The Effectiveness of CALL in Helping Persian L2 Learners Produce the English Vowel /ɒ/

Meisam Rahimi
rahimi.meysam@gmail.com
University of Isfahan, Iran

Mansoor Tavakoli
mr.tavakoli14@gmail.com
University of Isfahan, Iran

ABSTRACT

Recent research in pronunciation training has indicated a growing interest in the application of computer-based speech-production techniques. This paper tries to evaluate the effectiveness of a console-based Praat script that utilizes acoustic data in real time to help Persian L2 learners to improve their production of the English vowel /ɒ/. This vowel is not among the Persian vowels and believed to be difficult for Persian learners of English to perceive and produce. A group of 30 Persian ESL learners was recruited - 15 learners were randomly assigned to the experimental group and 15 to the control group. Over a three-week period, the experimental group received acoustic-articulatory training and was exposed to the CALL software for receiving acoustic feedback, while the control group was exposed only to auditory input on the target sound. The groups were given a pretest to ensure their comparability, an immediate posttest to evaluate the effectiveness of the feedback provided and a generalization test to see whether the participants were able to generalize the possibly developed knowledge to new contexts. The results of the study showed a significant improvement in the performance of the participants in experimental group in terms of both the posttest and the generalization test. These findings lend support to the feasibility of the use of much simpler and more available CALL tools than those reported in previous research for foreign language segmental acquisition and its effectiveness in generalization of the acquired skills to new contexts.

Keywords: pronunciation; CALL; articulation analysis; acoustic training; acoustic-articulatory feedback

INTRODUCTION

While pronunciation research findings have asserted that such components as stress, intonation and L2 phonemic contrasts are of pedagogical value (Patten & Edmonds, 2013), pronunciation instruction has received less attention compared to literacy skills, grammar and vocabulary (Derwing & Munro, 2005; Patten & Edmonds, 2013). Of the limited attention paid to pronunciation instruction, the bulk of it has been focusing on suprasegmental instruction (Tanner & Landon, 2009), with vowel instruction being marginalized. The marginalization can be accounted for on a number of grounds. Firstly, teaching vowels is relatively challenging due to the fact that we cannot always give a clear description of their articulatory properties and also because special instrumentation is required for observing their articulation (Wang & Munro, 2004). Secondly, either instructors do not know (or think they don't) how to deal with it effectively (Derwing & Munro, 2009) or/and current communicative methods consider the traditional drills as inappropriate for pronunciation instruction (Saalfeld, 2011). Last but not least, few well-controlled studies have reported empirical findings of vowel pronunciation training and its effects (Derwing & Munro, 2005).

Only recently have researchers become more interested in second language vowel pronunciation training and more specifically the application of speech-production-based techniques in this area (Ouni, 2013). These techniques are based on the idea that learners' metacognitive knowledge of how non-native phonemes are articulated, or their noticing articulatory dissimilarities between native and non-native phonemes, can foster their production. The frequent teaching method employed in this regard has been articulator placement or imitation (Patten & Edmonds, 2013). With regard to assessment, learners' improvement has been evaluated either directly through measuring the acoustic features of the learners' speech production or indirectly through evaluating the acoustic realization of learners' pronunciation. The method usually undertaken in the first approach is to provide a real-time visual feedback for learners' actual articulation through specialized software.

The major advantage of the modern methods in vowel pronunciation training in comparison with traditional methods is the potential of speech technology in identifying problematic areas in learners' pronunciation (Engwall & Bälter, 2007). Two major methods for the identification of articulation errors have been identified namely: phoneme classification and articulatory feature detection (Engwall, 2012). In phoneme classification, specific categories are set up for both correct and incorrect pronunciations into which the learners' speech production might be classified (Eskenazi, 2009). On the other hand, feature detection tries to find aspects of acoustic signal which can yield articulatory information (Teppermann & Narayanan, 2008). In phoneme classification, the output obtained from each detector is compared to the target criterion and finally the deviating features deliver the intended feedback (Strik, Truong, de Wet, & Cucchiarini, 2009). The feature detection is preferred to phoneme classification because it allows focusing the feedback on the particular features which deviate from the target norm, instead of reliance on general differences between phonemes.

Considering the disadvantages, previous research on computer assisted vowel pronunciation training can be criticized from two major perspectives. First, the audio or visual feedback is typically delivered through instruments (e.g., ultrasound machines, electropalatographs and electromagnetic articulographs) which have originally been developed for therapeutic purposes (Abberton & Fourcin, 1975). In other words, most of these techniques require an expert to tune and prepare the instruments before they can be used by non-experts. Consequently, these feedback techniques are not yet practical or sufficiently convenient to be used widely in language learning contexts and even for speech therapy on a large scale. For these reasons, finding alternative approaches which can be freely and easily available to general public are of interest to both researchers and language learners. Second, the previous research has largely ignored investigating learners' ability to generalize their new pronunciation skills to new contexts, if any.

This study evaluates the effectiveness of a console-based Praat script that calculates the first and second formants of the EFL learners' pronunciation of the target vowel. Praat is freely available and requires no linguistic or non-linguistic expertise in its application to both vowel pronunciation teaching and assessment. Specifically speaking, this study tries to examine the effectiveness of the feedback on the tongue position in EFL learners' pronunciation of the target vowel on the learners' pronunciation improvement with regard to the vowel. It is assumed that much less complex and more available tools than those employed in the previous studies can be used in vowel pronunciation pedagogy and assessment. Moreover, this study attempts to investigate whether learners are able to generalize their new skill acquired through the CALL tool to new contexts.

