# ARTICULATION THERAPY FOR CHILDREN WITH CLEFT PALATE USING VISUAL ARTICULATORY MODELS AND ULTRASOUND BIOFEEDBACK

Z. Roxburgh [(1)], J. M. Scobbie [(1)], and J. Cleland [(1), (2)]

[(1)] CASL, Queen Margaret University, Edinburgh
[(2)] University of Strathclyde, Glasgow

## ABSTRACT

Visual biofeedback tools, such as Electropalatography (EPG), are recommended for assessing and treating speech sound disorders (SSDs) associated with Cleft Palate (CP). However, EPG is not suitable for all clients, due to dependencies on stable dentition and timing of palatal repair. Ultrasound is becoming increasingly popular for its use in treating SSDs, with no reports on its dependency on structure of the vocal tract. However its clinical application in the CP population remains to be tested.

We compared Visual Articulatory Models (VAMs) with Ultrasound for the treatment of SSDs in two children with repaired submucous CP. Both children received two blocks of therapy each with eight sessions, with the first block using VAMs and the second using ultrasound. Results showed that both children improved overall, with more improvement found in the first block of therapy using VAMs.

**Keywords**: Ultrasound, Visual Articulatory Models, Phonetic Transcription, Cleft Palate

## 1. INTRODUCTION

Visual biofeedback (VBF) techniques have frequently been used in the assessment and treatment of speech production for speakers with Cleft Palate (CP). Electropalatography (EPG) is recommended in the Royal College of Speech and Language Therapy (RCSLT) Clinical Guidelines [15] for assessing and treating Speech Sound Disorders (SSDs) associated with CP.

Studies have investigated the benefits of using instrumental techniques, such as EPG, for identifying compensatory articulations in speakers with CP. Errors such as *increased tongue-palate contact, retraction to velar or glottal placement, fronted placement, complete closure (loss of grooving), open pattern, double articulations, increased variability* and *abnormal timing* [8]. By using instrumental techniques such as EPG, clinicians are able to identify errors which are not provided by phonetic transcriptions alone. Although phonetic transcription is deemed *gold standard* in identifying speech errors in CP, issues with phonetic transcription alone have been previously noted in the literature [16]. Inaccuracies in phonetic transcriptions can result in misdiagnosis and subsequent inappropriate intervention [10].

However EPG also has drawbacks, with each participant requiring an individual palate at a cost of around £400 (~€520). EPG is also not suitable for all speakers with CP, due to requirements of secondary surgery, ongoing dental, orthodontic and maxillary input.

Ultrasound Tongue Imaging (UTI) is a similar technique. With no costs for individualised hardware required and no need for stable dentition, UTI seems an obvious choice as an alternative to EPG for the CP population.

UTI has been used as a VBF tool since around 1980 [17]. More recently, UTI has been used to investigate compensatory articulations in speakers with CP [5], [9], with quantitative measures proposed to analyse UTI data of compensatory articulations [19]. However its clinical application remains to be tested.

Although UTI provides VBF, it does not provide a context for lingual tongue movement. Visual Articulatory Models (VAMs) [3], [11], [12] provide a context for lingual patterns by demonstrating dynamic tongue movement against passive articulators. They also provide additional information on velopharyngeal (VP) mechanisms and voicing. Recent advances in technology allow for such models to be commercially available on mobile devices, such as iPads [1].

Speech Trainer 3D [18] offers a VAM for iDevices [1] at a low cost for clinicians and clients. It provides dynamic animated videos for all English Phonemes and American-English vowels. As well as dynamic videos, it allows users to focus on particular aspects of speech production (e.g. lip movement, tongue movements, oral/nasal airflow and voicing) by providing an explanation of processes to achieve a particular phoneme. However, this VAM is not based on any anatomical data and phonetic inaccuracies, such as timing and place of articulation, are evident.

