

Introduction

Ultrasound has been used as a diagnostic medical tool since 1956. In 1969 Kelsey, Minifie and Hixon proposed using ultrasound to image the tongue (Ultrasound Tongue Imaging: UTI) as a safer alternative to x-ray imaging for viewing the articulators in both typical and disordered speech production. UTI is an instrumental articulatory technique that can be used to view and record movements of the surface of the tongue from the root to near the tongue tip. This chapter will describe ultrasound equipment and analysis followed by an overview of the type of speech errors which can be viewed with UTI.

Ultrasound refers to soundwaves beyond the range of human hearing. In diagnostic medical ultrasound, a probe (also known as transducer) consisting of a piezoelectric crystal emits pulses of ultrasonic soundwaves between 2-5MHz through the tissues of the body. Soundwaves have reflective properties, therefore after each pulse is emitted there is a short delay before an echo is received from reflecting surfaces. This echo is converted to a strength and distance and plotted to give an image, allowing us to view the borders of tissues of different density.

There are two main applications of UTI in clinical phonetics. The first is to use the real-time dynamic images as a biofeedback tool (see Cleland and Preston, in press, for an overview) for treating children and adults with speech sound disorders. In this application, the speaker sees their own tongue moving in real time and, guided by a Speech-Language Pathologist, uses that information to modify incorrect articulations. To use ultrasound in this way the clinician need only have a basic B-mode medical ultrasound system and a suitable probe. This is in contrast with the second application of UTI, which is for phonetic research or detailed assessment of speech disorders. In order to study the complex dynamics of speech, methods of synchronising ultrasound with the acoustic signal at a fast enough frame rate to capture speech events are required. There are now many studies using ultrasound as a tool for studying typical speech, with common applications including sociophonetic research into tongue-shapes for /r/ and vowel changes (see for example, Lawson, Scobbie & Stuart-Smith, 2011). Advances in ultrasound hardware coupled with advances in software for analysing ultrasound have led to ultrasound becoming standard equipment in the phonetics lab. However, clinical phonetic applications are only now emerging, perhaps because researchers have focused mainly on the biofeedback applications in clinical populations.

Ultrasound Equipment

Ultrasound Tongue Imaging

Ultrasound equipment consists of a probe (transducer) and central processing unit connected to either a stand-alone unit of the type seen in hospitals, or, more commonly in the phonetics laboratory, a PC or laptop. A variety of probes are available from ultrasound manufacturers. However, to be useful for measuring tongue movement the probe must both fit under the speaker's chin and image at adequate depth (the distance the ultrasonic waves can penetrate through the body, normally around 70-90mm is required) with an adequate field of view to image as much of the tongue as possible (Lee, Wrench, & Sancibrian, 2015). The most commonly used probes are microconvex and 20mm convex probes. The microconvex is smaller and fits well under small children's chins, allowing more of the tongue tip to be viewed, however imaging at depth can be poor. A convex probe gives a brighter image at depth, making it more suitable for larger children and adults. Figure one shows the field of view images by both types of probes.

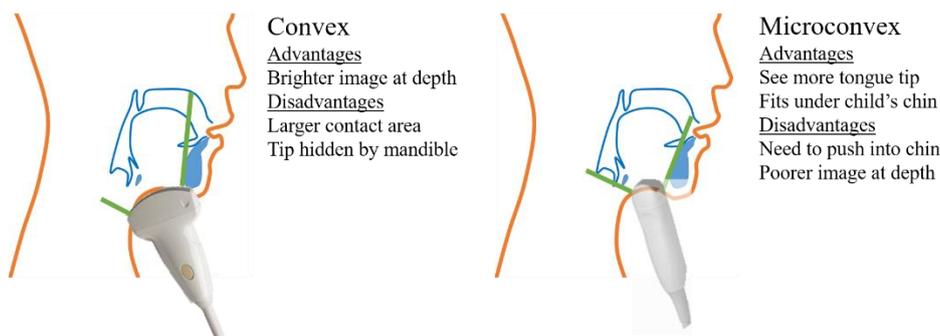


Figure 1: Comparison of convex (left) and microconvex probes (right). With thanks to Alan Wrench for supplying the image.