ISSN: 1675-8021

LITERATURE REVIEW

It has been asserted that technology has great potentials for language learning (Gabarre, Gabarre, Din, Shah, & Karim, 2014; Greenhow, Robelia, & Hughes, 2009) in various learning contexts (de Andrés Martínez, 2012). Amongst all, pronunciation has been a traditionally neglected facet, yet, an important aspect of L2 learning (Sturm, 2013). The effectiveness of technology in foreign language learning is strongly supported by the findings of studies from computer-assisted pronunciation training (Golonka, Bowles, Frank, Richardson, & Freynik, 2014).

Many researchers have suggested that speech instruction should focus on aspects and features that bring about confusion or hinder communication if pronounced improperly (Derwing & Munro, 2005, 2009). Miller (2012) asserts that "the speaker needs to physically produce the sounds of the target language with enough accuracy to be understood" (p. 49). American Council on the Teaching of Foreign Languages/National Council for Accreditation of Teacher Education (ACTFL/NCATE) Standards for the Preparation of Foreign Language Teachers (ACTFL, 2002) include in its phonology standards the learners' understanding the rules of the sound system and their ability to diagnose their own target language pronunciation difficulties. In other words, ACTFL and NCATE consider phonological awareness and accurate pronunciation of the target segmentals vital enough to be among its standards for L2 teachers. However, SLA research on speech has not been able to reach a consensus on a method to phonological instruction (Chan, 2010).

Many years of laboratory work on second language phonetic learning precede the application of CALL in pronunciation instruction (Wang & Munro, 2004). The application of Automatic Speech Recognition (ASR) technology has proven to entail facilitatory effects in pronunciation improvement. The results of studies on programs that compare the acoustic analysis of learners' recorded speech with that of native speakers have also demonstrated noticeable improvement in learners' prosody and vowel pronunciation (Carey, 2004; Hardison, 2004). For instance, in investigations conducted on French and Japanese college-level students, Kay's Computerized Speech Lab (CSL) was used to evaluate learners' improvement in pitch, prosody and duration aspects (Hardison, 2004; Hirata, 2004). The findings demonstrated a significant improvement in prosody and segmental accuracy and the learners' ability to generalize their newly acquired skills to new contexts. In another Kay-product study, Sona-Match was employed to investigate the impact of visual feedback on ESL learners' pronunciation of vowels (Carey, 2004). The results of the study showed significant improvement in the pronunciation of the participants in the experimental group after the treatment. Engwall (2012) tried to precisely estimate a learner's articulation through estimating the important articulatory features using acoustic-to-articulatory inversion. He further investigated whether learners were able to imitate the articulatory changes suggested by a virtual pronunciation teacher through audiovisual feedback. In this study, ultrasound imaging was used to monitor the articulatory changes made by seven learners who receive audiovisual feedback. The findings of the study indicated short-term changes in articulation of the participants who received the feedback. The results of Quintana-Lara's study (2012) revealed the effectiveness of acoustic spectrographic instruction on the production of the English phonological contrast /i/ and /ɪ/. While Quintana-Lara employed the Praat for analysis, the undertaken analytical procedure (creating spectrograms of speech samples, identifying the target segments in a spectrogram, and identifying the acoustic features (F1, F2) of the target segments) seems demanding for language learners. In a more recent study, Patten and Edmonds (2013) examined the efficacy of phonetic training on native Japanese speakers' production of American /r/ through spectrographic visual feedback provided by Visi-Pitch™ IV acoustic analysis software (KayPENTAX, model 3950). Native Japanese learners of English receiving the spectrographic visual feedback demonstrated a significant improvement in the pronunciation of /r/.

As the review of the above studies indicates, almost all training and assessment instruments employed (except for Quintana-Lara's study (2012) where the procedure seems burdensome) are specialized instruments originally intended for therapeutic or other technical purposes. These instruments are not only largely unavailable to language learners but also to speech therapists on a large scale. Even if the instruments are available, either an expert is often required to tune and prepare them before they can be used or non-experts would need to receive linguistic or non-linguistic training for using them. As a result, the application of alternative instruments which are available to general public and do not require special knowledge for implementation are of interest to researchers, instructors, and language learners. Moreover, except for the studies conducted by Hardison (2004) and Hirata (2004) in which Kay's CSL was used, all studies ignored the investigation of learners' ability in generalizing their newly acquired pronunciation skills to new contexts. As a result, it might be illuminating for both pronunciation teachers and learners to know about the effectiveness of the application of much simpler and more available tools than those utilized in previous research in pronunciation teaching on learners' ability to generalize their new pronunciation skills to new contexts.

