HELPING CHILDREN LEARN NON-NATIVE ARTICULATIONS: THE IMPLICATIONS FOR ULTRASOUND-BASED CLINICAL INTERVENTION

Joanne Cleland[1], James M.Scobbie[2], Satsuki Nakai[2] & Alan Wrench[3]

[1]University of Strathclyde, UK, [2]Queen Margaret University, UK, [3]Articulate Instruments Ltd., UK
joanne.cleland@strath.ac.uk; jscobbie@qmu.ac.uk; satsuki@ovod.net; awrench@articulateinstruments.com

ABSTRACT

An increasing number of studies are examining the effectiveness of ultrasound as a visual biofeedback device for speech production training or therapy. However, no randomised control trials exist. We compared the success of typically-developing children learning new articulations with and without ultrasound biofeedback. Thirty children aged 6-12 were randomly assigned to 2 groups: Group U were taught novel (non-English) consonants and vowels using ultrasound in addition to imitation, modelling, articulatory descriptions and feedback on performance. Group A were taught the same speech sounds, using the same methods but in the absence of ultrasound visual biofeedback.

Results showed that both groups of children improved in their production of the novel sounds with the exception of the high back vowels [u,ɯ]. No advantage for Group U was found, except for the palatal stop [c].

Keywords: Ultrasound, Visual Biofeedback, Speech Sound Disorders, Second Language Learning.

1. INTRODUCTION

Instrumental phonetic techniques were introduced into Speech and Language Therapy in the 1980s [3]. Techniques such as electropalatography (EPG) and ultrasound provide real-time dynamic visual biofeedback (VBF) of the articulators, which can be used to treat speech sound disorders (SSDs). While EPG has led the way as a VBF technique for speech therapy, ultrasound has also been used as VBF (Ultrasound Visual biofeedback: U-VBF) since around the same time [8]. Proponents of these techniques suggest that when people with speech disorders are able to *see* their own erroneous articulations, and modify them in real time, therapy outcomes are improved and previously intractable speech disorders are remediated. VBF fits well with theories of motor learning, providing "knowledge of performance" and hence allowing speakers to change and stabilise speech motor programmes [6]. It is also compatible with traditional articulatory and phonological therapy approaches [2] and can be used to demonstrate complex articulations that are normally difficult to describe.

Whilst EPG has been a popular technique in the lab, the costs and logistics of custom-made palates has been a barrier to adoption in the clinic. U-VBF overcomes some of these practical problems by being low cost and suitable for children at any stage of dental development. Both techniques, however, suffer from a lack of evidence of effectiveness. While a large number of case or small group studies have shown that EPG has positive outcomes, no randomised control trials (RCTs) exist. The evidence for U-VBF is even weaker, with around 18 studies in the literature, most often treating disordered /r/. RCTs are needed to evaluate the effectiveness of VBF, but they are difficult to design due to the heterogeneity of children with SSDs and difficulties with differential diagnosis.

This study modelled elements of a mini-RCT by using ultrasound to teach typically-developing children novel articulations in a pseudo-therapy context and comparing it with traditional articulatory techniques (i.e. motor-based approaches) to teach the same novel sounds. As well as being analogous to speech therapy, this may also have applications for second language learning.

Few studies use articulatory techniques for second language teaching, especially not in children (though there has been recent increased interest in Talking Heads and other technologies, see Golonka [4] for a review). Ouni [5] compared the ability of adult speakers of French to control simple tongue gestures, in the absence of sound, with and without U-VBF. Pre-test results showed that no participant was able to reproduce all of the 12 tongue gestures correctly. After 15-20mins of training with ultrasound the experimental group improved on 10/12 of the tongue gestures, whilst the control group made no improvement on any of the gestures. This suggests that even a short training session with ultrasound can result in positive changes. However, the method used by Ouni [5] differs from typical speech therapy in several ways. Firstly, the tongue gestures were dissociated from speech and secondly participants were not given feedback from the experimenter on their performance. In speech therapy VBF it is usual for the speech and language therapist to work alongside the client and provide feedback on performance. It is possible that if Ouni had done this the effects of the ultrasound training would have been greater. However, since the study used a “treatment verses no treatment” design (while the experimental group had 15mins of training the control group had a rest period) it is difficult to discern if ultrasound provides an advantage over articulatory training without U-VBF.

In this study, we sought to closely emulate a therapeutic context where children work alongside a speech and language therapist with positive reinforcement on performance. Moreover, we sought to establish whether U-VBF confers an advantage over more traditional articulatory techniques developed in the speech therapy clinic by comparing both types of feedback in a randomised design.

