ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 12, No. 8, pp. 1545-1555, August 2022
DOI: https://doi.org/10.17507/tpls.1208.10

Ultrasound Overlay Video Effectiveness in Teaching Pronunciation to Young EFL Learners

Fatimah H. Alshehri
Ministry of Education, Abha, Saudi Arabia

Abstract—**Professionals in the language teaching field have recently shown increased interest in using ultrasound technology as a visual feedback tool in pronunciation instruction. It is difficult for students to interpret and for larger learner groups and independent learners to apply. Further, without specialised equipment and expertise, this technology cannot be incorporated into teaching and learning contexts. Developing ultrasound overlay videos is an attempt to address the limitations of the current technology in language teaching contexts. This study compares the /p/ and /b/ production by learners who received ultrasound overlay video training, as compared to those who didn't. All learners participated in recordings and perception quizzes before and after the training and 11 days post-training for the treatment group. No effects were noted regarding the perception and pronunciation of the segments between the groups. The same was observed during their follow-up, suggesting that ultrasound overlay videos may not be effective in helping young female learners perceive and pronounce word segments and retain the benefits of language instruction.**

Index Terms—**pronunciation instruction, ultrasound overlay videos, ultrasound, visual feedback**

I. Introduction

Effective communication requires a proficient use of the phonological elements of the target language (Moghaddam et al., 2012). Thus, placing great emphasis on knowledge of phonology and phonetics is essential to enhance pronunciation (Hameed & Aslam, 2015). Despite learners' differences in learning L2, the age factor is one of the areas that highly affect pronunciation proficiency (Çakır & Baytar, 2014; Derwing, 2008; Derwing & Munro, 2005; Gilakjani & Ahmadi, 2011; Piske et al., 2001). Based on the widely held Theory of Critical Period hypothesis (Lenneberg, 1967), native-like proficiency, especially in pronunciation, can be achieved between the ages of 2 and 13 (Loewen & Reinders, 2011). The interference of learners' native languages subsequently gets stronger as they get older (Damayanti, 2008). Clearly, mispronunciation habits acquired in childhood becomes resistant to change, and thereby changing them takes a lot of time and effort (Gilakjani & Ahmadi, 2011). Young learners are more likely to achieve flawless, native-like pronunciation with accurate instruction (Reid, 2016).

Thus, it is important to implement special efforts to pronunciation teaching (Çakır & Baytar, 2014), and, most importantly, ensure the effectiveness of the instruction itself (Derwing, 2008). Auditory input, which the listen-and-repeat method heavily draws on, is insufficient to develop metalinguistic awareness (Neri et al., 2002) and straightforwardly map pronunciation from sound to articulation. Besides, Broughton et al. (1993) argue that accurate imitation depends mostly on how accurately someone hears what is imitated. This applies to the segment /p/ which is recognised as a problematic consonant that is commonly mispronounced as /b/ by native Arabic speakers learning English (Elmahdi & Khan, 2015; Hameed & Aslam, 2015). With advances in technology, however, the possibilities of providing accurate and visual feedback on pronunciation are increasing more than ever before. The past decade has witnessed an increased interest in using ultrasound as a visualisation tool in fields such as language teaching, speech therapy, and articulatory phonetics (Bliss et al., 2018).

Ultrasound can be defined as a technology that "emits ultra-high-frequency sound through a transducer or 'probe' containing piezoelectric crystals" (Gick et al., 2008, p. 309). Speech ultrasound imaging is a reflection of the sound traveling through the tongue, which is obtained through an ultrasound transducer held against the skin of the neck (Gick et al., 2008). The ultrasound also creates an image of the tongue through high-frequency sound waves emitted by an ultrasound probe (Byun et al., 2014). According to Byun et al. (2014), it is a technique that "allows the client and clinician to observe tongue position and shape to directly cue changes in tongue position or shape and evaluate whether the client has achieved the intended changes" (p. 2,103).

Throughout this research the term ultrasound overlay videos (UOVs) will be used to refer to those videos that show a speaker's tongue movements in speech along with a profile view of the speaker's head (Bliss et al., 2017). UOVs are identified by Yamane et al. (2015) as those that display speech production by incorporating speech tongue movements taken through mid-sagittal ultrasonic images with externally profiled views of the head.

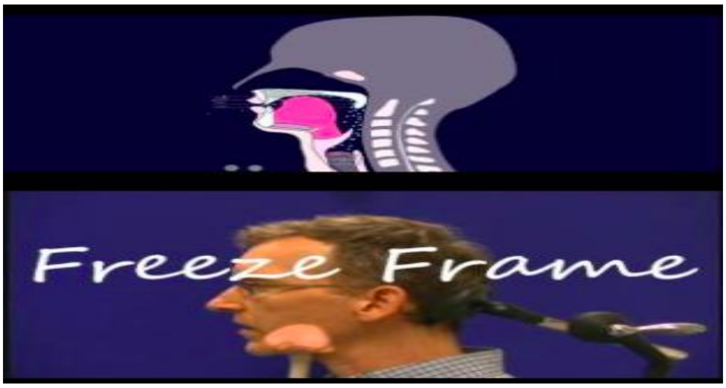

Figure 1 Screenshot of ultrasound overlay videos which explain with visual aid how linguistic sounds are articulated and provide understanding of articulatory differences between similar sounds. Source: https://enunciate.arts.ubc.ca.