Frame rate is also important. It is possible to achieve framerates $>300\text{Hz}$, however, imaging at this frame rate comes at the cost of spatial resolution and field of view. Wrench and Scobbie (2011) discuss the framerates required for imaging high velocity articulations such as clicks and conclude that the optimum framerate required is 60 to 200Hz. In clinical phonetic applications the lower framerate is adequate unless users specifically want to study clicks or trills. It should be noted that some ultrasound systems used in biofeedback research have framerates much lower than 60Hz and are therefore unlikely to be useful for studying speech dynamics from recorded data, although they are very useful for viewing tongue movements live. More recent systems used in biofeedback studies, for example the Micrus/Sonospeech system (see www.articulateinstruments.com), supply frame rates of around 100Hz and are therefore suitable for both biofeedback and clinical phonetic applications.

Ultrasound Tongue Imaging

There are two key technical considerations when using UTI for clinical phonetic applications. The first is synchronising the ultrasound to the acoustic signal and recording both for playback and analysis. The second is determining, or deciding not to determine, a co-ordinate space for the resultant tongue shapes. A co-ordinate space can only be determined by either fixing the probe relative to the head using a headset or correcting for probe movement using motion tracking. In contrast, hand-holding the probe is convenient for the phonetician and likely to be more comfortable for speakers. However, the data will consist of a series of tongue curves which are not located in the same co-ordinate space because it is impossible to hold the probe completely still by hand. Even consecutive frames may have small differences in translation and rotation of the tongue shape. It is also quite likely that the probe will drift to an off-centre position, leading to a view which is not mid-sagittal and prone to artefacts. Some researchers use a small camera or mirror as a method of checking that the probe is roughly in the midline and it is possible to use data acquired in this way either for qualitative observations or by applying scalar metrics which are not sensitive to translation and rotation. These types of metrics give numeric values based on single tongue curves and are sensitive to tongue-shape features such as the amount of tongue bunching and/or the number of tongue inflections (see below). However, such metrics still presume that the probe is held in the midline and may not be sensitive to twisting (yaw and pitch) or lateral displacement. An alternative approach is to stabilise the probe relative to the head or correct for head movement. Several research labs have systems for doing so, though not all have been used for clinical research. The Haskins Optically Corrected Ultrasound System (HOCUS, Whalen et al., 2005) uses a spring loaded holder for the ultrasound probe and optical tracking to correct for speaker head movement. A similar system, Sonographic and Optical Linguo-labial Articulation Recording system (SOLLAR, Noiray et al., 2018) adapts this technology into a child-friendly spaceship-themed set up. Both systems require attaching tracking dots to the speaker's head and probe and require two experimenters for data collection. Neither of these systems have yet been used for clinical phonetic research, though it would be possible to do so. Other labs stabilise the probe using a headrest or headset and have developed in-house software for analysing data which is often made freely available (see for example www.ultraspeech.com or UltraCATS, Gu et al., 2004). A complete ultrasound system for phonetic research and biofeedback is available from Articulate Instruments Ltd. (www.articulateinstruments.com). This small portable system is a USB device which connects to a PC or laptop. Software, Articulate Assistant AdvancedTM, can record synchronised ultrasound, audio, and lip camera data (and other channels such as electropalatography if desired). This system has been used in several clinical phonetic studies

Ultrasound Tongue Imaging

and intervention studies. In contrast to the optically corrected systems, the Articulate Instruments system makes use of a probe stabilising headset. This allows the user to ensure that measurements of tongue shape are not affected by probe movement but has the disadvantage that it can become heavy over time and may restrict jaw movement. Figure two shows headsets used in clinical studies in Scotland.



Figure 2: Headsets from Articulate Instruments Ltd. used in clinical studies arranged from left to right chronologically.

Understanding the Ultrasound Image in Typical Speakers

The surface of the tongue can be imaged in either a mid-sagittal or coronal view. Ultrasound is absorbed by bone and does not travel through air, therefore in both these types of images the mandible and/or hyoid bone casts a dark shadow and the air boundary at the surface of the tongue is a bright white line. Because ultrasound is reflected soundwaves, users should be aware that the ultrasound image is not a photograph, rather it is vulnerable to artefacts which can be difficult to interpret, especially if the probe is off-centre or indeed if the speaker has very unusual anatomy. The hard palate and the pharyngeal wall are not visible in the ultrasound image. It is possible to obtain a tracing of the hard palate by asking the speaker to swallow some water. This tracing can then be super-imposed on the image if the speaker is wearing a probe-stabilising headset or if a system for correcting for head movement is used. If the probe is held by hand then it is not possible to add a hard palate reference trace due to probe movement. Figure 3 shows an ultrasound image taken from the maximum point of articulation of [k]. Most UTI studies use the mid-sagittal view with the tongue tip to the right. Unlike electromagnetic articulography and electropalatography (EPG), the image shown is an anatomically correct representation of the tongue from near the tongue tip to near the root. It can be used to image all lingual articulations, although the shadow from the mandible can obscure the tongue tip, making dentals, interdental, and linguolabials difficult to see in their entirety. Ultrasound holds advantages over EPG for imaging vowels and post-velar articulations since EPG shows only tongue-palate contact from the alveolar ridge to the

