A Speech Therapy Animation and imaging Resource (STAR)

Eleanor Lawson†, Joanne Cleland†, Jane Stuart-Smith‡, Brian Aitken‡, Janet Beck •

† University of Strathclyde,
‡ University of Glasgow,
• Queen Margaret University, Edinburgh

ABSTRACT

The Speech Therapy Animation and imaging Resource (STAR) is a Web-based anglophone Speech and Language Therapy resource, that hosts over 1,000 videos showing the visible and hidden movements of speech articulators, using (i) ultrasound tongue imaging (UTI) with lip camera video, (ii) magnetic resonance imaging (MRI), and MRI-based animation. The *STAR* resource focuses on children's speech therapy and comprises of (1) a teaching and training website, which will host UTI videos of disordered and nondisordered child speech, nondisordered adult speech, and MRI videos of modelled speech (2) an in-clinic website hosting vocal-tract animations to aid speech therapy and speech practise at home.

STAR will be the first resource of its kind, providing Speech and Language Therapy students and teachers with real examples of imaged and animated vocal tract movement in disordered and nondisordered speech.

Keywords: Speech and Language Therapy, Ultrasound-Tongue-Imaging.

1. VISUAL APPROACHES TO SPEECH THERAPY

SLT training and practice has been dominated by auditory and descriptive approaches. However, there is increasing evidence that vocal-tract visualisation can provide breakthroughs in speech error remediation, see §§1.1-1.2. Recent research into visually-aided speech therapy has focussed on two main remediative tools that reveal hidden vocal organ movement (i) Visual Biofeedback (VB) (ii) Visual animated articulatory models (VAAMs).

1.1 Visual Biofeedback

VB provides the speech therapist and client with real-time visual information about the position and movement of the client's tongue during speech production, through e.g. UTI, using standard medical ultrasound machines**.** UTI VB therapy helps with both diagnosis and remediation of speech-sound errors. Research involving VB therapy shows breakthroughs for clients with persisting speech sound disorders - errors that show no improvement, despite years of conventional speech therapy, see [1], [2], [3]; including those caused by congenital sensorineural hearing impairment [4], and childhood apraxia of speech (CAS) [5]. However, it has been noted that viewing visual UTI video examples of target articulations plays a significant role in speech remediation, even without the feedback element of the client seeing their own tongue movement [3]. This finding may be due to the fact that vocal-tract visualisation is a more intuitive method of understanding correct articulations, avoiding complex descriptions that clients, especially children, may struggle to understand.

1.2 Visual animated articulatory models (VAAMs)

VAAMs show midsagittal cut-aways of the head and vocal tract, or partially-transparent 3-D heads, revealing the movements of hidden articulators. They provide the speech and language therapist (SLT) and their client with a visual representation of a speech target. VAAMs have proved to be effective in therapy for children with hearing loss [6], and for remediation of persisting speech-sound disorder [7]. While VB requires investment in technology and training, VAAM-based therapy is comparatively cheap. Roxburgh [9] compared the effects of UTI VB therapy versus VAAM-based therapy (the mobile/i-pad app "Speech Trainer 3D" [8], for children with submucous cleft palate. Results showed no advantage of VB over the VAAM [9].

Research using UTI VB tools show huge potential in Speech Therapy, but UTI is used extremely rarely by SLTs, especially in the U.K. Studies involving VAAMs in clinical settings also show that the provision of accurate visual models of the hidden speech articulators can have a breakthrough role in remediation of speech-sound disorders, but quality and accuracy of commercial VAAMs is a potential issue, i.e. it is not always clear how animated articulatory movements have been generated and whether they are based on real vocal-tract movement. The addition of normative

articulatory visual models (imaged and animated) to the SLTs arsenal, has the potential to make speech therapy easier, reduce caseloads, and, crucially, to improve client outcomes. Additionally, animated and imaged speech materials of nondisordered and disordered speech has the potential to greatly improve SLT training; providing teachers and students with accurate information about how the vocal organs move inside the vocal tract in both nondisordered and disordered speech.