UOVs could contribute to pronunciation instruction through both top-down and bottom-up teaching methods by demonstrating articulatory settings and enabling learners to view the tongue while producing the sounds (Gick et al., 2008). Furthermore, UOVs provide opportunities for constant listening and watching, thereby leading to improved pronunciation (Bliss et al., 2017). The research undertaken in this study seeks to investigate the effectiveness of UOVs on the pronunciation achievements of EFL learners.

II. Literature Review

The literature review will examine the main issues regarding ultrasound use in teaching pronunciation. It will focus on three major themes which emerge repeatedly throughout the literature reviewed. These are teaching novel and challenging sounds, teaching contrasts (two sounds with minimal phonetic difference), and controlling tongue movements. Despite the literature addresses these issues in a variety of contexts, this research will mainly target its application to non-clinical interventions. The use of ultrasonic technology in pronunciation adjustments, together with its benefits to pronunciation instruction, will be evaluated.

A. Teaching Novel and Challenging Sounds

With mixed results, a growing body of literature has investigated ultrasound efficacy in L2 pronunciation instruction. Meadows (2007) explored the potential of ultrasound in addressing common challenging sounds (/o:/, /e:/, and /ɴ/) for English speakers learning Japanese. In a mixed design, four native English speakers received four lessons of ultrasound-based and conventional instruction. The training involved instructional handouts and ultrasound real-time modeling of the sounds. The pre- and post-test productions, although not statistically different, were rated by native speakers as moving toward native-like attainment. This insignificance was possibly due to inadequate numbers of recorded tokens and training sessions. Moreover, with no valid experimental design with control and treatment groups, it is difficult to determine which instruction contributed to the general improvements of the target articulations.

Cleland et al. (2015) argued that ultrasonic visual feedback confers no advantage over traditional articulatory techniques in learning novel articulations. Their study compared the progress of 30 typically developing children randomised on study entry in learning non-English speech sounds with and without ultrasound visual biofeedback. The new articulations were taught through modelling, imitation, feedback and descriptions in addition to using ultrasound for the treatment group. No significant difference for ultrasound was observed between the two groups in the pre-test, teaching condition, or post-test except for the consonant /c/. This rather contradictory result may be attributed to the short 15- to 20-minute training.

In contrast to Cleland et al. (2015), Wu et al. (2015) reported a significant difference between treatment using ultrasound images followed by real-time tongue imaging and traditional, colloquial explanations of the French /r/. Two groups of five Chinese learners were involved in an experiment where preliminary perception tests confirmed their difficulty in identifying between French /r/ and Chinese /x/. Participants were recorded before and after training and better production of the phoneme was significantly found for the treatment group. The observed data confirmed that real-time ultrasound sped up the departure in articulation from Mandarin /x/ to French /r/ more efficiently than the traditional method. This does not appear to corroborate some previous studies (Cleland et al., 2015; Meadows, 2007) that suggested no benefit from ultrasound visualisation in improving L2 language learners' pronunciation. However, there was not a native French speaker involved to identify and rate the produced /r/.

In the same vein, White et al. (2017) examined the effectiveness of ultrasound visual feedback and traditional instruction in teaching dark /l/, an English syllable-final variant, to eight native Cantonese speakers. In an experimental design, two groups received identical instructions on the articulation of the target phoneme, but the experimental group also received ultrasound visual feedback lessons, through the ultrasound scanner, of almost 30 minutes in total. The post-test data highlight a significant improvement in the target sound, indicating a departure from back tongue to front tongue gesture to /l/, with an advantage for ultrasound training over non-ultrasound training.

B. Teaching Contrasts

In a one-session investigation, Gick et al. (2008) tested how ultrasound would potentially benefit the English approximant contrasts /l/ and /r/ in three native Japanese learners. The sounds were targeted word-initially, medially, and finally through six vowel contexts in a 1-hour session of ultrasound training, along with pre- and post-assessments. The training included ultrasound video recordings of participants' best and troublesome productions compared to images produced by the authors in terms of tongue shape and specific movements of tongue parts. Interestingly, the participants successfully produced the problem segments and generalised the gains in the post-assessment. The results indicated that even a short ultrasound period might have the potential for L2 speech production training. However, the observed improvements were likely caused by the phonetic knowledge the participants applied, as linguistics students, to the training. Moreover, they would have been more convincing if some variables (acquisition age, residency in an English-speaking country, and vowel contexts) had been controlled.

Similar to earlier findings, Tsui (2012) detected a strong evidence for the efficacy of direct ultrasound visual feedback to back up the lingual configurations of the liquids /l/ and /ɹ /, absent from Japanese phonology. Six native Japanese underwent four 45-minute training sessions over two weeks for these segments in different vowel contexts and word positions. The training involved ultrasound modelling and practice of the tongue positioning of /l/ and /ɹ /. The speech samples obtained two weeks after the training and rated by experienced listeners revealed more accurate productions than pre-training tokens. Most importantly, the gains obtained were maintained and generalised to two new words in the follow-up sessions. This appears to support the Noticing Hypothesis (Schmidt, 1990), where awareness is necessary to acquire L2 features, and with visual feedback, language learners can perceive phonemic differences that they would otherwise overlook. The validation of the results, however, was flawed by the absence of randomised and controlled trials and consistency in perceptual tasks. Additionally, the use of a repeated carrier sentence, although useful as a context, may have interacted with the productions of /l/ and /ɹ / or produced nonsensical sentences.

d'Apolito et al. (2017) study shows a significant advantage for articulatory training (ultrasound images of the tongue gestures and real-time biofeedback of the tongue position) over perceptual training in the acquisition of the American English contrast /ɑ–ʌ/. A comparison between the two was performed to determine their effects on the productions of nine Italian learners divided into three groups. Post-test outcomes showed improvement of the target contrast for the experimental group, suggesting that the 1-hour ultrasound session was more effective and stable than the perceptual training. This result agrees with previous findings in the literature (Gick et al., 2008; Tsui, 2012) in terms of short-term ultrasound training benefits on pronunciation.