PRONUNCIATION NORMALIZATION

Generally speaking, if a set of vowels with the same phonetic quality are pronounced by different speakers, while the absolute values of formant frequencies will be different across speakers, their relative position on a formant chart will be similar (Ladefoged & Johnson, 2011). Figure 1 demonstrates the first two formant frequencies for the vowels in *heed, hid, head, had, hod, hood, who'd* as spoken by two native speakers of California English:

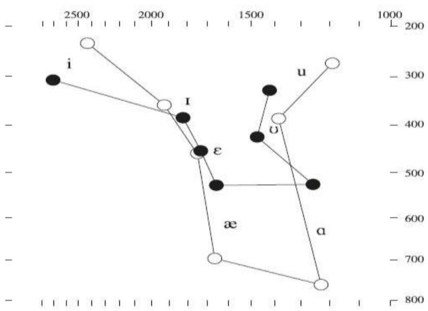

FIGURE 1. A formant chart showing some of the vowels of two speakers of Californian English. The frequency of the first formant is plotted on the ordinate (the vertical axis), and the frequency of the second formant is plotted on the abscessia (the horizental axis).

This cross-personal variation in formant frequencies is observed through different genders (male and female), as well as through different age groups (children and adults) (Stevens, 2000). This variation is due to the size of the speaker's vocal tract (Fant, 2004). Fant (2004) argues that "The F-pattern frequencies are to a first approximation inversely proportional to the length of the speaker's vocal tract from the glottis to the lips" (p. 171). He states that the average formant frequencies for children are higher than adults due to their smaller heads and that "the average female-male difference is of the order of 20%" (p. 171). Moreover, variations have been observed in formant frequencies in an individual's production of the same vowel in the same context but in different occasions and various contexts e.g., CVC and CVCV (Hawkins & Midgley, 2005). To make things more complicated, there might also be some variations in the production of a specific vowel across different accents of a single language or even some vowels might be absent in a particular variation of a language (Ferragne & Pellegrino, 2010). Considering these facts, it would be inappropriate to establish a single criterion for receiving feedback in all contexts. To consider these parameters and to make the criterion more realistic, it would be better to consider a range for a formant as our criterion rather than the mean of the formant, that is, for example, instead of considering 376 as the mean of the first formant for vowel /ʊ/, it would be better to take the range 314-438 as our criterion for correct pronunciation.

Bearing in mind the mentioned elements (i.e. age, gender, language variation, and formant ranges), the first and the second formant frequencies of the target segment (/ɒ/) pronounced by 20 male native English speakers between 20 and 40 years old were measured. The means and the standard deviations were calculated and the range mean ± one standard deviation was considered as the norm reference for the correct pronunciation in this study.

VOWELS ACOUSTIC PROPERTIES AND ARTICULATION

According to Ladefoged and Johnson (2011), height and backness are the two features of vowel quality which can be used to contrast vowels in nearly every language. According to Stevens (2000), the major articulatory operation for producing a high vowel with a first formant low frequency is to raise the tongue body so that a relatively narrow constriction in the oral cavity is created. On the other hand, low vowels are created through lowering the tongue body and constricting the vocal tract in the vicinity of the tongue root. As the tongue height decreases, the frequency of the first formant increases so that F1 for non-low vowels is intermediate between that for a low vowel and that for a high vowel and the first formant of low vowels is maximally high. It can be seen that the first formant corresponds inversely to what we called, in articulatory terms, *vowel height* (Stevens, 2000).

Independent of the tongue height, the tongue body displacement leads to a common acoustic consequence. According to Stevens (2000), "forward movement of the tongue body causes an increase of the second-formant frequency to a maximum value" (p. 282). On the other hand, a backed tongue body yields an F2 value that is maximally low so that it could be concluded that the higher the tongue body the higher this maximally low F2 value would be. Consequently, it can be seen that the second formant frequencies are much higher for the front vowels than they are for the back vowels (Ladefoged & Johnson, 2011). The general assumption behind showing formant frequencies of the target vowel and the related articulatory changes to learners is that they will imitate the suggested changes or implicitly improve their perception of the to-be-learned phonemes and thus their production.

METHOD

PARTICIPANTS

A total of 30 male Persian EFL learners in Iran participated in the study. Their average length of English learning experience was 6.3 years. The participants' ages ranged from 22 to 35 years old (mean = 27.6) at the time of the study. The English majors were from the field of English language and literature, and the non-English majors were from the fields of Humanities, Mechanical Engineering, and Computer Sciences. All the participants started to learn English formally at high school (at the age of thirteen). None of the participants claimed to have received any form of phonetic training.

All participants volunteered to participate in the study. All the study related training and practice took place outside of class time. A pure tone hearing test (The test is available online at: http://www.audiocheck.net/testtones_hearingtestaudiogram.php) demonstrated normal hearing ability for all participants. All participants reported normal learning and vision abilities and no knowledge of a third language. None of the participants reported any study-abroad experience. Additionally, none of them claimed any language contact with native English speakers.