1.2 Aims

As a first step to determining whether U-VBF is helpful in teaching children new articulations, we designed a mini-RCT. We sought to determine whether typically-developing children can be taught to produce new consonants and vowels in a short training session with either U-VBF (Group U) or articulatory training (Group A). Our research questions were:

1. Can children imitate novel speech sounds accurately without training (Pre-test)? Hypothesis: low levels of accuracy, for both groups.
2. Are children able to imitate novel speech sounds following training (Teaching Condition) with U-VBF (U) or Articulatory training (A)? Hypothesis: Group U>Group A.
3. Can the children show evidence of retention of the speech sounds (Post-Test)? Hypothesis: Teaching>Post-Test>Pre-Test, for both groups.

2. METHOD

Data presented here is from a larger set of experiments designed to look at the effectiveness of U-VBF in typical speakers and children with SSDs. A subset of data reporting a holistic judgement of accuracy in typical children is reported here.

2.1 Speakers

Most clients in the speech therapy clinic are children and for this reason participants were 30 typical Scottish English children aged 6;0 to 11;8 (M=8.79, SD=1.56; Males=14). Children were randomised on entry to either Group U (n=14) receiving U-VBF, or Group A (n=16) receiving articulatory teaching.

2.2 Materials

Table 1 shows the non-Scottish English speech sounds taught to both groups of children. Speech sounds were selected to be easily discernible on mid-sagittal ultrasound allowing us to predict an advantage for Group U. Model articulations were provided by a female phonetician ultrasound recorded saying each of the speech sounds in isolation (or with a schwa following the stops) and in context of [aCa] or [dV]. Audio was extracted from the videos for Group A as a model. Real words providing comparison articulations were recorded from the children using orthographic (or occasionally verbal) prompts: *home, he, who, huge, Sam, sham, tap, cap, chap, gap, bap*.

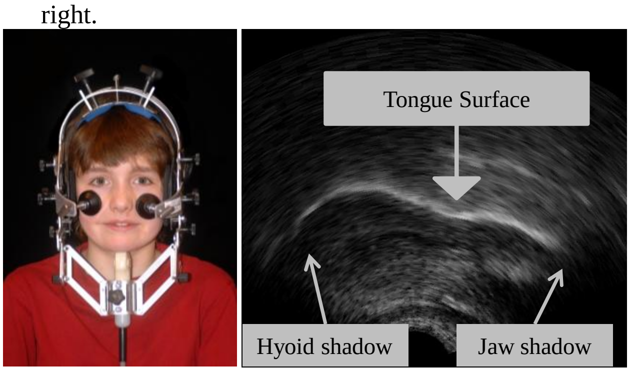
Table 1: Novel (Non-English) Speech sounds

Sound	Tongue-shape to be different from:	Possible target:
u	“who”, [ʉ]	Higher than “home”, [o]
ɯ	“who”, [ʉ] , “he”, [i]	Higher than “home”, [o]; Spread lips
y	“who”, [ʉ]	“he”, [i]
c	“cap”, “tap” [t], [k]	“huge”, [ç]
ʂ	“Sam”, “Sham” [s], [ʃ]	(clear retroflexion on UTI)
b͡g	“bap”, [b]	“gap”, [g]; Lip closure

2.3. Ultrasound Set-Up

Ultrasound data was acquired from both groups using an Ultrasonix SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced software™ [1]. The echo return data was recorded at 121fps with a 112.5 degree field of view (FOV) in the mid-sagittal plane. Fig 1 shows a typical ultrasound image (right). Speakers were fitted with a headset (Fig. 1, left) to stabilize the ultrasound probe. Simultaneous acoustic and lip-camera recordings allow us to identify lip-rounding.

Figure 1: Ultrasound head-set (left); typical ultrasound image (right) with tongue-tip to the right.



2.4. Procedure

Both groups of children underwent three different tasks or conditions: pre-test imitation; a teaching condition where novel speech sounds were taught; and post-test imitation where no feedback on performance was provided by the SLT. Group A was blinded to the ultrasound display during all conditions whilst Group U viewed the ultrasound at all times following a brief orientation to the display. The whole procedure took around 35mins.

2.4.1 Pre-Test Imitation Task

The children heard each segment once accompanied by either the ultrasound video (Group U) or a visual symbol (Group A) and were asked to imitate it immediately. First the segment in isolation was played, then their response was recorded and then the same segment was played in context [aCa] or [dV]. All of the consonants were recorded first, followed by all the vowels.