Ultrasound Tongue Imaging

boundary of the hard and soft palate. However, ultrasound typically only images a single mid-sagittal (or coronal) slice, therefore features such as tongue bracing is better imaged with EPG. 3D ultrasound systems are available, but they are as yet unable to image with the temporal resolution 2D systems are capable of.

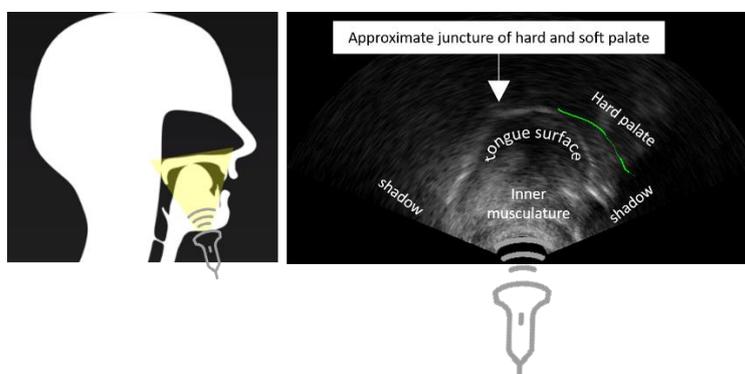


Figure 3: Mid-sagittal ultrasound image taken from the maximum point of contact of [k]. Tongue tip is to the right.

An understanding of ultrasound tongue shapes in typical speakers is required in order to interpret ultrasound images from clinical populations. This is especially true when making qualitative judgements about potentially disordered tongue shapes in speakers with motor speech disorders or unusual anatomy (for example, speakers who have had partial glossectomy). The ultrasound image is not normalised, therefore the display for each speaker shows a slightly different size and shape of tongue. Because of this, there are a large number of speaker-dependent factors involved in interpreting ultrasound. Moreover, the lack of normalisation highlights the wide-range of tongue shapes that are possible. For example, American English /r/ may have at least six different possible tongue shapes (Boyce et al., 2015). Tongue-shapes for vowels will also differ greatly, even within one language, due to both dialect features and speaker-specific anatomy. This makes comparing ultrasound images between speakers a challenge, with most phonetic studies choosing to look at intra-speaker measures. Nevertheless, there are some commonalities in tongue-shapes for specific speech sounds. The www.seeingspeech.com website provides a very useful clickable IPA chart of ultrasound movies collected from two different speakers. Table 1 shows example ultrasound images from a range of consonants compared to MRI images from the same speaker.

Articulation	Ultrasound	MRI
Alveolar Stop [t]		

Ultrasound Tongue Imaging

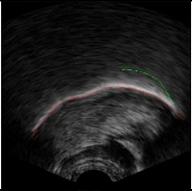
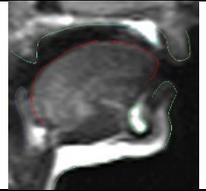
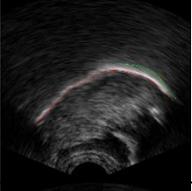
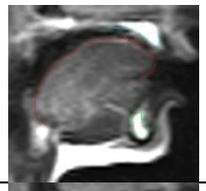
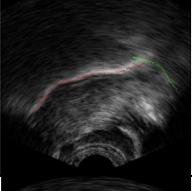
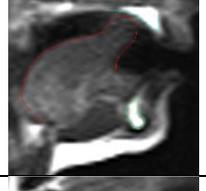
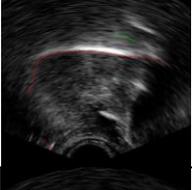
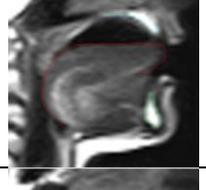
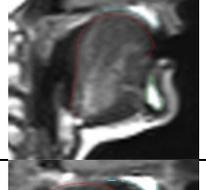
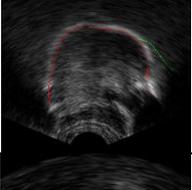
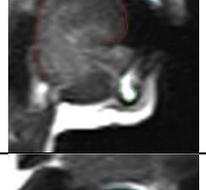
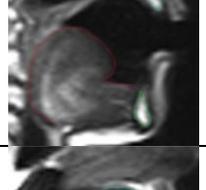
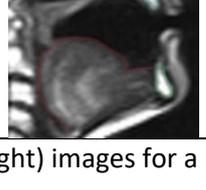
Alveolar fricative [s]		
Post-alveolar fricative [ʃ]		
Retroflex approximant [ɻ]		
Dark l [ɫ]		
Palatal fricative [ç]		
Velar Stop [k]		
Uvular stop [q]		
Pharyngeal fricative [ħ]		