## 1.2. Aims

We aimed to test and compare the clinical application of Speech Trainer 3D and Ultrasound in the treatment of SSDs associated with CP. We aimed to answer the following questions:

1. Will children's speech improve post-therapy when using VAMs and U-VBF?
2. What are the qualitative differences between the two tools presented?

## 2. METHODOLOGY

We present two single-case studies of two males with repaired submucous cleft palate. Both children received six assessment sessions: two initially, two finally, and two interspersed between a block of eight sessions of VAM therapy followed by a block of eight sessions of ultrasound visual biofeedback (U-VBF) therapy.

### 2.1. Participants

*Andrew (pseudonym), Hemifacial Microsomia (lower part of one side of the face is underdeveloped) with Microtia (under developed external portion of the ear) and a mild unilateral conductive hearing loss (HL), 9;2 years:* Backing /n/ to [ɲ] or [ŋ] with suspected double articulations at referral. He had previously received extensive therapy to target his production of /n/, with no success.

*Craig (pseudonym), 6;2 years*: Few high pressure consonants, backing /k/ to glottal placement and fronting /g/ to [d] or [n], with possible double articulations at referral. Craig had not received any previous therapy on production of velars. Previous therapy had targeted bilabial consonants and alveolar fricatives, which had not resolved at the time of referral.

### 2.2. Recording Procedure

All assessment sessions were recorded with simultaneous ultrasound, audio and lip-camera. Prior to the first block of therapy, the SLT was blinded to the ultrasound data so not to influence the treatment choices in the VAM condition. Ultrasound data was acquired using an Ultrasonix SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced™ software [2] version 2.14 which internally synchronised the ultrasound and audio data. The echo return data was recorded at ~121 frames per second (fps), i.e. ~8ms

per frame with a 135 degree field of view (FOV) in a mid-sagittal plane.

A bespoke version of AAA was developed to allow us to use the software for therapy. This included features such as saving and calling up target tongue-shapes based on an individual's own productions and quick playback of participant's attempts at articulations during therapy or for analysis afterwards.

At each of the assessment sessions the participants completed the phonology subtest of the Diagnostic Evaluation of Articulation and Phonology [7]. Following this, an untreated wordlist/s targeting each child's specific lingual errors was selected from a battery. Post-therapy, a wordlist containing treated words was also recorded.

### 2.3. Analysis

Narrow phonetic transcriptions were performed using the acoustic and lip-camera data pre-therapy and the acoustic, ultrasound and lip-camera data post-therapy. This allowed us to calculate a Percent Consonant Correct (PCC) and Percent Target Consonant Correct (PTCC) at each time point of assessment.

Using AAA software [2], plosives were annotated at the burst and vowels, fricatives and nasals were annotated at their acoustic midpoint. For each segment, the nearest ultrasound frame to the burst or midpoint was selected and a spline was fitted using the automatic function in AAA. A qualitative analysis of the overall quality of the ultrasound data was also carried out. Lip-camera data was analysed by looking at the images at the burst or midpoint to identify aspects of speech production which was not identified through acoustic or ultrasound data, e.g. linguolabial or interdental productions or excess lip rounding.

PCC and PTCC scores derived from using only acoustic and lip-camera data were compared to PCC and PTCC scores derived from using added ultrasound data to identify whether more errors were identified when instrumental analysis was implemented.

## 3. THERAPY

The first block of therapy used an iPad app Speech Trainer 3D [18] as a VAM and the second block of therapy used U-VBF. Typically, the first 30mins of each session focused on using either Speech Trainer 3D or Ultrasound, and the second 30mins on traditional table-top activities, such as minimal pairs [4], focusing on the same target phoneme to build in generalisation.