These findings fit those of Tateishi and Winters (2013), who suggested that ultrasound training can improve the production of a non-native speech contrast. The researchers explored the significance of ultrasound visual feedback in the production and perception of the English contrasts /r/ and /l/ produced by 10 native Japanese speakers. After ultrasound images were taken for production, discussions began to promote participants' intellectual involvement in the training process and self-awareness of articulations. The five-session experiment led to more distinct productions of the phonemes, with a more native-like /l/ than /r/, but with no improvement noted in perception accuracy. This inconsistency might have been caused by the few training sessions and the reliance on production ability rather than perception tests to assess perception. However, these data must be interpreted with caution because of the few training sessions, and absence of randomised trials to validate the results.

Likewise, Pillot-Loiseau et al. (2015) investigated the effects of ultrasound visual feedback, at the word level, on the French contrast /y–u/ produced by seven Japanese learners. The subjects participated in a 12-week conventional pronunciation training with three additional 45-minute ultrasound lessons. Productions were recorded before and after training and between these recordings, ultrasound feedback of position and shape was given for the experimental group who were seen again two months after second recording. Productions improved after training and further improvements were noticed later. The researchers concluded that ultrasound visual feedback can change the production of challenging vowels over time depending on the utterance type. The results suggest that the adopted articulatory approach of production training may improve perception and, in this regard, contradict the earlier findings of Tateishi and Winters (2013).

C. Controlling Tongue Movement

Ouni's (2014) study investigated whether real-time ultrasound visual feedback would raise human awareness of tongue control. Using a randomised design, 24 native French speakers performed 12 tongue gestures of the three phonemes /a/, /i/, and /k/ in the absence of sound. After a 15-minute observation of real-time ultrasound visual feedback, the experimental group performed 10 out of 12 gestures, with no improvement for the control group on any of them. These findings match those of d'Apolito et al. (2017), Gick et al. (2008), Tsui (2012), and White et al. (2017) regarding the benefit of short sessions of practice with ultrasound on pronunciation.

The literature review shows the potential of ultrasound imaging technology in L2 pronunciation as a visualisation tool. While most of the works reviewed above agree on ultrasound's efficacy in acquiring, discriminating, or adjusting the tongue movement for novel and challenging speech sounds, others hold the view that it presents no advantage in learning pronunciation compared with other instructional contexts. A key problem with much of the literature regarding this issue is that most ultrasound imaging is difficult to employ in most teaching conditions due to the prohibitive cost of acquiring enough ultrasound equipment for large classes. The ultrasound is also difficult for learners to interpret without specialised training. Due to technological challenges, most studies have tended to focus on one-on-one settings, formal laboratory teaching, and adults rather than younger learners.

A closer examination of the literature also reveals a number of gaps and shortcomings. The generalisability of some published research is problematic, as it is limited to a small sample size (d'Apolito et al., 2017; Gick et al., 2008; Meadows, 2007; Pillot-Loiseau et al., 2015; Tateishi & Winters, 2013; Tsui, 2012; White et al., 2017; Wu et al., 2015). Ouni (2014) failed to provide adequate proof of his findings that pronunciation improvement is related to visual feedback. The empirical design of Meadows (2007), Tateishi and Winters (2013) and Tsui (2012) did not validate the results in randomised control trials. Moreover, d'Apolito et al. (2017), Gick et al. (2008) and White et al. (2017) did not consider long-term effects on knowledge at the phonological level.

III. Methodology

Within the specific context of young learners, this research seeks to answer the following questions:

1. What is the difference in the perception of /b/ and /p/ when pronunciation is taught with and without UOVs?

Hypothesis: High accuracy occurs with UOVs training.

2. What is the difference in the pronunciation of /b/ and /p/ when they are taught with and without UOVs?

Hypothesis: High accuracy occurs with UOVs training.

An important aspect in this study relates to these questions: The need to study the implementation of UOVs in pronunciation practice contexts, despite being in demand, is in its embryonic stages in terms of research and usage in primary education. The necessity to assess the effectiveness of UOVs as a resource of language teaching and learning is acknowledged by Bliss et al. (2017). The literature review chapter also identifies a gap in the existing knowledge in that, until recently, the impact of UOVs on the pronunciation of young learners has been unclear. This study seeks to contribute to this growing area of research by investigating the impact of UOVs on the pronunciation of young learners. According to the researcher's knowledge, no previous study has examined this area. The opportunity, therefore, to investigate their efficiency with young learners should significantly contribute not only to this field in general but also to a richer understanding of their broader effects on the young in particular.

A. Data Collection

(a). Site and Sample Selection

The study was conducted in a public primary school that controls any factors of language experience, making it suitable for the experiment. Primary public schools are known for their limited and late introduction of language classes compared with private schools. The students, specifically Grade 6, became the core of the recordings and perception quizzes. This group also provided an achievable approach to the experiment. Questionnaires, similar to those used by Tsui (2012) were provided to obtain information on the participants' characteristics and language experience.

