. Retrieved from http://bura.brunel.ac.uk/bitstream/2438/6592/1/FullTextThesis.pdf.

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. https://doi.org/10.1177/0265659012467473.

Giolas T, & Epstein A. (1963). “Comparative intelligibility of word lists and continuous discourse”. *Journal of Speech and Hearing Research* 6 (4): 349–358. https://doi.org/10.1044/jshr.0604.349.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-278. 10.1037/0033-295X.105.2.251.

Goldinger, S., Cutler, A., McQueen, J. M., & Zondervan, R. (2000). The role of perceptual episodes in lexical processing. *Paper presented at the SWAP Spoken Word Access Processes*.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, *72*, 17–34. https://doi.org/10.1016/j.wocn.2018.10.005.

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. https://doi.org/10.1016/j.wocn.2015.06.007.

Huensch, A. (2016). Perceptual phonetic training improves production in larger discourse contexts. *Journal of Second Language Pronunciation*, *2*, 183–207. https://doi.org/10.1075/jslp.2.2.03hue.

Hustad, K. C. (2008). "The relationship between listener comprehension and intelligibility scores for speakers with dysarthria." *Journal of Speech, Language & Hearing Research, 51* (3) 562-573. 10.1044/1092-4388(2008/040).

Hwang, H., & Lee, H.-Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. *Proceedings of the 18th International Congress of Phonetic Sciences.* Retrieved from https://pdfs.semanticscholar.org/b1de/abb7eb1385c1731895f1a4a7fc5a9d144ce3.pdf.

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. https://doi.org/10.1121/1.4926561.

Kartushina, N., & Martin, C. D. (2019). Talker and Acoustic Variability in Learning to Produce Nonnative Sounds: Evidence from Articulatory Training. *Language Learning*, *69*, 71–105. https://doi.org/10.1111/lang.12315.

Lambacher, S.G., Martens, W.L., Kakehi, K., Marasinghe, C.A. & Molholt, G. (2005). "The effects of identification training on the identification and production of American

English vowels by native speakers of Japanese." *Applied Psycholinguistics,* 26 (2): 227-247. https://doi.org/10.1017/S0142716405050150.

Lambacher, S., Martens, W., & Kakeki, K., (2002). The influence of identification training on identification and production of The American English mid and low vowels by native talkers of Japanese. *Paper presented at the 7th International Conference on Spoken Language Processing*, Denver, August, 245–248.

Lane, H. (1963). "Foreign accent and speech distortion". *Journal of the Acoustical Society of America,* 35 (4): 451-453. http://dx.doi.org/10.1121/1.1918501.

Lengeris, A. (2008). The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. *Poster presented at the 2nd ISCA Workshop on Experimental Linguistics*, ExLing, Athens, August.

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 talker variability in learning new perceptual categories." *Journal of the Acoustical Society of America,* 94 (3): 1242-1255. 10.1111/j.1467-7687.2011.01118.x.

Logan, J. S., & Pruitt, J. S. (1995). "Methodological issues in training listeners to perceive non-native phonemes." In *Speech perception and linguistic experience: Theoretical and methodological issues*, edited by W. Strange, 351-378. Timonium, MD: York Press.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 89–108. https://*doi*.org/10.3758/CABN.2.2.89.

Munro, M. J., & Derwing, T. M. (1999). "Foreign accent comprehensibility and intelligibility in the speech of second language learners". *Language Learning* 49 (1): 285-310. https://doi.org/10.1111/0023-8333.49.s1.8.

Nishi, K. & Kewley-Port, D. (2007a). "Second language vowel perception training: Effects of set size, training order, and native language". *Paper presented at the 16th International Congress of Phonetic Sciences*, Saarbrucken, August, 1621-1624.

Nishi, K. & Kewley-Port, D. (2008). "Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training." *Journal of Speech, Language, and Hearing Research, 51* (6): 1480-1493. doi.org/10.1044/1092-4388(2008/07-0109).