### 3.1. Therapy Block 1 - Articulatory Animations

The first and second therapy sessions for each child used Speech Trainer 3D to demonstrate the constituent parts of the vocal tract and to demonstrate and label the phonemes of English being targeted. Both children were able to label parts of the vocal tract and label target phonemes (e.g. "a /k/ is a back (velar), quiet (voiceless), mouth (oral) sound") within the first two sessions. Both children were able to discriminate between alveolar and velar plosives and nasal stops during visual discrimination tasks within two sessions. Production practice was individualised, but followed a motor-based approach similar to Preston [13]. As both children were not stimulable for the target articulation, therapy began by targeting the phonemes they were able to achieve and describing and modelling the differences between these phonemes and the target phonemes.

### 3.2. Therapy Block 2 – Ultrasound Visual Biofeedback (U-VBF)

As shown in [6], the first therapy session focused on learning to associate the movement of the ultrasound image on the screen with the movement of their own tongue by demonstrating tongue shapes already in their inventory. During this first session, ultrasound was also used as VAM and ultrasound images were compared to those in Speech Trainer 3D. As in the first block of therapy, production practice was individualised and followed a motor based approach. Again, therapy began by using ultrasound to reinforce correct productions. As both children were able to produce a perceptually acceptable target within the first block of therapy, in at least one condition (e.g. VC or CV), their own best attempt from the pre-therapy assessment recording was used as a target tongue-shape.

## 4. RESULTS

Fine phonetic transcriptions show that both children made improvements in PTCC overall, with higher increase in PTCC scores during the first block of therapy using Speech trainer 3D.

### 4.1. Andrew

Analysis of his untreated wordlist pre-therapy showed that Andrew was consistently substituting /n/ with [ŋ] in Word Initial (WI), Word Medial (WM) and Word Final (WF). Analysis of the DEAP also showed that he was inconsistently velar fronting and had /s/ distortions, whereby he produced /s/ as [ʃ]. Post-therapy, Andrew was able to achieve [n] in WI position in /no/. In the maintenance session,

Andrew was still inconsistently velar fronting and had /s/ distortions. Andrew also had VP friction on production of fricatives. Figures 1 and 2 present Andrew's PCC and PTCC scores across all assessment sessions, showing very little improvement overall.

**Figure 1:** Andrew's DEAP Phonology PCC Scores (grey sections denote blocks 1 and 2 of therapy)
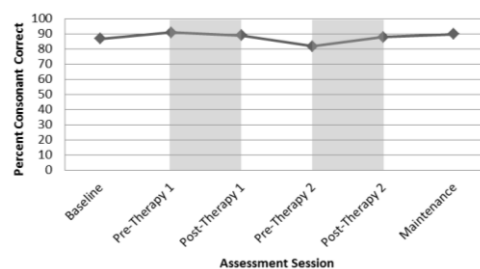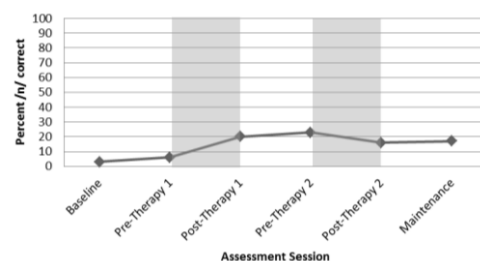


**Figure 2:** Andrew's PTCC Scores (grey sections denote blocks 1 and 2 of therapy)



### 4.2. Craig

Narrow transcriptions of Craig's data showed that he was making multiple speech errors pre-therapy. Table 1 presents the errors identified by the DEAP pre-therapy.

**Table 1:** Errors identified by the DEAP at Baseline

| Process | Examples | Consistency |
|---|---|---|
| Deaffrication | /ʤ/ → [d] | 1 |
| Cluster Reduction | /st/ → [d] /sk/ → [ʔ] /kɹ/ → [ɹ] /sb/ → [(s)b] /skw/ → [w] | 5 |
| Fronting | /s/ → [θ] /z/ → [ð] /ʃ/ → [θ] /g/ → [d] /k/ → [d] | 15 |
| Stopping | /s/ → [d] /z/ → [d] /v/ → [b] | 4 |
| Voicing | /f/ → [v] /p/ → [b] /t/ → [d] | 9 |

Analysis of his untreated wordlist showed that Craig's production of velar plosives was inconsistent. He would inconsistently front velar plosives to [d] (WI) or [n] (WM) or back to glottal placement (WM or WF). Alveolar [t] was consistently produced as a glottal stop. Possible double articulations were identified through perceptual analysis.