Because of time constraints, a total of 10 students who speak Arabic as a native language were randomly split into two groups: the comparison group (CG) with no training and the experimental group (EG) with UOV training. In their questionnaires, all the participants reported no hearing problems or speech or language disorders. All the participants were aged between 11 and 12 (mean age = 11.30) and were female.

All the participants had no experience living in an English-speaking country. They were asked to rate their productions of the segments /p/ and /b/ in the questionnaire. The segments were rated on a four-point scale: (1) not at all on target, (2) somewhat on target, (3) almost on target, and (4) exactly on target. Regarding motivation to participate in the study, the questionnaire involved another four-point self-rating scale. The participants were asked to choose among (1) not motivated at all, (2) somewhat motivated, (3) extremely motivated, and (4) very motivated. Table 1 demonstrates the subjects' characteristics based on their answers to the questionnaire.

Table 1
Subjects' Characteristics Obtained From Questionnaires

Subjects	Age	Age of First Exposure	Living in English-speaking Country	Spoken English in Daily Life	Motivation
EG1	12	3 years	No experience	75%	Very motivated
EG2	12	9 years	No experience	25%	Extremely motivated
EG3	11	10 years	No experience	25%	Very motivated
EG4	11	4 years	No experience	50%	Extremely motivated
EG5	11	5 years	No experience	25%	Very motivated
CG6	11	10 years	No experience	25%	Somewhat motivated
CG7	11	6 years	No experience	25%	Very motivated
CG8	12	9 years	No experience	50%	Very motivated
CG9	11	7 years	No experience	50%	Very motivated
CG10	11	5 years	No experience	25%	Very motivated

Quantitative data were primarily collected through pre- and post-training recordings and perception quizzes. Perception quizzes were utilized since production interacts with perception (Flege, 1995) and recordings allowed for measuring production intelligibility and accuracy. This approach achieved triangulation at its simplest level by providing data from more than one standpoint (Cohen et al., 2002).

(b). Validity of the Perception Quiz

The Pearson coefficient was calculated to identify the validity of the study tool, whereas the correlation coefficient was calculated between every item and the total degree of the quiz, as shown in Table 2.

TABLE 2
PEARSON CORRELATION FOR THE QUIZ ITEMS OF /B/ AND /P/ PERCEPTION WITH THE TOTAL DEGREE OF THE QUIZ

Items	Pearson correlation	Items	Pearson correlation
1	.709**	8	.562**
2	.682**	9	.529**
3	.612**	10	.672**
4	.628**	11	.790**
5	.503**	12	.541**
6	.741**	13	.587**
7	.760**	14	.595**

** Correlation is significant at the 0.01 level

Tables 3 shows that all the statements are significant at the 0.01 level, meaning a high internal consistency as well as high and adequate validity indicators that can be trusted when applying the current study.

(c). Reliability of the Perception Quiz

To check the reliability of the study tool, a retest measure was used as shown in Table 3.

TABLE 3
RETEST FOR MEASURING THE STABILITY OF THE STUDY TOOL

	Number of items	Pearson correlation
Overall reliability	14	0.862**

Table 3 shows that the study quiz has a statistically acceptable stability. The value of the overall stability coefficient (Pearson correlation) was 0.862, which is a high degree of stability and indicates consistent results when applied to the present study.

(d). Equivalence of Groups

To determine group equivalence, the Mann–Whitney test was used to compare the mean scores of the EG and CG in the pretest, as shown in Tables 4 and 5.

TABLE 4
MEAN AND STANDARD DEVIATION OF THE PRETEST FOR BOTH EG AND CG IN THE PERCEPTION QUIZ OF /b/ AND /p/

Groups	*n*	Mean	Standard deviation
Control	5	6.80	2.17
Experimental	5	7.20	2.05

TABLE 5
MANN–WHITNEY TEST RESULTS OF THE PRETEST CONCERNING THE DIFFERENCES BETWEEN THE MEAN SCORES OF BOTH EG AND CG IN THE PERCEPTION OF /b/ AND /p/

Groups	*n*	Mean rank	Sum of ranks	*z*	*p*
Control	5	5.30	26.05	0.217	0.828
Experimental	5	5.70	28.50		

It is apparent from Table 5 that there are no statistically significant differences between the mean scores of the EG and CG in the pretest of /b/ and /p/ perception. The previous result indicates that there is consistency between the groups.

(e). Baseline Assessment

The assessment process, in both the pretest and posttest sessions, and the training were conducted in the school. One week prior to the commencement of the study, each participant's guardian signed a consent and participants were individually assessed through a number of tasks:

A. Perception quiz
B. Audio recording of word list #1
C. Audio recording of word list #2
D. Audio recording of word list #3

The participants identified /p/ or /b/ in 14 minimal pairs in a perception quiz. Each segment pair was depicted through pictures and displayed side by side in Microsoft Office PowerPoint 2016 and on sheets handed to the subjects. An audio recording by a native speaker was played for one word of the pair, and then the participants marked the image representing the spoken word. They were allowed to replay the recording whenever needed. To avoid orthographical interference, no written words were shown.

Once the quiz was completed, a list containing 24 words, 12 each for /p/ and /b/, was used for recording. The two segments were elicited in different word positions and vowel contexts and were presented initially, medially, and finally, with four words each. The vowel contexts included were low-front /æ/, mid-back /ʌ/, mid-central /ɜ:/, mid-front /e/, high front /i:/, and /ɪ/ as well as the diphthongs /eɪ/ and /əʊ/. Following this, each participant was recorded three times per word to increase the reliability of the elicited samples. This process of making a representative sample through three tokens per word was adopted from Lotto et al. (2004); Aoyama et al. (2004); and Tsui (2012). In the analysis, data from the third elicitation were used, given the increased familiarity with the word list and the reduced need for cueing.