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*, 461–472. doi.org/10.1121/1.3593366.

Pisoni, D. B., Lively, S. E. & Logan, J. S. (1994). "Perceptual learning of non-native speech contrasts: Implications for theories of speech perception". In *Development of speech*

perception: The transition from speech sounds to spoken words, edited by J. Goodman and H. C. Nusbaum, 121–166. Cambridge, MA: MIT Press.

Purcell, E. & Suter, R. (1980). Predictors of pronunciation accuracy: *A reexamination. Language Learning*, *30*, 271-287. doi: 10.1111/j.1467-1770.1976.tb00275.x.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033- 2909.86.3.638.

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. https://doi.org/10.1121/1.4812767.

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. https://doi.org/10.1017/S0142716417000418.

Schneider, W., Eschman, A. and Zuccolotto, A. (2002a). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A. and Zuccolotto, A. (2002b). *E-Prime Reference Guide.* Pittsburgh: Psychology Software Tools Inc.

Sinkevičiūtė, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in*

Second Language Acquisition, 1–26. http://doi.org/10.1017/S0272263119000263.

Smiljanić, R., & Bradlow, A. (2011). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness" *Journal of Acoustical Society of America, 130* (6): 4020–4032. 10.1121/1.3652882.

Strange, W., & Dittmann, S. (1984). "Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English." *Perception and Psychophysics, 36* (2): 131–145. http://dx.doi.org/10.3758/BF03202673.

Thomson, R. I. (2011). "Computer assisted pronunciation training: targeting second language vowel perception improves pronunciation." *Computer Assisted Language Instruction Consortium Journal, 28* (3) 744-765. 10.11139/cj.28.3.744-765.

Tikofsky RS, & Tikofsky RP. (1964). "Intelligibility as a measure of dysarthric speech." *Journal of Speech and Hearing Research, 7*: 25–333. 10.1044/jshr.0704.325.

Uther, M., Uther, J., Athanasopoulos, P., Singh, P. & Akahane-Yamada, R. (2007). "Mobile Adaptive CALL (MAC): A lightweight speech-based intervention for mobile language learners." *Paper presented at the Interspeech 2007 Conference for the International Speech Communication Association*, Antwerp, August. 2329-2332.

Varonis, E., & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition*, *4*, 114-136. 10.1017/s0272263112000812.

Wang, Y., Jongman, A., & Sereno, J.A. (2003). "Acoustic and Perceptual evaluation of Mandarin tone productions before and after perceptual training." *Journal of the Acoustical Society of America, 113* (2) 1033–1043. https://doi.org/10.1121/1.1531176.

Wong, J. W. S. (2012). Training the Perception and Production of English /e/ and /ae/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *14th Australasian International Conference on Speech Science and Technology*, (December), 37–40.

Wong, J. W. S. (2014). The Effects of High and Low Variability Phonetic Training on the Perception and Production of English Vowels / e / - / æ / by Cantonese ESL Learners with High and Low L2 Proficiency Levels. *Proceedings of the 15th Annual Conference of the International Speech Communication Association*, 524–528. Retrieved from https://repository.hkbu.edu.hk/hkbu_staff_publication/ 6234.

Wong, J. W. S. (2015b). The Effects High-Variability Phonetic Training on Cantonese ESL Learners' Production of English Vowel Contrasts - *An Acoustic Analysis. Phonetics Teaching and Learning Conference*, 107–111. Retrieved from https://repository.hkbu.edu.hk/hkbu_staff_publication/6235/.

Yamada, R., Strange, W., Magnuson, J. Pruitt, J., & Clarke, W. (1994). "The intelligibility of Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English." Paper presented at the *International*

Conference of Spoken Language Processing for the Acoustical Society of Japan, Yokohama, September 2023-2026.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R. &Näätänen, R. (2010). "Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English." *Journal of Cognitive Neuroscience, 22* (6): 1319-1332. 10.1162/jocn.2009.21272.