Within his maintenance session, Craig was able to achieve velar plosives, although production remained inconsistent. As he was able to achieve more high pressure consonants, this resulted in increased VP friction on production of velars. Fig 4 shows that DEAP scores increased with more accurate productions of velars and alveolars treated in the study; and bilabials and ladiodentals, treated in parallel by his usual SLT. Figures 3 and 4 present Craig's PTCC scores across the assessment sessions.

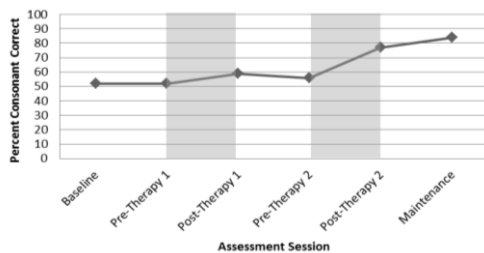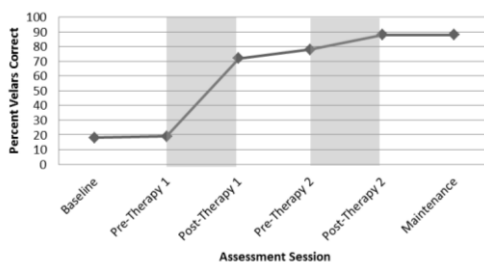**Figure 3:** Craig's DEAP Phonology PCC Scores (grey sections denote blocks 1 and 2 of therapy)



**Figure 4:** Craig's PTCC Scores (grey sections denote blocks 1 and 2 of therapy)



## 5. DISCUSSION

Results show that both of the children's PTCC decreased post-therapy using ultrasound, with phonetically trained listeners agreeing that pre-therapy tokens sounded "closer to the English target" than post-therapy when using U-VBF [14].

Speech Trainer 3D provides an animated articulatory model with a context for lingual patterns by displaying dynamic tongue movements in relation to passive articulators. However, there are inaccuracies in the model, which is not based on anatomical data. Despite these inaccuracies, both children made improvement within the first block of therapy with the animated model, highlighting that perhaps models only need to approximate what the vocal tract might be doing, rather than being anatomically correct. Within one session, both children were able to label areas of the vocal tract, label speech sounds and describe their own speech patterns. When moving on to using U-VBF, both children were able to identify tongue shapes using ultrasound within a few sessions. However, when using ultrasound as a biofeedback tool both children found this to be a difficult concept and did not understand how to use the biofeedback to alter their tongue shape. Although the children may not have benefitted from the biofeedback directly, it was useful for the SLT implementing therapy to provide positive reinforcement and auditory feedback alongside the visual biofeedback, allowing the SLT to adapt therapy and tailor to the needs of each child.

However, ultrasound images for both children were of poor quality, due to their anatomy, and were difficult to interpret both during therapy and for analysis. U-VBF is proposed as an alternative to EPG, which relies on stable dentition and timing of palatal repair. However, it should be noted that anatomical features, such as the size and shape of the jaw, scarring on the hard palate and velum and facial symmetry should be accounted for when selecting clients for longitudinal ultrasound recordings. Both of the children had small jaws, which resulted in missing tongue tip data due to the mandible shadow. Andrew had Hemifacial Microsomia, which made fitting the headset difficult at times. Craig had a very small head, making-probe stabilisation difficult as his head was too small for the Articulate Instruments headset, which is designed primarily for adults. As the headset was too big, or was not able to sit straight on both children, issues with probe movement were not resolved.

## 6. CONCLUSION

This paper has presented a qualitative comparison of two visual feedback tools used alongside articulation therapy for two children with repaired submucous cleft palate.