Table 1: Ultrasound (left) and MRI (right) images for a range of consonants from the IPA

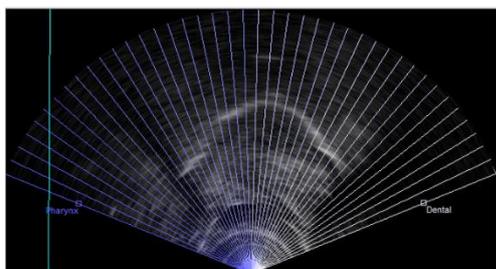
Analysing Ultrasound Data

The type of analyses that are possible depend on the type of ultrasound system used to acquire the data. Most types of analysis involve first tracking the surface of the tongue and fitting a

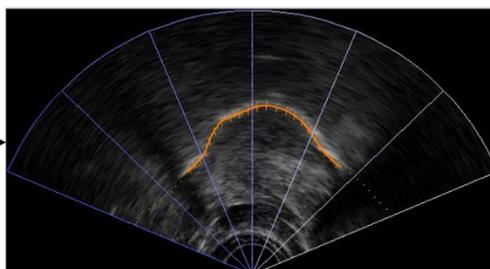
Ultrasound Tongue Imaging

spline with a number of discrete data points to the surface. Until fairly recently phoneticians fitted splines to ultrasound images by hand using specialist software (for example, Articulate Assistant Advanced (AAA), or UltraCats). This was a time consuming process and as such necessitated choosing only a limited number of frames for analysis, for example choosing only the burst of stops or midpoint of fricatives. There are now a number of algorithms available for fitting splines to the tongue surface automatically (including in AAA software), or semi-automatically, allowing all ultrasound frames to be used if dynamic analysis is required. Once splines are fitted, either polar or Cartesian coordinates can be exported and used to compare tongue shapes quantitatively. Polar coordinates are often used since the probe can be used as the origin and splines can be fitted along radial fanlines. Figure 4 gives an example of data acquired using the Articulate Instruments Ltd ultrasound system with headset stabilisation. The top left panel shows 62 fanlines emanating from the probe. First, the phonetician annotates the speech data using either acoustic or articulatory landmarks. In this example, the phonetician has annotated stop closures at the burst. Next, a spline is automatically fitted to the surface of the tongue (panel 2). From here either the co-ordinates can be exported and compared statistically or multiple splines can be compared visually (panel 3). If headset stabilisation is used, it is possible to average multiple splines for comparison (panel 4), see Articulate Assistant Advanced “t-test function”. If headset stabilisation is not used (either because it is not available or because it is not desirable for young children or speakers with clinical diagnoses) then co-ordinates are exported and scalar metrics are used to compare tongue-shapes. For example, the Dorsum Excursion Index (DEI, Zharkova, 2013) can be used to compare the degree of tongue dorsum movement between two different target phonemes, irrespective of translation and rotation.

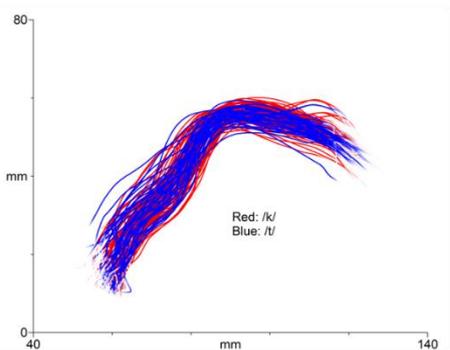
Ultrasound Tongue Imaging



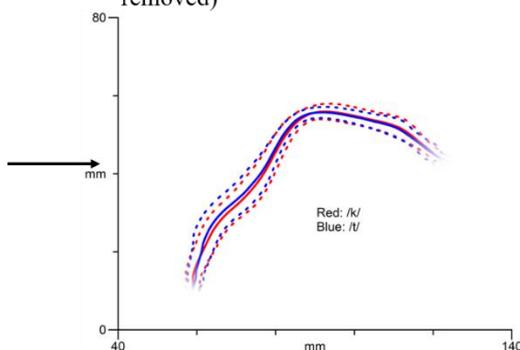
Raw ultrasound image with radial fanlines (62) superimposed



A spline is fitted to the surface of the tongue. 62 measurement points are available. In this diagram redundant datapoints (those in shadow) are removed)



Multiple splines can be exported and compared. In this plot multiple productions of /k/ and /t/ are shown.