TRAINING AND ASSESSMENT TOOLS

Apart from the instruments mentioned in Section 2, there are many interactive systems based on speech technologies that have been commercialized. The following are some which are based on serious speech technology research: NativeAccent (Carnegie Speech, www.carnegiespeech.com) for teaching pronunciation, ATR (www.atr.jp) for speech perception and production, Versant (Pearson, www.ordinate.com/) for assessing the fluency of non-native speech, integrated courses such as (Rosetta Stone Inc, www.rosettastone.com), Better Accent Tutor (BetterAccent, http://www.betteraccent.com/), Dragon Naturally speaking (Nuance, http://www.nuance.com) and Alelo (www.alelo.com) for language and culture tutoring. The major issue with these programs is that they have limited functions and have been designed mainly for English language teaching so that they are not available for learners of other languages. These programs are also commercially available and require relatively special hardware and supporting software, making them unaffordable or inaccessible for a range of language learners within the classroom and possibly other contexts. On the other hand, Praat (Boersma & Weenink, 2013) is a programmable, multipurpose, freeware program for the reconstruction and analysis of acoustic speech signals and speech analysis. This software offers a wide range of standard and non-standard procedures, including articulatory synthesis, spectrographic analysis, and neural networks. Being resourceful, open source, and available for free, this software has attracted the attention of many researchers as well as practitioners in areas other than phonetic and phonological analysis such as language teaching. The user interface is relatively simple and user-friendly so that it can be tailored and utilized for teaching a whole array of segmental and suprasegmental elements not only in English but also in any other language. Accordingly, Praat was chosen as the basic tool in this study.

In order to make the receiving of feedback a more straightforward process with less delay time and distractions and to provide a better overall picture of the participants' pronunciation improvement, a Praat script was written. The script, a batch file, Praat console application (Praatcon) and all audio files were put in the same folder. When the batch file was double-clicked, a list (figure 2) of the first and second formants of the audio files was displayed in a command prompt window. As for assessment, the spectrograms (figure 3) produced by the Praat program were used in order to extract the formants from the streams of sounds.

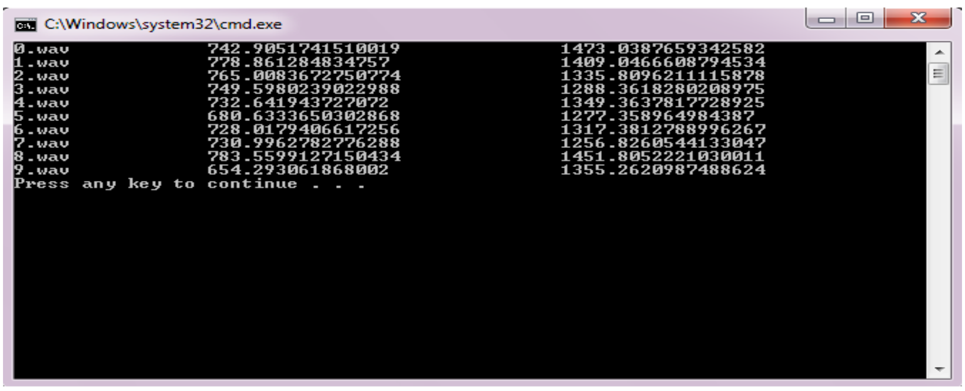

FIGURE 2. The list of the first and second formants in a command prompt window

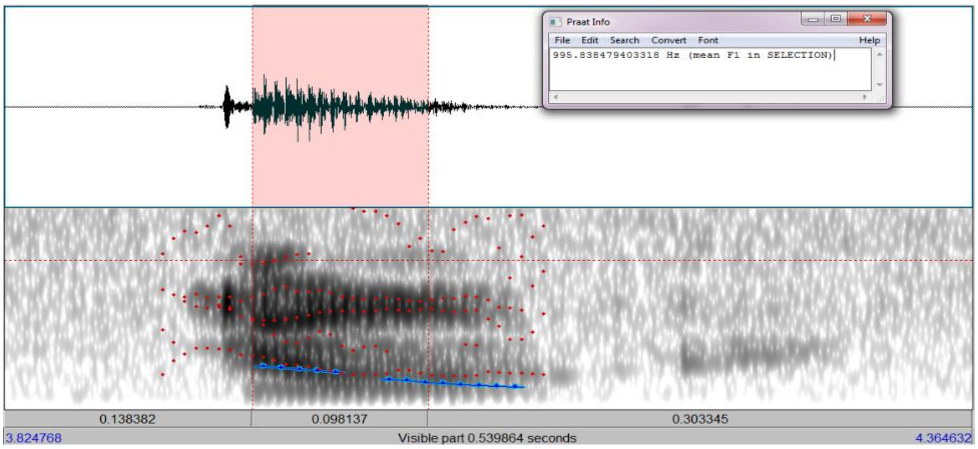

FIGURE 3. Spectrogram of the word "dog" produced by the Praat program and the first formant of the target vowel extracted.

THE STIMULI

Both training and generalization stimuli contained /ɒ/-medial (e.g., fog: /fɒg/) words. Words with only medial position were selected to avoid the distraction, which might have been resulted from the diversity in vowel position. The consonants were common to both Persian and English. The chosen words were monosyllabic and consonants in the syllables appeared as both single consonants and consonant clusters (See Appendix). Monosyllabic words were selected because of their relatively simple pronunciation. Both training and generalization stimuli were presented randomly to participants in all groups and all tests.

PROCEDURES

The training lasted for a period of three weeks. The sessions were conducted twice a week, between fifteen to twenty minutes per session on Saturdays and Tuesdays. Participants were randomly assigned to an experimental research group and a control group. The experimental and control groups had only the first and the last sessions (the pretest and the posttest) in common. Sixty-five stimuli were selected for this study. The training stimuli were used in pretest and during the period between pretest and posttest for practicing in both the experimental and control groups and the generalization stimuli were only used in the posttest. Each stimulus was written on an index card. The index cards were presented in a randomized order across different phases of the study (pretest, training and posttest).