Using Microsoft Office PowerPoint 2016, pictures of the chosen words were displayed. Afterwards, word elicitation was performed without any carrier phrases to avoid any interaction between the production of target sound and that of the carrier phrase. Again, to avoid the effects of orthography on pronunciation, the participants were shown the list only once to identify any unfamiliar words and later during elicitation whenever needed. The participants were recorded over two consecutive days because of the limited time available.

(f). Treatment Sessions

Each participant in both groups had four five-minute training sessions, with UOVs being provided for the EG and no intervention for the CG. The sessions were held over four consecutive weeks and included modeling, having the subjects practice the sounds using silent gestures and then their voice, and finally checking the word lists, with the provision of UOVs for EG. The first training session was composed of the following:

A. Instruction on the gestural configurations of /p/ and /b/ for both groups
B. Description of how UOVs work for the EG
C. Practice with both groups with UOVs being provided for the EG

Each participant in the EG received an informational illustration on how the technology works and an orientation of the tongue tip and root on UOVs. Furthermore, they were provided with a link to the University of British Columbia website (http://enunciate.arts.ubc.ca/) where they could review the videos. Afterwards, the training began with instructions on the lingual components of the segments accompanied by UOV demonstrations. To avoid list effects, target sounds with different word positions were distributed across the list (Gick et al., 2008). The participants had the opportunity to practice the learned articulation position in a pattern that progressively increased in difficulty, thereby helping to tailor training to their learning styles:

A. Segments in isolation without voice
B. Segments in isolation with voice
C. Segments in word-initial position in open syllables
D. Segments in word-initial position in closed syllables
E. Segments in word-final position in closed syllables
F. Segments in word-medial position

In every session, the EG had the opportunity to practice with UOVs. UOVs of the segments were displayed and the researcher demonstrated the correct configurations of the two. After that, participants practised the correct positioning of /p/ and /b/ in reading the lists. The progression through training stages depended on how the participants progressed in each training meeting. At the end of every session, lists of the words were handled to the participants for practice at home. They were instructed to practice the word lists up to a maximum of 10 minutes each day. As reported by them, the word lists were practiced between 2 and 10 minutes (mean = 6.6 minutes) between 1 and 10 days (mean = 2.9 days) from the last training session to the post-assessment.

(g). Post-Assessment

To assess the learned segments, participants were seen again individually two weeks after the training. As in Tsui's study (2012), the lingual components of the two segments were practiced for five minutes prior to the assessment. This stage involved the same perception quiz and word list included in the initial assessment as well as the same procedure.

(h). Follow-Up Assessment

The EG participants were seen 11 days following the post-assessment session for an individual follow-up evaluation. This final session included the same perception quiz and recording list.

B. Data Analysis

The analysis was a two-pronged approach of describing and analysing both the recorded word judgments and the perception quizzes for both groups. To judge the recorded tokens, two native English speakers were recruited as they are adept at detecting non-native pronunciation (Derwing & Munro, 2005). As Tsui's study (2012), productions were judged on a four-point rating scale. The choices included were first recording is better, second recording is better, both equally accurate, and both equally inaccurate. The pre- and post-training audio files of each target segment were inserted into Microsoft PowerPoint 2016 and the judges were asked to play the two words on each slide unknowing which assessment the audio file belonged to. By randomising the presentation of the words, the listeners were unable to identify the pre- or post-training tokens.

The judges were able to replay each word a maximum of three times to ensure thoughtful ratings. However, they were urged to make their judgments after the first listening. The listener's choice was recorded immediately. Each judge spent approximately 2 hours listening and judging 24 words per subject for pre-and post-assessment and 24 words per EG participant for the follow-up session. To judge perceptual ability, the scores of 14 minimal pairs in the pre- and post-assessments were included in the analysis. Only correct choices on the answer sheet were calculated for each participant. The number of correct choices of pre-assessment was compared to that of post-assessment. To measure the difference, the data were statistically analysed using the Mann–Whitney U and Wilcoxon tests as it is the best test for two independent samples that are small and not normally distributed (Nachar, 2008).

As for recordings, the results were examined by comparing the EG findings against the CG findings. A total of 24 words for each participant (12 pre-training and 12 post-training) in both groups were used in the analysis. The words included /p/ and /b/ in word-initial, word-medial and word-final positions two words each. The target words for the analysis were semi-randomly chosen to ensure that the final analysis included all word shapes (Tsui, 2012). Similarly, data analysis of the follow-up session followed the same procedure, comparing the post- and follow-up training gains of the EG.

Reliability

To measure the inter-rater reliability, a comparison was made between the two listeners' choices. An agreement was considered to exist whenever the two judges rated the same word as belonging to the same category identified above, while disagreement was noted when one judge's response for a word was different from the other one. The inter-rater reliability was 78% for both /p/ and /b/ words in the pre- and post-assessment and 76% for the follow-up session. According to Cohen et al. (2002), when agreement is between 64-81% of the data being analyzed, the level of agreement is strong (McHugh, 2012). Inter-rater disagreements were mostly in words rated as more accurate in post-training by a judge and equally accurate or inaccurate by the other. It was rare that the two judges disagree that a particular word would fit the category of pre-training more accurate or post-training more accurate.