Splines can be averaged. This plot shows mean splines for /k/ and /t/ and the standard deviations (dotted lines)

Figure 4: Example ultrasound analysis. Panel 1 (top left) shows the raw ultrasound image with fanlines emanating from the probe. Panel 2 (top right) shows a spline fitted to the surface of the tongue with measurement (data) points represented by orange spokes

Table 2 summarises the key indices which have been used in the literature to quantify or compare tongue shapes with reference to papers using them on clinical populations. It is important to note that even if probe stabilisation is not used it is often still important that the UTIs are acquired with both the mandible and hyoid shadow in view. This is because most measures require the phonetician to determine end points of the tongue spline and this is not possible if the probe is rotated such that a large portion of the tongue front or root is not in view.

Ultrasound Tongue Imaging

Measure	Description	Independent of head movement	Example use
T-test function (Articulate Instruments Ltd)	Determines whether there is a statistically significant difference between two sets of tongue curves along a fanline.	No	Identification of covert contrast in children with persistent velar fronting (Cleland et al., 2017)
Mean Nearest Neighbour Difference (Zharkova & Hewlett, 2009)	Quantifies the difference between two tongue curves	No	Coarticulation and stability in people who stutter (Belmont, 2015)
Dorsum Excursion Index DEI (Zharkova, 2013)	quantifies the extent of excursion of the tongue dorsum, conceptually represented by the point on the tongue curve located opposite the middle of the straight line between two curve ends	Yes	Identification of covert contrast in velar fronting (McAllister Byun et al., 2016)
LOC _{a-i} (Zharkova, 2015)	Quantifies excursion of the tongue in relation to the back of the tongue.	Yes	Covert contrast of s and f (both stopped to dental) (Zharkova et al., 2017)
Curvature Degree Curvature position (Aubin & Menard, 2006)	Quantifies the extent of maximal excursion within a single tongue shape Determines where on the tongue curve the excursion is	Yes	Covert contrast of s and f (Zharkova et al., 2017)
Procrustes analysis (Dawson et al., 2016)	Gives the sum of the squared differences between two tongue shapes, following translation, rotation, and scaling	Yes	Tongue complexity in vowel production in speakers with Down Syndrome compared to resting state (Carl, 2018)
Modified Curvature Index (Dawson et al., 2016)	Quantifies curvature degree and number of inflections in a tongue shape	Yes	Tongue complexity in vowel production in speakers with Down Syndrome (Carl, 2018)

Ultrasound Tongue Imaging

Discrete fourier transformation (Dawson et al., 2016)	Differentiation of tongue shapes between targets	Yes	Differentiation of tongue shapes for target vowels in speakers with Down syndrome (Carl, 2018)
NINFL Number of Inflections (Preston et al., 2019)	Count of the number of tongue curvature changes in a single tongue contour	Yes	Tongue complexity in rhotic distortions (Preston et al., 2019)
Peak Velocity (Heyde et al., 2015)	Calculates displacement and velocity of tongue movement along a single measurement vector	No	Kinematic analysis in people who stutter (Heyde et al., 2015)
KT Crescent (Scobbie & Cleland, 2017)	Quantifies dorsal velar constriction spatially in relation to a same-speaker alveolar baseline	Yes	Degree of separation between /k/ and /t/ in a child with persistent velar fronting (Cleland & Scobbie, 2018)
Width Measure and Tongue Length (Roxburgh, 2018)	Quantifies the mean and maximum distances between two curves measured radially from the probe origin. Zones of significance are compared to the proportion of visible tongue length.	Yes	Degree of separation between velar and alveolar consonants in children with submucous cleft palate (Roxburgh, 2018)
Anteriority index 3D Ultrasound (Bressmann et al. 2005)	Volumetric index indicating relative position of the main mass of intrinsic tongue tissue in the oral cavity	No	Tongue protrusion following partial glossectomy (Bressmann et al. 2005)
Concavity index 3D Ultrasound (Bressmann et al. 2005)	Measures how convex or concave the shape of the tongue is along the whole length of the tongue volume	No	Tongue grooving following partial glossectomy (Bressmann et al. 2005)
Asymmetry index. 3D Ultrasound	Volumetric measure of the difference in lateral tongue height between each side of the tongue	No	Tongue symmetry following partial glossectomy (Bressmann et al. 2005)