As a prerequisite for the study, all participants were required to have access to computer speakers, and a microphone. Since participants might not have been familiar with the stimuli, they were informed that the stimuli had only one vowel and that would be the target in this study. Moreover, the participants listened to the pronunciation of the training stimuli by a British English speaker (the pronunciation of the words by the Oxford Advanced Learners' Dictionary, 8th Edition) in order to further familiarize them with the target vowel.

During the first session, all the participants completed a background questionnaire and a consent form. Both groups were also given a pretest to ensure that they were comparable. Moreover, considering the fact that the target vowel was utilized by many educated Persian speakers in the area of Tehran (Handbook of the International Phonetic Association, 2007, p. 130), the pretest was also conducted to ensure that the target vowel did not already exist in the vowel system of the participants. To this end, the participants were required to read the related stimuli while their utterances were recorded. The recordings were analysed through the Praat program by experimenters in order to establish the basis for future comparison.

During the second session, the experimenters introduced the Praat program, the Script and the concepts of the first and second formants to the participants in the experimental group. Afterwards, samples of native English speakers' pronunciations were analysed using Praat. Participants were taught how to record their utterance using Praat. Later, under the supervision of the experimenters and through guided practice, the participants recorded themselves and analysed the recordings through the software. Experimenters observed the process carefully in order to eliminate any ambiguities. The participants in the control group were introduced to the target vowel. They were required to listen to the English native pronunciation of the vowels in isolation and embedded in words and then repeat them. These participants were required to go through the same procedure for the rest of the sessions up to the posttest.

As for the participants in the experimental group, the third session was dedicated to the introduction of the target segment (/ɒ/). At the beginning of the session, the acoustic characteristics of the target segment were first explained to the participants and then they practiced. The practice phase included listening to the target segment in isolation and embedded within words, repeating the segment and the words several times, recording oneself while pronouncing the segment and receiving feedback through the system. This process was repeated throughout the session. Participants were required to try to attain F1 and F2 values within the norm reference ranges.

During the fourth session, the 15 new training stimuli were introduced and practiced along with the 15 stimuli from the previous session. During the fifth session, the acoustic features of the target vowel were first reviewed and then participants practiced pronouncing the vowel in isolation and embedded in the words they were provided with, repeating the stimuli and receiving computerized feedback. During the sixth session, the posttest was administered to all the participants in both experimental and control groups. In the posttest, the participants were required to read the training and control stimuli while being recorded. Besides, a generalization test was conducted. The generalization test was intended to evaluate the performance of the participants on new, untrained test items. The recordings were analysed through the Praat program by experimenters in order to identify the possible articulatory changes. All the tests were conducted individually. The recordings were created through the Praat program on a Lenovo computer, using a Genius GHP-430F microphone. The "record mono sound" setting and "11,025 Hz" were selected based on its usage in a study by Patten and Edmonds (2013). All data collections were run individually.

DATA ANALYSIS AND RESULTS

In the scoring procedure, first the F1 and F2 means and standard deviations of the native speakers' pronunciations of the target segment (/ɒ/) were calculated (See table 1). Then, the formant measurements which fell within the first, second and third standard deviations were assigned 3, 2, and 1 point(s), respectively.

TABLE 1. Means and standard deviations of F1 and F2 of the vowel /ɒ/ as spoken by native speakers of English.

Vowel	F	Mean Freq. ($\bar{x}$)	St. dev. (S)	$\bar{x}$ + S	$\bar{x}$ - S
/ɒ/	1	599	67	532	666
	2	891	159	732	1050

To evaluate the magnitude of change in the production of the target segments, the effect sizes (Cohen's *d*) were calculated (Beeson & Robey, 2006). Effect sizes are beneficial in allowing comparisons across and within conditions and participants to scale relative improvement. The frame of reference intended here for interpreting effect sizes is the rule of thumb provided by Oswald and Plonsky (2010) for interpreting Cohen's *d* effect sizes in the SLA field (*d* = .4 is small, *d* = .7 is medium, and *d* = 1.0 is large), which is very similar to the Cohen's as the best known frame of reference (small effects ($d \leq 0.20$), medium effects ($0.20 < d < 0.80$), and large effects ($0.80 \leq d$).

In order to ensure the obtained results in the posttest were merely due to the applied treatment, the means of the formant measurements of the participants were compared with the pronunciation norm range (see Pronunciation Normalization). The results of the comparison indicated that the participants' means was not in the norm range, hence indicating the lack of the vowel in the phonetic system of the participants.

ANALYSIS AND RESULTS OF THE TRAINING STIMULI

Two repeated measures analyses of variance (ANOVA) (one for each acoustic measurement, F1 and F2) and four two-tailed t-tests (two for each acoustic measurement, F1 and F2, respectively) were obtained.