Moreover, without the judges being informed, two tokens for each subject were replayed as soon as the judgments for all the listed words had been conducted to determine intra-rater reliability. The repeated words were 14% of the whole pre- and post-assessment tokens and 16% of the whole follow-up tokens. The intra-rater reliability in the pre- and post-assessment was 60% for Judge 1 and 85% for Judge 2. In the follow-up session, it was 80% for Judge 1 and 100% for Judge 2.

IV. Results

A. *Perception Quiz Results*

The perception quiz included 14 minimal pairs, and the analysis targeted subjects' correct responses in the quiz. The research in this regard sought to determine the effectiveness of UOVs in the perception of /p/ and /b/ through the following question: "What is the difference in the perception of /b/ and /p/ when pronunciation is taught with and without UOVs?" To determine the difference in the perception of /b/ and /p/ when pronunciation is taught with and without UOVs, the Mann–Whitney test was used, as shown in Tables 6 and 7.

Table 6
Mean And Standard Deviation Of The Posttest For The Difference In The Perception Of /b/ And /p/ When Pronunciation Is Taught With And Without UOVs

Groups	*n*	Mean	Standard deviation
Control	5	8.20	2.17
Experimental	5	8.20	1.30

Table 7
Mann–Whitney Test Results Of The Posttest Concerning The Differences In The Perception Of /b/ And /p/ When Pronunciation Is Taught With And Without UOVs

Groups	*n*	Mean rank	Sum of ranks	*z*	*p*	Eta squared
Control	5	5.70	28.50	0.213	0.831	0.0
Experimental	5	5.30	26.50			

Tables 7 and 8 reveal that there were no statistically significant differences in the perception of /b/ and /p/ when pronunciation was taught with and without UOVs. The mean score for both groups was 8.20. No effectiveness of UOVs for the perception of /b/ and /p/ was observed for female young learners. As shown in Table 7, the value of the eta squared in the posttest was 0.0. This did not surpass 0.14, the value indicating the educational importance of statistical results in psychological and educational research (Murad, 2000). The value of eta squared revealed no effect of UOVs on the perception of /b/ and /p/ for female young learners.

B. *General Results of the Accuracy of /p/ and /b/ Recordings*

Of all the /p/ productions, 26% were rated by both judges as more accurate in post-training, whereas 10% of the words were judged as more accurate in pre-training. Both judges rated 31% of the words as equally accurate and 10% as equally inaccurate. Accurate post-training productions of the segment /p/ were targeted in the analysis of both groups.

The next part of the recordings was concerned with /b/ productions. Both judges rated three words (5%) of all /b/ productions as more accurate in post-training. Also, only 1% across word positions was judged as more accurate in pre-training. Most of the ratings (48%) indicate an equal accuracy of both pre- and post-productions. Both judges rated 21% of all the /b/ tokens as equally inaccurate in both assessments. Accurate post-training productions of the segment /b/ were targeted in the analysis for both groups.

On the questions of "What is the difference in the pronunciation of /p/ when it is taught with and without UOVs?" and "What is the difference in the pronunciation of /b/ when it is taught with and without UOVs?" Tables 8, 9, 10 and 11 demonstrate the difference using the Mann–Whitney test.

TABLE 8
MEAN AND STANDARD DEVIATION OF THE DIFFERENCE IN THE PRONUNCIATION OF /p/ WHEN IT IS TAUGHT WITH AND WITHOUT UOV

Groups	*n*	Mean	Standard deviation
Control	5	1.80	1.48
Experimental	5	1.40	1.34

TABLE 9
MANN–WHITNEY TEST RESULTS CONCERNING THE DIFFERENCES IN THE PRONUNCIATION OF /p/ WHEN IT IS TAUGHT WITH AND WITHOUT UOVS

Groups	*n*	Mean rank	Sum of ranks	*z*	*p*	Eta squared
Control	5	5.80	29.0	0.328	0.743	0.02
Experimental	5	5.20	26.0			

TABLE 10
MEAN AND STANDARD DEVIATION OF THE DIFFERENCE IN THE PRONUNCIATION OF /b/ WHEN IT IS TAUGHT WITH AND WITHOUT UOV

Groups	*n*	Mean	Standard deviation
Control	5	0.0	0.0
Experimental	5	0.60	0.89

TABLE 11
MANN–WHITNEY TEST RESULTS CONCERNING THE DIFFERENCES IN THE PRONUNCIATION OF /B/ WHEN IT IS TAUGHT WITH AND WITHOUT UOV

Groups	*n*	Mean rank	Sum of ranks	*z*	*p*	Eta Squared
Control	5	4.50	22.50	1.491	0.136	0.12
Experimental	5	6.50	32.50			

The results shown in the tables indicate that there were no statistically significant differences between both groups' pronunciation of /p/ and /b/ when they were taught with and without UOVs. For /p/, the mean scores for the CG and EG were 1.80 and 1.40, respectively. As for /b/, the mean scores for the CG and EG were 0.0 and 0.60, respectively. These results reveal no effectiveness for UOVs in the pronunciation of /p/ and /b/ for female young learners. The value of the eta squared was 0.02 for /p/ and 0.12 for /b/. These did not surpass 0.14, the value indicating the educational importance of statistical results in psychological and educational research (Murad, 2000).