2. FROM SEEING SPEECH TO STAR

2.1 Clinical use of Seeing Speech

Seeing Speech www.seeingspeech.ac.uk [10] was developed as a Phonetics teaching and learning resource, hosting UTI and MRI videos of speech, and animation (based on MRI vocal-tract recordings frame-by-frame). However, *Seeing Speech* is not specifically designed for SLT needs; it does not have an optimal interface for use in clinic and is missing many important resources for SLTs. A user survey conducted between 2019 – 2020 revealed that 18% (N=140) of *Seeing Speech* user-respondents were either SLTs, SLT teachers, researchers, or students, using the site for: continuing professional development (CPD); revision, in preparation for sessions with patients; in clinic and to recommend to clients and their carers. Specific feedback from SLTs on the usefulness of *Seeing Speech* stated that viewing speech articulations helped with diagnosis; feedback to clients/carers; provided a visual support for children to better explain tongue placement, and improved understanding of non-English speech sounds. An informal survey of local Scottish SLTs, asked how *Seeing Speech* had changed practice. SLTs reported greater understanding of articulatory movement for themselves, and improved description and feedback to clients and their carers. SLTs reported that priorities for a clinical resource, similar to *Seeing Speech*, would be: a more child/carer-friendly interface; addition of missing items such as affricates; animated examples of speech sounds for clients and their carers to view; materials to help clients practise at home; addition of the *extensions to the International Phonetic Alphabet for disordered speech* (extIPA) chart [11]; imaged examples of different disorders – e.g. cleft palate, speech sound disorders, cerebral palsy etc.; imaged examples of speech from different age groups; electropalatographic images; different ways of producing some sounds e.g. /s/ made with the tongue tip down; more animations showing the difference between /ʃ/ and /s/; more speech imaging examples in the coronal plane, especially for sibilants and /r/.

2.2 The STAR website

Funding was obtained from the *Economic and Social Research Council*'s *Secondary Data Analysis Initiative* to create a vocal-tract imaging resource, after the model of *Seeing Speech*, but with a children's Speech and Language Therapy focus: the *Speech Therapy Animation and Imaging Resource* (STAR). *STAR* will be launched in November 2023 and will comprise of two websites:

(1) www.seeingspeech.ac.uk/speechstar/ - a site for Speech and Language Therapy teaching and training that will provide users with access to imaged and animated video examples of midsagittal vocal tract movements in disordered and nondisordered speech.

(2) www.speechstar.ac.uk - an in-clinic site for child Speech and Language Therapy use that will provide users with quick and easy access to midsagittal vocal-tract animations of English speech samples to: help explain vocal-organ movement to clients and their carers; identify speech-production targets, and to aid practise at home.

An additional aim of the *STAR* project is to make curated UTI recordings, collected during research projects for over fifteen years, available to key user groups who might benefit from having access to them (see table 1).

Table 1: Projects from whose corpora UTI/MRI video were selected for the STAR databases. Dark grey cells indicate clinical-focussed child-speech projects.

Project title, date, imaging modality	Speaker age	No. of speakers
Dynamic Dialects (2014-15) (UTI+lip video/MRI)	Adult (18-65)	54
Changes in shape, space and time (2016-19) (UTI+lip video)	Adult (18-50)	32
Ultrax (2011-14) (UTI+lip video)	Child (5-13)	68
Ultraphonix (2015-16) (UTI)	Child (6-13)	20
Visualising Speech (2017-18) (UTI)	Child (3-15)	39
Ultrax2020 (2017-20) (UTI)	Child (5-16)	50

2.2.1 STAR content

The *STAR* teaching and training site will contain:

i. Clinically-focussed introductory materials explaining how the resource can be used; describing and showing how UTI, the MRI recordings were collected, and how the speech animations were created.
ii. An interactive clickable extIPA chart showing midsagittal vocal tract movements in modelled disordered speech sounds, imaged using UTI, MRI and in animated form.
iv. The Edinburgh MRI modelled speech corpus – a new database of modelled speech recorded with MRI (see §2.2.2 below).
iii. A set of filterable UTI (and lip camera) speech databases: (1) a non-disordered speech database of key words, (2) a disordered child-speech sentences database, and (3) a child speech-error database (containing words and phrases).