The two t-tests revealed no significant differences on the pretest performance of the participants for both F1 and F2, $t(28) = 0.453$, $p = .741$ and $t(28) = 0.512$, $p = 0.695$, respectively. However, the results of a two-tailed t-test revealed a significant difference between groups on the posttest, $t(28) = 5.745$, $p = 0.001$, $t(28) = 6.682$, $p < 0.001$. Cohen's *d*s were 0.9 and 1.01, respectively, indicating a large effect size according to our frame of reference. This suggests that while the groups were comparable at the beginning of the study, at the end of the study, the groups were no longer equivalent in their ability to accurately pronounce the target segments.

Another two-tailed t-test on rate of improvement (posttest score - pretest score) demonstrated a significant difference between groups for both acoustic measurements (F1 & F2), $t\ (28) = 9.435$; $p < 0.001$, $t\ (28) = 10.215$; $p < 0.001$. Cohen's *d*s were 1.324 and 1.126, respectively, indicating a large effect size according to our frame of reference. In other words, the rate of improvement for the experimental group was significantly higher than that of the control group.

A repeated measure ANOVA with time (pretest and posttest) as within-group factors, and group (experimental and control) as between-group factors was conducted for each acoustic measurement (F1 and F2) to see whether or not there was a significant difference between the experimental and control groups after treatment. Considering the first acoustic measurement (F1), an effect was observed for time as the within-subjects factor (F (1, 28) = 82.637, p < 0.001). There was also an interaction effect for time × group (F (1, 28) = 79.121, p = 0.000). Regarding the second acoustic measurement (F2), the analysis revealed an effect for time as the within-subjects factor (F (1, 28) = 92.543, p < 0.001). Besides, there was also an interaction effect for time × group (F (1, 28) = 83.769, p = 0.000). The observed effects for time and time × group confirm the results of the t-tests reported above.

ANALYSIS AND RESULTS OF GENERALIZATION STIMULI

Like the previous section, two repeated measures analyses of variance (ANOVA) (one for each acoustic measurement, F1 and F2) and four two-tailed t-tests (two for each acoustic measurement, F1 and F2) were obtained.

Considering both F1 and F2, the two two-tailed t-tests run on the pretest performance showed no significant differences, t (28) = 0.493, p = 0.782 and t (28) = 0.482, p = 0.712, respectively. However, the results of a two-tailed t-test revealed a significant difference between groups on the posttest, t (28) = 5.128, p = 0.012, t (28) = 6.169, p = 0.008. Cohen's *d*s were 0.8 and 0.91, respectively, indicating large effect sizes according to our frame of reference. This suggests that not only the performance of the trainees developed significantly from the pretest to the posttest, but they were also able to generalize the improvement to new contexts. What was noticeable was that the effect sizes for the generalization test are smaller than those in the previous section. This effect could be attributed to the amount of practice advocated to the training stimuli.

Another two-tailed t-test on the pretest and posttest scores demonstrated a significant difference between groups for both acoustic measurements (F1 & F2), t (28) = 8.856; p = .001, t (28) = 9.275; $p < 0.001$. Cohen's ds were 0.92 and 1.01, respectively, indicating a large effect size. As shown in the previous section, the rate of improvement for the experimental group was significantly higher than that of the control group. Once again, the results of the above t-tests indicate that the participants in the experimental group were able to generalize the relative improvement in the pronunciation of the target segment to new contexts.

A repeated measures ANOVA with time (pretest and posttest) as within-group factors, and group (experimental and control) as between-group factor was conducted for each acoustic measurement (F1 and F2) of the target vowel in the generalized stimuli to see whether or not there was a significant difference between the experimental and control groups after treatment. Considering the first acoustic measurement (F1), an effect was observed for time as the within-subjects factor (F (1, 28) = 88.372, $p < 0.001$). The interaction between the factors time × group was also significant (F (1, 28) = 83.662, p = 0.000). Regarding the second acoustic measurement (F2), the analysis revealed an effect for time as the within-subjects factor (F (1, 28) = 89.413, $p < 0.001$). There was also an interaction effect for time group (F (1, 28) = 81.385, p = 0.000). Like the previous section, the observed effects for time and time × group confirm the results of the above-reported t-tests.

Participants also reported no access to native speaker models and media, with regard to native speaker models. The daily reports of the participants also indicated their daily practice as required.

DISCUSSION AND CONCLUSION

The current study developed pronunciation practices enhanced by computer generated feedback through a Praat script, and assessed its usefulness in helping non-native learners to improve their English vowel pronunciation. The results of the study revealed that the performance of the participants improved significantly in both training and generalization groups after the implication of the treatment.