V. DISCUSSION

A. Perception Accuracy of /p/ And /b/

Learning to improve production of foreign sounds results in adjusting the perception of the same sounds (Kartushina et al., 2015) as each perceptual unit is a produced or intended gesture (Best, 1995). In pronunciation, both auditory and visual information greatly benefit L2 learners in terms of acquiring speech sounds and patterns (Bliss et al., 2016). The information provided by visual representations can serve as scaffolding and help to overcome perceptual limitations imposed on production (Kartushina et al., 2015). However, no evidence for UOVs' influence, as a visualizing tool, on female young learners' perception accuracy of /p/ and /b/ was found.

The accuracy score reveals that the treatment was unsuccessful to establish categories for the segments in participants' perception. A possible explanation for this result may be the lack of perceptual training of the targets. In this study, perception was only assessed but not trained; therefore, it was not possible to demonstrate that production training provided by UOVs positively affects the perception of /p/ and /b/. Even though these results differ from those of some other studies (Pillot-Loiseau et al., 2015) suggesting that ultrasound can improve perception, they are consistent with those of Tateishi and Winters (2013) showing that ultrasound has no positive influence on perception accuracy.

Given that these findings are based on a limited number of minimal pairs (14 minimal pairs in total), the results should be treated with the utmost caution. This small number limits an adequate representation of the subjects' actual ability. It is not inconceivable that different evaluations would have been obtained if a larger number of minimal pairs had been included in the assessments. Despite this, an important insight emerged from the perception quiz. As in Tsui's study (2012), none of the word-initial (WI), word-medial (WM), or word-final (WF) positions in which /p/ and /b/ appeared demonstrated a difficulty for perception.

B. Pronunciation of /p/ and /b/

Contrary to expectations, the results of using UOVs in the pronunciation instruction of /p/ and /b/ did not reveal significant differences between the two groups. This result matches the findings of Cleland et al. (2015) and Abel et al. (2016). There are several explanations for the lack of differences between the groups. In L2 acquisition, some theoretical perspectives assume that perception is necessary to produce new sounds or discriminate between L2 and L1 sounds and achieve accuracy through identifying phonetic features of a given sound or a contrast between two sounds (Best, 1995; Flege, 1995). As perception accuracy was not achieved in this study, it was possibly difficult for participants to perceive the difference between the two sounds, thereby leading to no significant improvements in production. This further supports the idea of Flege’s Speech Learning Model (1995), which state that no accurate L2 production surpasses perceiving L2 and L1 (dis)similarities.

Additionally, the lack of group differences was likely due to the subjects’ short training sessions (20 minutes in total). Participants also reported short intervals of homework practice with UOVs and with word lists provided throughout training of 2–10 minutes (mean = 6.6 minutes) for 1 to 10 days (mean = 2.9 days). However, in contrast to earlier findings by d’Apolito et al. (2017), Gick et al. (2008), Ouni (2014), Tsui (2012), and White et al. (2017), this research result showed that practice sessions, even short, with UOVs’ visual feedback had no benefit on the segments’ accuracy.

VI. Conclusion

In sum, the comparison between the two groups indicated no advantage of UOVs in the perception accuracy of /p/ and /b/. The evidence from this finding supports the idea that using the visual feedback of UOVs does not ameliorate perception accuracy for young learners. Contrary to previous findings, no effects of UOVs training on young learners were noted for the pronunciation of the targets. This leads to the conclusion that short UOV training and practice has no benefit on young learners’ pronunciation of the segments.

Importantly, the findings significantly differ from previous results reported in the literature in that the benefits of UOV training were not only unobserved but also not even retained at the 11-day follow-up. The follow-up, despite being soon after the training, also suggests that UOVs are ineffective for young learners to retain the segments over time. Together, these results provide important insights into using UOVs as a pronunciation tool. Within the context of young learners, the tool is not useful in teaching and discriminating between /p/ and /b/.

Limitations and Recommendations

With a small sample size, the findings of this study might not be transferable to larger populations. Future studies on the current topic with large sample sizes are recommended for all levels, including beginners, and not just advanced primary learners. Another barrier is the difficulty in accessing participants, particularly on busy school days, when they are needed. Therefore, the training duration, 20 minutes in total, was shorter than it was planned for as the subjects were seen during short intervals throughout the day. Such a duration is perhaps not long enough to determine the effectiveness of UOVs in pronunciation instruction for young learners. Therefore, future studies with long training sessions with UOVs combined with perceptual training to ameliorate production should be considered.

Although the results do not support the usage of UOVs for female young learners, more research is needed to evaluate the technology for other speech sounds and retention over time. In future investigations, it would also be interesting to undertake other computer-assisted pronunciation resources with challenging phonetic segments for English as a Foreign Language learners such as /ŋ/, /v/ and /tʃ/ as that topic may be more responsive to different tools. In addition, this research set out to compare UOVs instruction with traditional instruction. Further studies employing large randomised controlled trials and comparing UOVs with other computer-assisted pronunciation tools remain to be seen.

References

[1] Abel, J., Bliss, H., Gick, B., Noguchi, M., Schellenberg, M., & Yamane, N. (2016). *Comparing Instructional Reinforcements in Phonetics Pedagogy*. Paper presented at the Proc. ISAPh 2016 International Symposium on Applied Phonetics.

[2] Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), 233–250.

[3] Best, C. T. (1995). “A direct cross-realist view of cross-language speech perception,” in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York Press, Baltimore, MD), pp. 171–204).

[4] Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, *4*(1), 129-153.

[5] Bliss, H., Burton, S., & Gick, B. (2016). Ultrasound overlay videos and their application in Indigenous language learning and revitalization. *Canadian Acoustics*, *44*(3), 136-37.