2.4.2 Teaching Task

Both groups were taught each segment in turn by the same SLT and their best attempt (as judged on line by the SLT) was recorded. Group U was able to watch the ultrasound video/audio model as many times as they wanted and could see the target tongue shape (a still frame) at all times. The SLT gave the children feedback on the correctness of their productions including tongue-shape, lip-shape and acoustic characteristics. In addition to the biofeedback, articulatory description and shaping from phonetically close speech sounds was used. Table 2 shows the types of prompts used for each segment and the order in which they were taught.

Table 2: Prompts

	Instructions/prompts
ʂ	Curl your tongue tip behind your top teeth. Put tongue in position for /r/ (where the child used a retroflex /r/) and blow
c	Make a sound between k and t. Put your tongue in position for [j], but make a short sound
b͡g	Put your tongue in position for /g/ and imagine you are saying /b/ and /g/ at the same time
y	Say /i/ and round your lips, keeping your tongue steady
u	Try to move your tongue as high up towards the back of your mouth as you can. Say [w] and try to achieve a similar tongue shape OR say [o] and slide your tongue back
ɯ	As for the previous vowel, but with smiley lips

Group A were taught the same speech sounds; both child and therapist were blind to the UTI at all times. They were allowed to hear the audio model as many times as they wanted and a distinct visual symbol (e.g. a “b+g” for the double articulation) for each segment was visible at all times. Just like Group U, children in Group A were given articulatory descriptions of the novel sounds summarised in Table 2.

The children’s best attempt was recorded and they were then asked to produce the sound between vowels [aCc] or after a stop [dV] using similar teaching techniques.

2.4.3 Post-Test

Both groups then repeated the Pre-test imitation task to determine whether they had retained the speech sounds. During this imitation task no feedback on correctness was given and the children were only allowed one attempt.

2.5 Annotation/Analysis

Using AAA software [1] each vowel and fricative was annotated at its acoustic midpoint. Stops were annotated at the midpoint of closure using the ultrasound and acoustic data. For each segment, the nearest ultrasound frame to the midpoint was selected and a spline indicating the tongue surface fitted to the image using the automatic edge tracking function in AAA software [1]. AAA allows multiple splines to be exported to a “workspace” to allow direct comparison of tongue shapes.

Each child’s best attempt at a segment was judged as correct or incorrect (1 or 0) by a phonetician blind to the grouping. Judgements were based on comparisons between ultrasound splines for an individual child’s English phonemes versus the image at the midpoint of the articulation of children’s attempts at the novel speech sounds and the video images of lips (to judge roundness for the vowels and lip-closure for the double articulation). Where multiple attempts at a target were recorded the “best” token was annotated, if any. If not, the last attempt was annotated.

3. RESULTS & DISCUSSION

Both groups pre-test were not able to accurately imitate non-English speech sounds: only 5% were judged as on target. However, after only around 20 minutes of teaching, performance improved significantly ($t(28)=-16.75$; $p<.0005$) to 55%. This suggests that it is possible to teach children novel articulations in a very short amount of time. This gain was somewhat maintained in the post-test condition, where the children were 32% accurate, again significantly more so than pre-test (t(28)=10; p<.0005). It is probable that further teaching would have led to consolidation of the speech sounds, in line with research on second language learning [9].

Figs 3 and 4 show the results for consonants (3) and vowels (4) per group. For both types of segment there was no significant difference between groups in either the pre-test, teaching or post-test condition (significance levels included in the figures), suggesting that ultrasound did not confer an advantage in our tasks.

Figure 3: Consonant results for groups U=U-VBF and A=Articulatory teaching.

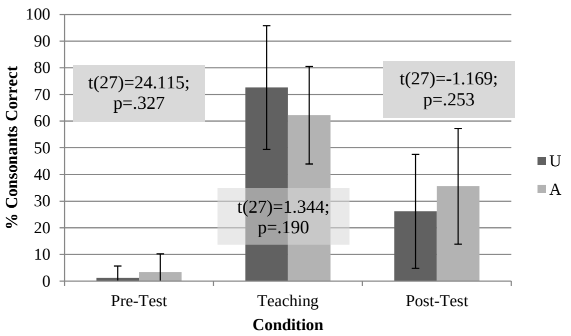
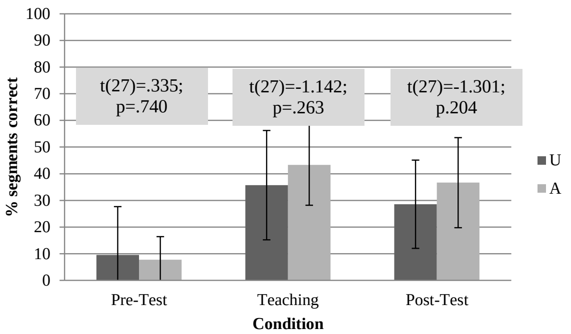


Figure 4: Vowel results for groups U=U-VBF and A=Articulatory teaching.