The findings of the present study confirm the previously reported positive effects of computer assisted phonological training regarding segmentals via acoustic and spectrographic instruction. The contribution this study makes to the body of research in this area is twofold. First, the instruments employed in previous studies were specialized tools originally developed for therapeutic (e.g., ultrasound machines, electropalatographs and electromagnetic articulographs) or other technical purposes. For instance, Carey (2004) utilized Sona-Match (a KayPENTAX product) to investigate the influence of visual feedback on second language learners' vowel pronunciation. The findings of the study showed a significant improvement in the pronunciation of the participants. Engwall (2012) used ultrasound imaging in order to estimate important articulatory features using acoustic-to-articulatory inversion and consequently to investigate the effectiveness of audiovisual feedback on learners' articulatory changes. This study demonstrated short-term improvements in articulation of the participants who received audiovisual feedback. In a more recent study, Patten and Edmonds (2013) utilized Visi-Pitch™ acoustic analysis software in order to provide Japanese second language learners of English with spectrographic visual feedback. The findings of the study suggested that spectrographic visual feedback can be considered a promising method for segmental training. Even in the faint case of availability of these instruments for pedagogical purposes, an expert is required to fine-tune and prepare them for use by non-experts. In other words, these feedback techniques are not yet practical for language learners, language teachers, most of the researchers in this area and even for speech therapists on a large scale. This study, on the other hand, provided supportive evidence for the effectiveness of a CALL tool which is freely available and requires no linguistic or non-linguistic expertise in its application to both vowel pronunciation teaching and assessment. The features of the tool make it generalizable to other segmentals not only in Persian but also in any other language and to both classroom and self-regulated contexts. In practical terms, this tool is best suited for self-regulated pronunciation practice and improvement as learners can receive the initial training in the classroom and then practice the target segmentals wherever and whenever it suits them. Regarding classroom contexts, special settings are required so that a learner's voice does not interrupt other learners' practice and even degrade the feedback they receive.

Second, previous research has largely dealt with the effects of articulatory instructions merely on training stimuli, ignoring the ability of learners in generalizing their newly developed skills to new contexts. Hardison (2004) conducted two experiments in order to investigate the effectiveness of computer-assisted prosody training and its generalizability to segmental accuracy and novel sentences. While this study reported on the generalization of the training, a specialized tool, that is, Kay's CSL was utilized to provide the intended training on suprasegmental elements. The Kay's CSL is an advanced hardware and software speech analysis system with a broad range of application in analyzing speech and voice, which has been mainly used by clinicians at leading medical centers and by university researchers. Hirata (2004) also employed the same tool in her assessment of the effectiveness of a pronunciation training program which provided visual feedback on pitch and durational contrasts to English second language learners of Japanese. As in the previous study, the trained group in this study showed improvement in the perception and production of both training and novel test materials. Unlike the majority of the research in this area and like the above two studies, this study assessed the generalizability of articulatory instruction. What differentiates this study from the studies conducted by Hardison (2004) and Hirata (2004) is the utilized instruments. In other words, Hardison (2004) and Hirata (2004) employed specialized instruments in order to apply their treatment, while in this study a much simpler, easier to obtain, use and cheaper instrument was used. The features of the tool make this specific CALL software and methodology generalizable to other segmentals not only in Persian but also in other languages and to both classroom and self-regulated contexts.

Another issue to be mentioned is the rationale behind the selection of only one vowel (i.e. /ɒ/) as the target segmental. The vowel /ɒ/ was selected as the target segmental due to its absence in Persian, making it difficult for Persian L2 learners of English to produce and perceive. Certainly, further research can investigate the effectiveness of the application of such simple instruments on articulatory instruction regarding other segmentals in other languages.

Overall, the results of the study imply that timely individualized computerized feedback is available for learners without any need for special requirements and can be used outside the classroom to help learners self-regulate their own learning in an autonomous way, accelerating their learning process and decreasing teachers' loads. Further research will be beneficial that focuses on investigating the effectiveness of such tools with regard to other segmentals and across learners with different levels of proficiency. The long-term retention of the effects of such training over several weeks can also be investigated in future study.

ACKNOWLEDGEMENTS

The authors are most grateful to Prof Sarah Hawkins, Prof John C. Wells, and Dr. Batool Alinezhad for their considerate care and attention and their invaluable contributions.

REFERENCES

Abberton, E. & Fourcin, A. (1975). Visual Feedback and the Acquisition of Intonation. In E.H. Lenneberg & E. Lenneberg (Eds.). *Foundations of Language Developments: A Multidisciplinary Approach* (2) (pp.157-65). Academic Press.

American Council on the Teaching of Foreign Languages (ACTFL) (2002). *ACTFL Program Standards for the Preparation Offoreign Language Teachers*. Yonkers, NY: Author.

Beeson, P.M., & Robey, R.R. (2006). Evaluating Single-subject Treatment Research: Lessons Learned from the Aphasia Literature. *Neuropsychology Review*. *16*(4), 161-169.

Boersma, P., & Weenink, D. (2013). Praat: Doing Phonetics by Computer (Version 5.3.29) [Computer program]. Retrieved September 30, 2013, from http://www.praat.org

Carey, M. (2004). CALL Visual Feedback for the Pronunciation of Vowels. *CALICO Journal*. *21*, 571-601.

Chan, A., Y.W. (2010). Advanced Cantonese ESL Learners' Production of English Speech Sounds: Problems and Strategies. *System*. *38*, 316-328.

De Andrés Martínez, C. (2012). Developing Metacognition at a Distance: Sharing Students' Learning Strategies on a Reflective Blog. *Computer Assisted Language Learning*. *25*(2), 199-212. DOI: 10.1080/09588221.2011.636056

Derwing, T.M., & Munro, M.J. (2005). Second Language Accent and Pronunciation Teaching: A Research-based Approach. *TESOL Quarterly*. *39*(3), 379-397.

Derwing, T.M., Munro, M.J. (2009). Putting Accent in its Place: Rethinking Obstacles to Communication. *Language Teaching*. *42*, 476-490.