Yorkston. K., & Beukelman, D. (1981). *Assessment of Intelligibility of Dysarthric Speech*. C.C. Publications; Tigard, OR.

Appendices: Supplementary Materials

List of words and sentences

Blip Bitter	Bleep Beater
Itch Fit Fist	Each Feet Feast
Lick Litter	Leak Litre
Sit Slip Ship	Seat Sleep Sheep

- He eats salad and is very <u>fit</u>.

- He has smelly <u>feet</u>.
- Please, take a <u>seat</u>.
- <u>Sit</u> down, please.
- He likes feeding the <u>sheep</u>.
- I love my <u>sleep</u> and cannot get up in the morning.

Figures

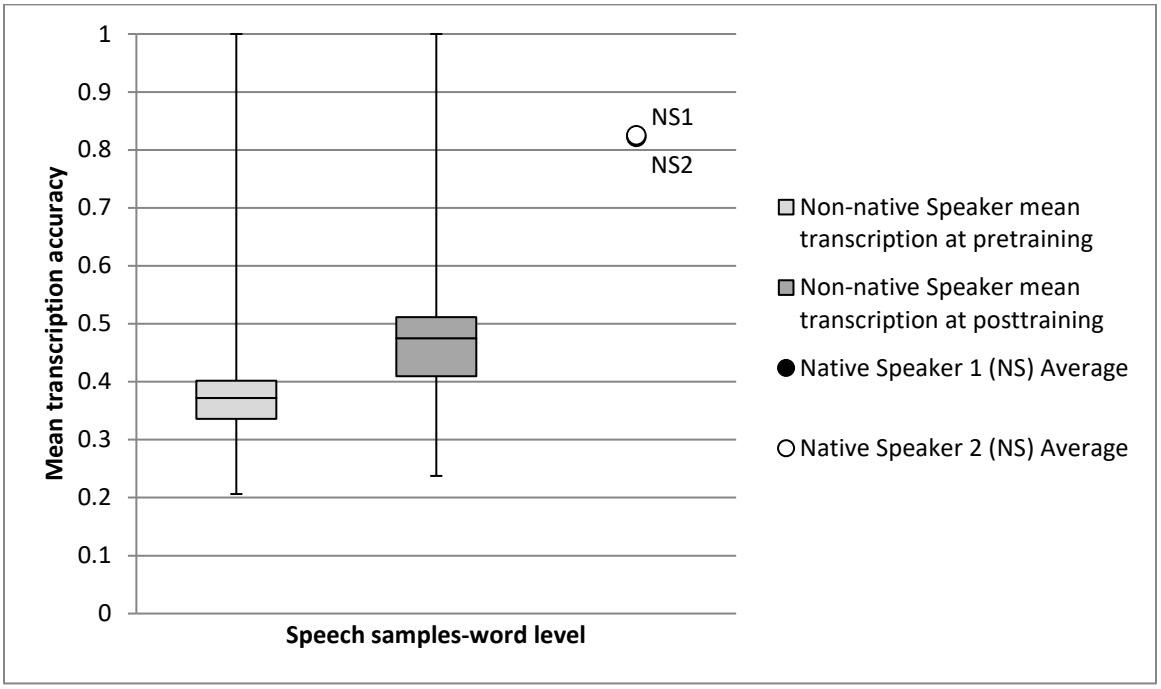

Figure 1

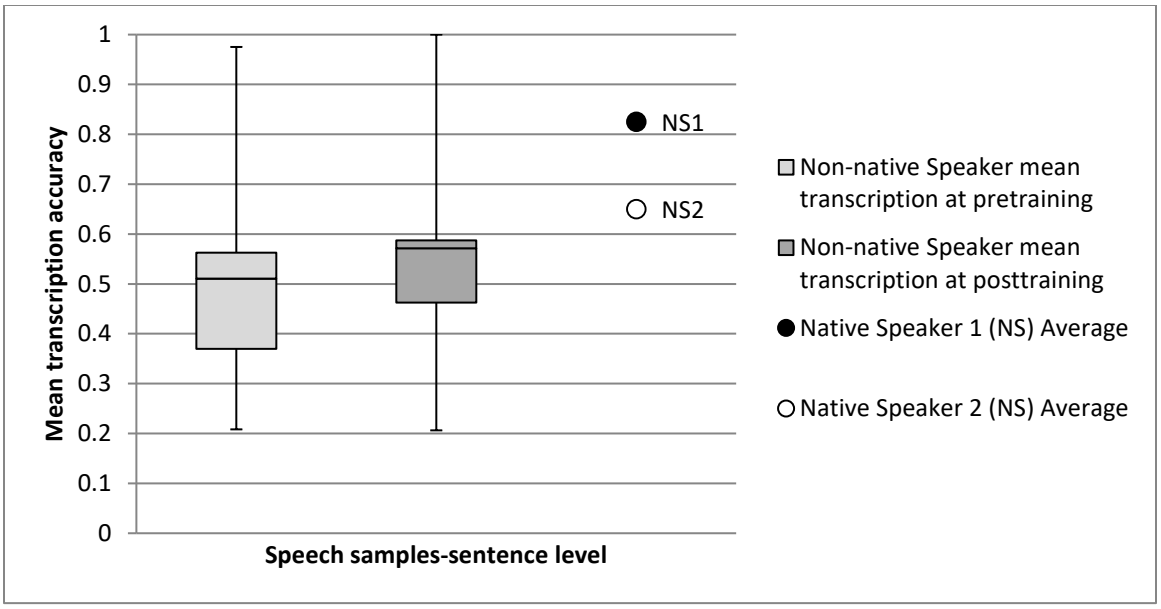

Figure 2

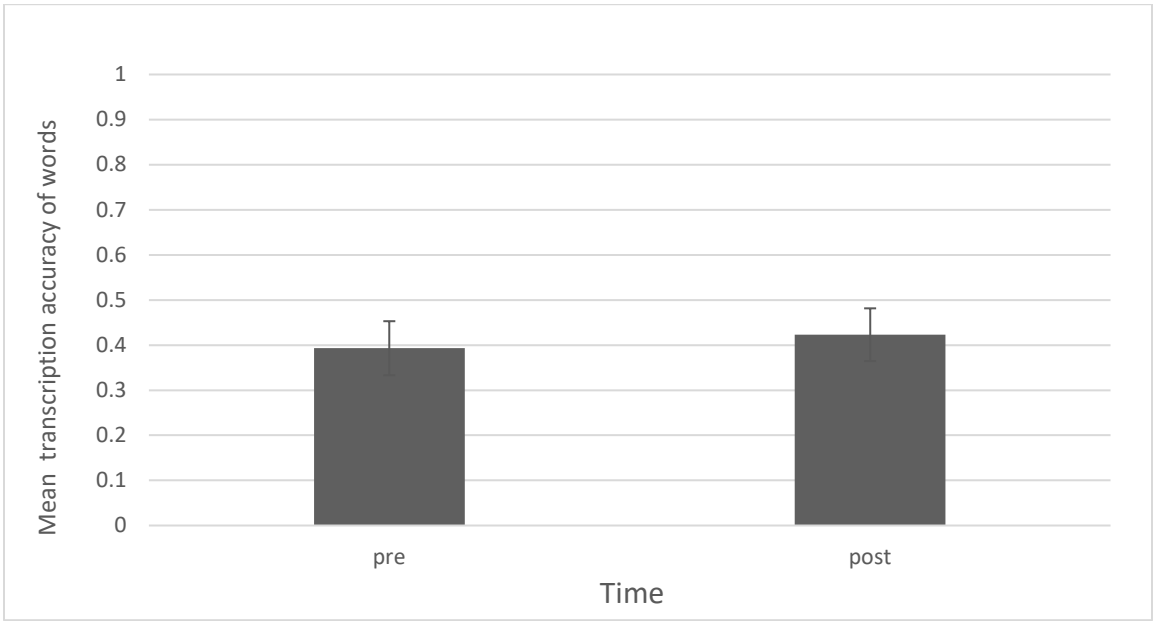

Figure 3

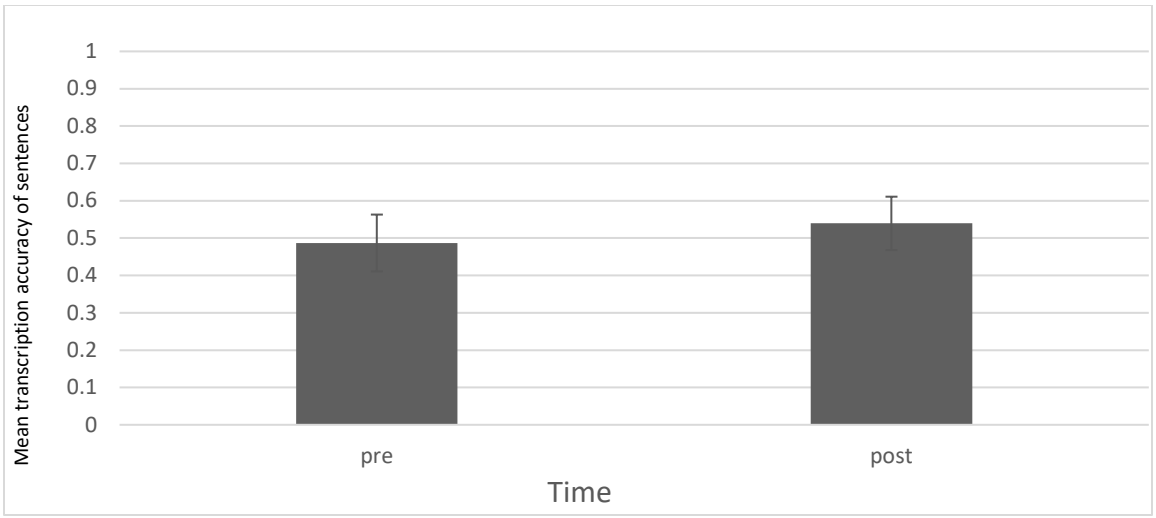

Figure 4

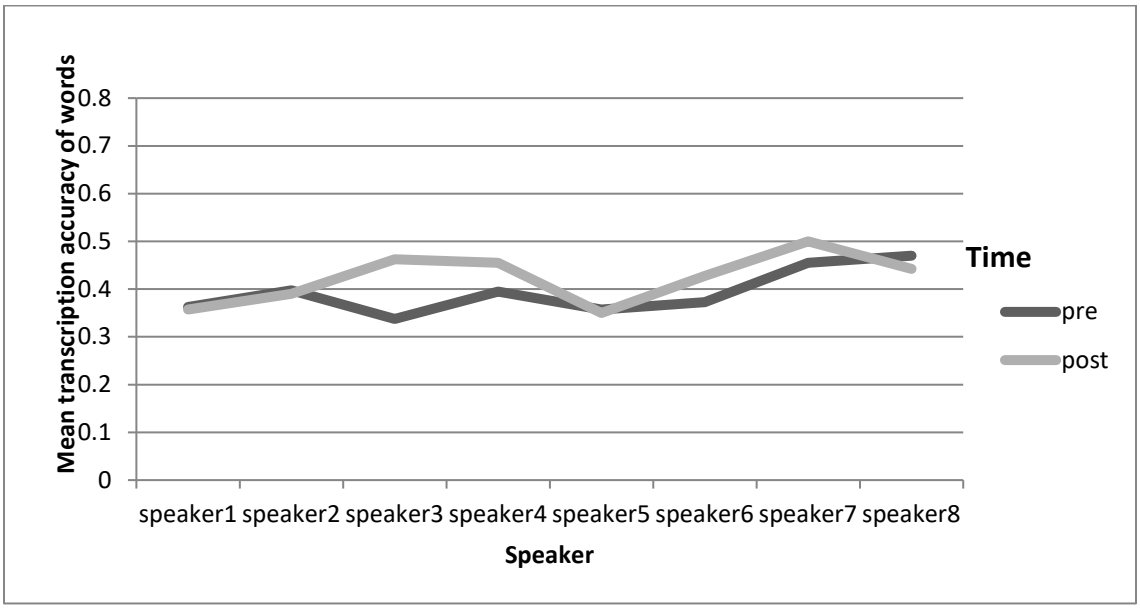

Figure 5

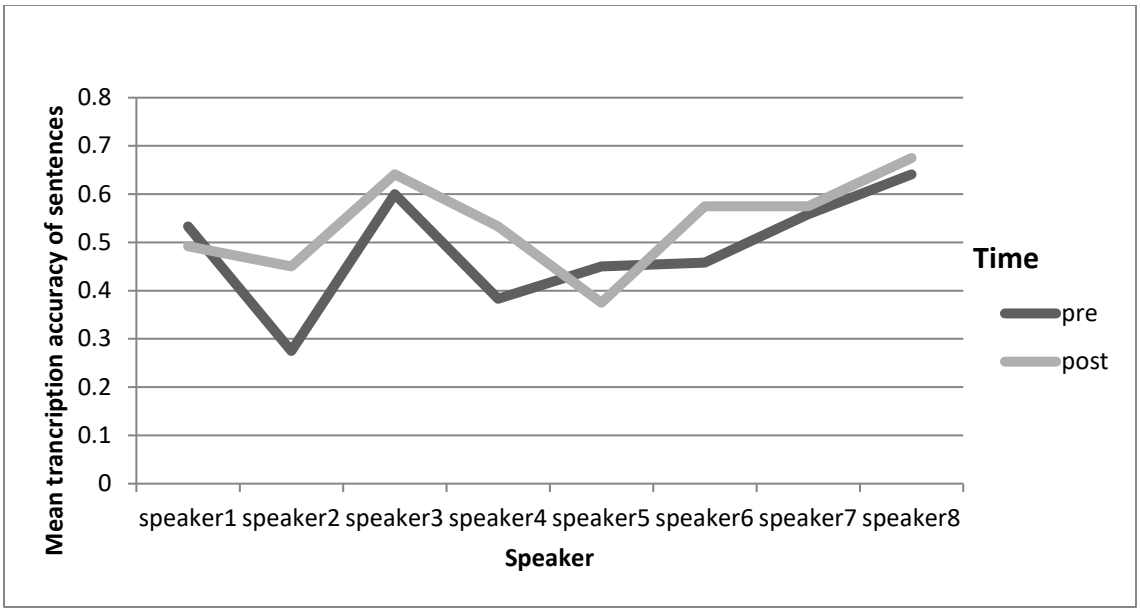

Figure 6

Figure captions

Figure 1: Average marker of the two native speaker' speech samples (NS1 and NS2) and boxplots of the non-native speakers' speech samples as transcribed by native listeners at word level.

Figure 2: Average marker of the two native speaker' speech samples (NS1 and NS2) and boxplots of the non-native speakers' speech samples as transcribed by native listeners at sentence level.

Figure 3: Transcription accuracy of Greek-speakers' speech samples by native speakers at word level. Error bars show +/- 1 standard error from the mean.

Figure 4: Transcription accuracy of Greek speakers' speech samples by native speakers at sentence level. Error bars show +/- 1 standard error from the mean.

Figure 5: Transcription accuracy of Greek speakers' speech samples by native speakers in interaction with time (pre; post) at word level.

Figure 6: Transcription accuracy of Greek speakers' speech samples by native speakers in interaction with time (pre; post) at sentence level.