[6] Bliss, H., Johnson, K., Burton, S., Yamane, N., & Gick, B. (2017). Using multimedia resources to integrate ultrasound visualization for pronunciation instruction into postsecondary language classes. *Journal of Linguistics and Language Teaching*, *8*(2), 47–62.

[7] Broughton, G., Brumfit, C., Pincas, A., & Wilde, R. D. (1993). *Teaching English As a Foreign Language*. London, United Kingdom: Routledge.

[8] Byun, T. M., Hitchcock, E. R., & Swartzb, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research*, *57*, 2116–2130.

[9] Çakır, İ., & Baytar, B. (2014). Foreign language learners' views on the importance of learning the target language pronunciation. *Journal of Language & Linguistics Studies*, *10*(1), 99-110.

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

[11] Cohen, L., Manion, L., & Morrison, K. (2002). *Research Methods in Education*: Routledge.

[12] d'Apolito, I. S., Sisinni, B., Grimaldi, M., & Fivela, B. G. (2017). Perceptual and Ultrasound Articulatory Training Effects on English L2 Vowels Production by Italian Learners. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 11*(8), 2120–2127.

[13] Damayanti, I. L. (2008). Is the younger the better? Teaching English to young learners in the Indonesian context. *Educare, 1*(1), 31-38.

[14] Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. G. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (Vol. 36, pp. 347-369). Philadelphia: John Benjamins Publishing Company.

[15] Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*, 379–397. doi:10.2307/3588486

[16] Elmahdi, O., & Khan, W. (2015). The Pronunciation Problems Faced by Saudi EFL Learners at Secondary Schools. *Education and Linguistics Research, 1*, 85. doi:10.5296/elr.v1i2.7783

[17] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 92*, 233–277.

[18] Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (Vol. 36, pp. 309–322). Philadelphia: John Benjamins Publishing Company.

[19] Gilakjani, A. P., & Ahmadi, M. R. (2011). Why is pronunciation so difficult to learn? *English Language Teaching, 4*(3), 74-83.

[20] Hameed, P. F., & Aslam, M. S. (2015). Pronunciation as a stumbling block for the Saudi English learners: An analysis of the problems and some remedies. *Theory and Practice in Language Studies, 5*(8), 1578–1585 doi:http://dx.doi.org.sdl.idm.oclc.org/10.17507/tpls.0508.06

[21] 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.

[22] Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley and Sons.

[23] Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition*. Houndmills, Basingstoke,Hampshire: Palgrave Macmillan.

[24] Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of/r/and/l. *From sound to sense, 50*(2004), C381-C386.

[25] McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica, 22*(3), 276-282.

[26] Moghaddam, M. S., Nasiri, M., Zarea, A., & Sepehrinia, S. (2012). Teaching pronunciation: The lost ring of the chain. *Journal of Language Teaching and Research, 3*(1), 215-219. doi:10.4304/jltr.3.1.215-219

[27] Murad, S. (2013). *Methods of Scientific Research "Designs and Procedures."* Cairo: Modern Book Publishing House.

[28] Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology, 4*(1), 13-20.

[29] Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer assisted language learning, 15*(5), 441-467.

[30] Ouni, S. (2014). Tongue control and its implication in pronunciation training. *Computer-Assisted Language Learning, 27*(5), 439–453.

[31] Pillot-Loiseau, C., Kamiyama, T., & Antolík, T. K. (2015). *French/y/-/u/contrast in Japanese learners with/without ultrasound feedback: vowels, non-words and words.* Paper presented at the International Congress of Phonetic Sciences (ICPhS) 2015.

[32] Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics, 29*(2), 191-215.

[33] Reid, E. (2016). Teaching English pronunciation to different age groups. In *Jazykovedné, literárnovedné a didaktické kolokvium xxxixi: Zborník vedeckých prác a vedeckých štúdií* (pp. 19-30): Bratislava Z-F LINGUA.

[34] Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics, 11*(2), 129-158.

[35] Tateishi, M., & Winters, S. (2013). *Does ultrasound training lead to improved perception of a non-native sound contrast? Evidence from Japanese learners of English.* Paper presented at the Proc. 2013 annual conference of the Canadian Linguistic Association.

[36] Tsui, H. M.-L. (2012). *Ultrasound speech training for Japanese adults learning English as a second language.* (Master Thesis), University of British Columbia,

[37] White, D., Gananathan, R., & Mok, P. (2017). *Teaching dark/l/with ultrasound technology.* Paper presented at the Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference, Calgary, Alberta, Canada.

[38] Wu, Y., Gendrot, C., Hallé, P., & Adda-Decker, M. (2015). *On Improving the Pronunciation of French/r/in Chinese Learners by Using Real-Time Ultrasound Visualization.* Paper presented at the ICPhS 2015 (18th International Congress of Phonetic Sciences).

[39] Yamane, N., Abel, J., Allen, B., Burton, S., Kazama, M., Noguchi, M., Gick, B. (2015). *Ultrasound-integrated pronunciation teaching and learning*. Ultrafest VII, Hong Kong.

Fatimah H. Alshehri was born in Saudi Arabia on May 10th, 1981. Mrs. Alshehri received a bachelor’s degree in English Language from King Khaled University in 2003 and a master’s degree in TESOL from King Saud University in 2020. She is currently pursuing PhD degree in Applies Linguistics with English Language Teaching from Southampton University.

She has a long experience in teaching English as a foreign language in Saudi Arabia and she works now in King Abdulaziz University. Her research interests include second language teaching, phonics and phonetics and teaching pronunciation to second language learners.