Engwall, O. (2012). Analysis of and Feedback on Phonetic Features in Pronunciation Training with a Virtual Teacher. *Computer Assisted Language Learning*. *25*(1), 37-64. DOI: 10.1080/09588221.2011.582845

Engwall, O., & Bälter, O. (2007). Pronunciation Feedback from Real and Virtual Language Teachers. *Computer Assisted Language Learning*. *20*, 235-262.

Eskenazi, M. (2009). An Overview of Spoken Language Technology for Education. *Speech Communication*. *51*, 832-844.

Fant, G. (2004). *Speech Acoustics and Phonetics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ferragne, E. and Pellegrino, F. (2010). Formant Frequencies of Vowels in 13 Accents of the British Isles. *Journal of the International Phonetic Association*. *40* (1), 1-34.

Gabarre, C., Gabarre, S., Din, R., Shah, P. M., & Karim, A. A. (2014). iPads in the Foreign Language Classroom: A Learner's Perspective. *3L: Language Linguistics Literature®*, *Southeast Asian Journal of English Language Studies*. *20*(1), 115-127.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L. & Freynik, S. (2014) Technologies for Foreign Language Learning: A Review of Technology Types and their Effectiveness. *Computer Assisted Language Learning*. *27*(1), 70-105. DOI: 10.1080/09588221.2012.700315

Greenhow, C., Robelia, E., & Hughes, J. (2009). Web 2.0 and Classroom Research: What Path should We Take Now? *Educational Researchers*. *38*, 246-259.

Hardison, D. (2004). Generalization of Computer Assisted Prosody Training: Quantitative and Qualitative Findings. *Language Learning and Technology*. *8*(1), 34-52. Retrieved October 25, 2006 from http://llt.msu.edu/vol8num1/hardison

Hawkins, S. & Midgley, J. (2005). Formant Frequencies of RP Monophthongs in Four Age Groups of Speakers. *Journal of the International Phonetic Association. 35*(2), 183-199.

Hirata, Y. (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning. 17*(3-4), 357-376. DOI: 10.1080/0958822042000319629

Ladefoged, P. & Johnson, K. (2011). *A Course in Phonetics (Sixth Edition)*. Boston, MA: Wadsworth**.**

Miller, J.S. (2012). Teaching French Pronunciation with Phonetics in a College-level Beginner French Course. *NECTFL Review. 69*, 47-68.

Oswald, F.L., & Plonsky, L. (2010). Meta-analysis in Second Language Research: Choices and Challenges. *Annual Review of Applied Linguistics*. *30*, 85-110.

Ouni, S. (2013). Tongue Control and its Implication in Pronunciation Training. *Computer Assisted Language Learning*. 1-15. DOI: 10.1080/09588221.2012.761637

Patten, I. & Edmonds, L. A. (2013). Effect of Training Japanese L1 Speakers in the Production of American English /r/ using Spectrographic Visual Feedback. *Computer Assisted Language Learning*. 1-19. DOI: 10.1080/09588221.2013.839570

Quintana-Lara, M. (2012). Effect of Acoustic Spectrographic Instruction on Production of English /i/ and /I/ by Spanish Pre-service English Teachers. *Computer Assisted Language Learning*. 1-21. DOI: 10.1080/09588221.2012.724424

Saalfeld, A.K. (2011). Acquisition of L2 phonology in Advanced Learners: Does Instruction Make a Difference? In *Proceedings of the Second Annual Conference on Pronunciation in Second Language Learning and Teaching*, pp. 144-152.

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Strik, H., Truong, K., de Wet, F., & Cucchiarini, C. (2009). Comparing Different Approaches for Automatic Pronunciation Error Detection. *Speech Communication, 51*, 845-852.

Sturm, J. L. (2013). Explicit Phonetics Instruction in L2 French: A Global Analysis of Improvement. *System. 41*, 654-662

Tanner, M. W. & Landon, M. M. (2009). The Effects of Computer-assisted Pronunciation Readings on ESL Learners' use of Pausing, Stress, Intonation, and Overall Comprehensibility. *Language Learning & Technology. 13*(3), 51-65.

Teppermann, J., & Narayanan, S. (2008). Using Articulatory Representations to Detect Segmental Errors in Nonnative Pronunciation. *IEEE Transactions on Audio, Speech and Language Processing. 16*(1), 8-22.

Wang, X. & Munro, M. J. (2004). Computer-based Training for Learning English Vowel Contrasts. *System. 32*, 539-552.

APPENDIX

Training Stimuli		Generalization Stimuli	
cock	cost	blot	mop
cod	non	bod	mot
col	odd	bonze	nod
cop	rod	chock	odd
god	rot	cocks	rot
jock	sop	con	slop
mod	stock	crock	sod
pop	tossed	doss	tock
rob	wok	dot	tod
rod	got	fog	tot
ROM	hop	glob	block
shod	on	hod	bond
smock	not	lob	bossed
sock	want	lop	clock
sod		mod	

ABOUT THE AUTHORS

Meisam Rahimi is a doctoral student of Applied Linguistics at the University of Isfahan, Isfahan, Iran. His primary research interests include computational linguistics, language acquisition, and computer assisted language learning.

Mansoor Tavakoli is an Associate Professor at the University of Isfahan, Isfahan, Iran. He has published many articles in various journals in the area of English language teaching and learning. His main areas of interest are language testing and research in applied linguistics.