Results show that both children made overall improvement in PTCC, with a higher increase in scores when using Speech Trainer 3D, suggesting that VAMs show promise as an effective tool. However, it should be noted that this study used a very small sample size, with both children having a repaired submucous CP, not taking into account other types of CP. Future studies should include a range of cleft types, taking into consideration anatomical features such as facial symmetry and the size of the jaw when selecting suitable participants for future ultrasound studies.

## 8. REFERENCES

[1] Apple. 2012. The new iPad. [online] Available at: http://store.apple.com/uk/browse/home/shop_ipad/family/ipad [Accessed March 16 2012].

[2] Articulate Assistant Advanced Ultrasound Module User Manual, 2013. Revision 215, Articulate Instruments Ltd.

[3] Badin, P., Tarabalka, Y., Elisei, F., Bailly, G. 2010. Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*. Vol 52. pp. 493-503.

[4] Barlow, J. A., Gierut, J. A. 2002. Minimal pair approaches to phonological remediation. *Seminars in Speech and Language*, *23(1)*, 57-67.

[5] Bressmann, T., Radovanovic, B., Kulkarni, G. V., Klaiman, P., Fisher, D. 2011. An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics*. 25(11-12). pp.1028-1033.

[6] Cleland, J., Scobbie, J.M. & Wrench, A., (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics and Phonetics*. Pp. 1-23.

[7] Dodd, B., Hua, Z., Crosbie, S., Holm, A., Ozanne, A. 2002. *Diagnostic Evaluation of Articulation and Phonology.* London: The Psychological Corporation.

[8] Gibbon, F.E. 2004. Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. *Clinical Linguistics & Phonetics*. Vol 18(4-5). pp.285-311.

[9] Gibbon, F.E., Wolters, M. 2005. A new application of ultrasound to image tongue behaviour in cleft palate speech. *Poster presentation at the Craniofacial Society of Great Britain and Ireland Annual Scientific Conference,* Swansea, UK (13-15 April 2005).

[10] Howard, S. 2004. Compensatory articulatory behaviours in adolescents with cleft palate: comparing the perceptual and instrumental evidence. *Clinical Linguistics & Phonetics.* Vol 18. pp.313-340.

[11] Kroger, B.J., Gotto, J., Albert, S., Neuschaefer-Rube, C. 2055. A visual articulatory model and its application to therapy of speech disorders: a pilot study. *ZAS Papers in Linguistics.* Vol.40: pp 79-94.

[12] Kroger, B.J., Graf-Borttscheller, V., Lowit, A. 2008. Two- and Three-Dimensional Visual Articulatory Models for Pronunciation Training and for Treatment of Speech Disorders. *ISCA.* Brisbane, Australia. September 22-26. pp. 2639-2642.

[13] Preston, J., McCabe, P., Rivera, A., Whittle, J., Landry, E., Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research.* 57(6). pp. 2102-2115.

[14] Roxburgh, Z., Cleland, J., and Scobbie, J.M. Multi-listener evaluation of articulatory intervention with visual articulatory models and ultrasound biofeedback in two children with repaired cleft palate. Under Review.

[15] Royal College of Speech and Language Therapists (RCSLT). 2005. *RCSLT Clinical Guidelines.* Oxon: Speechmark Publishing.

[16] Sell, D. 2005. Issues in perceptual speech analysis in cleft palate and related disorders: a review. *International Journal of Language and Communication Disorders*. Vol 40(2). pp.103-121.

[17] Shawker, T. H., Sonies, B. C. 1985. Ultrasound biofeedback for speech training. Instrumentation and preliminary results. *Investigative Radiology*. Vol 20, 90-93.

[18] Smarty Ears. 2011. *Speech Trainer 3D.* [online] Available at: http://smartyearsapps.com [Accessed November 2011

[19] Zharkova, N. 2013. Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal.* Vol 50. pp.76-81.