STAR databases are underpinned by the MySQL database management system [12]. The UTI databases are filterable by some or all of the following filters: *speaker code*; *age*; *sex*; *speech-disorder diagnosis*; *speech-error type*; *speech sound*; *place of articulation*; *syllable position*. Users will be able to select up to four videos to play side-by-side for comparison. All videos are accompanied by metadata: demographic and stimulus/target. The child speech-error database also contains a phonemic transcription of the target word and phonetic transcription of the utterance produced.

The *STAR* in-clinic site contains a reduced set of resources, more suited to use in-clinic and with child clients. This site will contain speech animations, which offer a more easily interpretable stimulus for children and their carers:

i. Animations showing the speech organs and how they move.
ii. Animations of a set of key English consonant sounds in syllables.
iii. Speech animations to support practise at home.

2.2.2 The Edinburgh MRI modelled speech corpus

Funding was obtained from the *Royal Society of Edinburgh* (RSE) and also financial support was given by *NHS Research Scotland* (NRS), through the *Edinburgh Clinical Research Facility*, to create a new MRI-based speech corpus for *STAR*. The Edinburgh MRI modelled speech corpus (EMMSC) consists of dynamic midsagittal vocal-tract MRI recordings of English speech produced by model talker Prof. Janet Beck. Materials consist of English minimal sets (recordings of consonants that are common English therapy targets: [s], [ʃ], [tʃ], [t] and [k] in onset and coda position with a variety of high, low, front and back vowel nuclei), and polysyllabic words containing these key consonant sounds.

MRI recordings were made at the *Edinburgh Imaging Facility*, *Royal Infirmary Edinburgh* (EIF RIE), using a 3 Tesla Siemens Prisma MRI scanner (Figure 1), with a head coil receiver.

Figure 1: Siemens Prisma MRI scanner at EIF RIE.

The recording protocol used for the EMMSC corpus maintains a similar number of midsagittal scans per second (6.25) to the MRI scans currently available on *Seeing Speech* (6.5), but increases image resolution by 70% to 320x320 pixels (Figure 2).

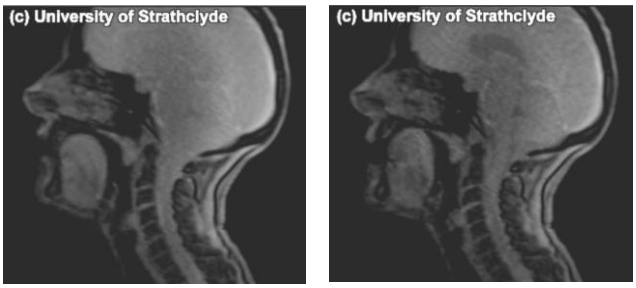

Figure 1: (left) an EMMSC midsagittal scan of a [t] constriction in "Betty", (right) a [k] constriction in "Becky".

The DICOM files obtained were turned into video using ImageJ software [13], then further manipulated (increased brightness and contrast, noise-cancelled audio and copyright marks added) using VirtualDub [14] with frameserver editor AviSynth [15].

Speech audio was recorded inside the MRI scanner using an OptoAcoustics FOMRI III dual-channel, fibre-optic microphone system. The microphone was fixed to the MRI head coil using

velcro straps and the microphone was positioned as close as possible to the speaker's lips using the microphone's goose neck. Further noise-cancelling was carried out using Audacity [16]; however, it was not possible to completely remove extraneous noise relating to the MRI machine without removing too much of the speech signal. For the purposes of exemplifing speech sounds to SLT students, teachers and child Speech and Language Therapy clients, we plan to add MRI videos with "clean" audio at a later date, i.e. having our model talker match her productions in the MRI machine, while being recorded in a sound studio. We will use the denoised MRI audio recordings as stimuli for our model talker to copy, while recording her simultaneously with UTI. We then examined the lingual articulation in the MRI and the UTI recording to make sure that they were comparable.