However, consonants were produced more accurately than vowels in the "teaching" condition (t(28)=-5.322; p<.0005) despite vowels being more accurate pre-test (imitation, t(28)=2.635; p=.014). This difference was not maintained in the post-test condition (t(28)=.361; p=.72). Studies of articulatory therapies show that targeting vowels is problematic because a lack of tongue-palate contact in vowels leads to reduced somatosensory feedback. This was borne out in this study, despite the addition of U-VBF, which images vowels easily. Additionally, closer inspection revealed that the increased accuracy in vowels (both groups) was due to a high success rate with [y] (7% on target in pre-test, rising to U=86% and A=93%). This segment differs from the other two vowels [u, ɯ] in two key ways. Firstly, high front vowels such as [y] involve significant tongue-palate contact, enhancing somatosensory feedback, and secondly, this vowel made use of a tongue shape, [i], already in the children's inventories (no true close back vowel exists for Scottish English, /u/ is fronted [7]).

In terms of specific consonants, an advantage was found for the ultrasound group in the teaching condition only of the palatal stop (t(27)=2.231; p=.03). Again, the tongue shape for this segment was already in the children's inventories (in [ç] in "huge") and it is possible the ultrasound enabled this group to more accurately achieve a similar tongue-shape for the stop since they were instructed to base the tongue-shape on [ç] or [j]. This advantage would not be predicted for [y] where both groups had access to visual information: lip rounding.

4. CONCLUSIONS

As predicted, children were able to approximate non-English speech sounds within a short teaching session, however, contrary to our expectations ultrasound did not provide an advantage, except for [c]. The children were largely unsuccessful at learning completely new tongue shapes and/or accessing areas of their articulatory space which they were unfamiliar with (close back vowels), but with extra time and training this should be possible [8]. Given extra time ultrasound may have accelerated this process.

Contrary to the growing body of literature in SSDs [2,6] U-VBF was no more effective than articulatory training. Whilst this highlights the importance of conducting large RCTs, with control arms of competing therapy approaches, it would be unwise to conclude that ultrasound will not be found to be beneficial in children with SSDs. While highly structured studies like this show that non-clinical populations can help address specific, detailed questions about U-VBF, further research on the nature of typical vs. clinical learning strategies is required. Children with SSDs selected for VBF differ from typical children in that they tend to have a history of *persistently* being unable to achieve new tongue-shapes, i.e. persistently substituting /r/ with [w]. Typically children receive multiple sessions of training, suggesting that in the current study children required further exposure from the U-VBF to produce accurate articulations of new tongue shapes. It is clear that we need clinical studies involving real cases which are not merely treatment/no-treatment, and yet which report realistic therapy. Future studies should aim to determine firstly whether or not U-VBF is more effective than other treatments and if so, what the required dosage is.

5. ACKNOWLEDGMENTS

This work was supported by EPSRC grant number EP/I027696/1. Thanks are due to our participants and Steve Cowen for technical assistance.

6. REFERENCES

1. Articulate Instruments Ltd. (2011). Articulate Assistant Advanced User Guide: Version 2.13. Edinburgh: Articulate Instruments Ltd.
2. Bernhardt, B., Gick, B., Bacsfalvi, P. & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics and Phonetics*, 19, 605 – 617.
3. Dagenais, P. (1995). Electropalatography in the Treatment of Articulation/Phonological Disorders. *Journal of Communication Disorders*, *28,* p303-329.
4. Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2012). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70-105.
5. Ouni, S. (2011). Tongue gestures awareness and pronunciation training. In *12th Annual Conference of the International Speech Communication Association-Interspeech 2011*.
6. Preston, J., Brick, N. & Landi, N. (2013. Ultrasound Biofeedback Treatment for Persisting Childhood Apraxia of Speech. *American Journal of Speech-Language Pathology*, *22,* p627-643.
7. Scobbie, J.M., Stuart-Smith, J., Lawson, E. (2012). Back to front: a socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Rivista di Linguistica / Italian Journal of Linguistics, 24,* 103-148.
8. Shawker, T. H. & Sonies, B. C. (1985). Ultrasound biofeedback for speech training. Instrumentation and preliminary results. *Investigative Radiology, 20,* 90-93.
9. Taimi, L., Jähi, K., Alku, P., & Peltola, M. S. (2014). Children Learning a Non-native Vowel– The Effect of a Two-day Production Training. *Journal of Language Teaching and Research*, *5*(6), 1229-1235.