3. EVALUATION

STAR has been co-designed from the outset with input from a Clinical Advisory panel of SLTs, and SLT trainers. Users will also be invited to evaluate the resources in a beta version of *STAR*, followed by a period of revision and improvement before the main launch with an online survey attached to the sites for ongoing feedback.

4. SUMMARY

Speech and Language Therapy research over the past decade has highlighted that adding vocal-tract imaging, or animation of the movements of the visible and hidden speech articulators, to the SLTs arsenal can be game changing for remediation of persistent speech sound disorders. Building on the success of *Seeing Speech*, *STAR* will be a web-based resource providing SLTs, SLT teachers, students, clients and carers access to curated databases of imaged and animated vocal-tract movement to improve understanding of how the vocal organs move in disordered and nondisordered speech.

7. REFERENCES

[1] Adler-Bock, M., Bernhardt, B., Gick, B. and Bacsfalvi, P., 2007. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *AJSLP*, 16(2), 128-139.

[2] Byun, T., Hitchcock, E. and Swartz, M., 2014. Retroflex Versus Bunched in Treatment for Rhotic Misarticulation: Evidence From Ultrasound Biofeedback Intervention. *JSLHR*, 57(6), 2116-2130.

[3] Cleland, J., Scobbie, J.M. and Wrench, A., 2015. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *CLP*, 29(8-10), 575-597.

[4] Bernhardt, B., Gick, B., Bacsfalvi, P. and Ashdown, J., 2003. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *CLP*, 17(3), 199-216.

[5] Preston, J., Brick, N. and Landi, N., 2013. Ultrasound Biofeedback Treatment for Persisting Childhood Apraxia of Speech. *AJSLP*, 22(4), 627-643.

[6] Massaro, D.W. and Light, J., 2004. Using visible speech to train perception and production of speech for individuals with hearing loss. *JSLHR*, 47(2), 304-320.

[7] Fagel, S and Madany, K., 2008. A 3D virtual head as a tool for speech therapy for children. Proceedings of the ninth annual conference of the international speech communication association (Interspeech 2008) 22-26 Sept. 2008, Brisbane, Australia, 2643-2646.

[8]Smartyears, 2017. Speech Trainer 3D. https://www.smartyearsapps.com

[9] Roxburgh, Z., 2018. Visualising articulation: real-time ultrasound visual biofeedback and visual articulatory models and their use in treating speech sound disorders associated with submucous cleft palate. (unpublished PhD thesis). Edinburgh: Queen Margaret University.

[10] Nakai, S., Beavan, D., Lawson, E., Leplatre, G., Scobbie, J.M. and Stuart-Smith, J., 2018. Viewing speech in action: speech articulation videos in the public domain that demonstrate the sounds of the International Phonetic Alphabet (IPA). *ILLT*, 12(3), 212-220.

[11] ICPLA (2008) Extensions to the International Phonetic Alphabet for disordered speech. https://www.internationalphoneticassociation.org/sites/default/files/extIPAChart2008.pdf

[12]Widenius, M. 2009. MariaDB v10.6.5 https://mariadb.org/

[13]Rasband, W.S. 1997. ImageJ v1.50i https://imagej.nih.gov/ij/

[14]Lee, A. (1998) VirtualDub v1.10.4. https://www.virtualdub.org/

[15] Rudiak-Gould, B., van Eggelen, E., Post, K., Berg, R., Brabham, I. (2000). AviSynth, v2.5. http://avisynth.nl

[16]Audacity Project. (2005). Audacity. v2.0.5 https://www.audacityteam.org/