Ultrasound Tongue Imaging

(Bressmann et al. 2005)			
Tongue Displacement	Cumulative displacement of maximum tongue height measured radially	No	Tongue displacement during production of /r/ before and after intervention
(Bressmann et al., 20116)			(Bressmann et al., 2016)

Table 2: Ultrasound Metrics used to quantify disordered speech

Clinical Populations

As a biofeedback tool UTI has been used most often with children with residual speech sound errors; persistent speech sound disorders; Childhood Apraxia of Speech; hearing impairment; cleft lip and palate; and Down syndrome (Sugden et al., 2019). It is less used as a biofeedback tool for adults with acquired disorders, though it has been used with adults with acquired apraxia of speech (Preston & Leaman, 2014) and glossectomy (Blyth et al., 2016; Girod-Roux et al., 2020). As a clinical phonetic tool UTI has potential utility in any studies where the phonetician believes that studying tongue movement directly is likely to add important information that phonetic transcription cannot provide. Therefore, it is mostly likely to be of use for investigating motor speech disorders; articulation disorders including cleft lip and palate; and dysfluency. For children with surface-level phonological impairments it can be important in uncovering subtle underlying motor speech constraints. Ultrasound is also used to identify subtle speech differences in speakers with visual impairments (Ménard et al., 2013), however, since visually impaired speakers do not present with a speech disorder this population is beyond the scope of this chapter.

Speech Errors Revealed by Ultrasound

Clinical phonetic applications of UTI broadly fall into four different types: 1. Comparisons with typical speakers. 2. identification of covert errors/supplementing phonetic transcription 3. identification of covert contrasts; and 4. quantifying change post-intervention.

Comparisons with typical speakers.

Studies comparing the speech of clinical populations to typical speakers using UTI are relatively sparse to date. Ultrasound offers the potential to answer theoretical questions about the nature of tongue movements in disordered speakers, yet normative data with which pathological data could be compared is relatively rare, especially in studies of children. UTI studies of typical speakers tend to answer specific questions, for example studies of coarticulation, rather than provide information about a wide variety of tongue movements. Nevertheless a few studies have attempted to compare ultrasound metrics between typical speakers and speakers from clinical groups. Heyde et al. (2016) compared peak velocity and duration of onset and offset of consonantal closure between persons who stutter and typical

speakers. They demonstrated that persons who stutter display different behaviour in the consonantal offset. In a number of studies comparing typical speakers with partial glossectomy patients Bressmann and colleagues have used 3D ultrasound to measure the functional topography of the tongue prior to surgery with the hope of predicting post-surgical outcomes (see for example Bressmann, Uy & Irish, 2004). Using scalar metrics (see table 2) Carl (2018) demonstrated that speakers with Down syndrome show reduced tongue shape curvature and complexity of some vowels compared to typical speakers.

An alternative approach to comparing clinical populations to typical speakers is to compare speakers with reduced intelligibility to speakers from the same clinical group with normalised speech. Cleland et al. (2019) report on error types identified using UTI in 39 children with cleft lip and palate. Many of the children presented with few or no errors in their speech. Given the anatomical differences in cleft lip and palate, children with normalised speech serve as a better control group than typical speakers. The error types identified in this study are explored in detail in the next section.

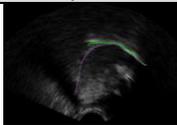
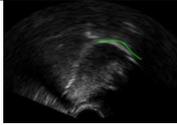
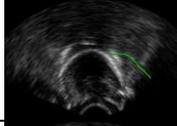
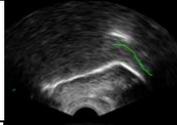
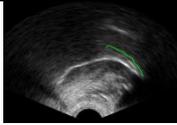
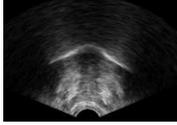
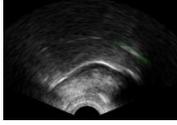
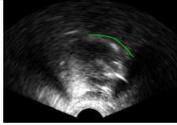
Covert errors/supplementing phonetic transcription

There is a long history of using instrumental phonetic techniques to supplement phonetic transcription and describe covert errors. Cleland et al. (2019) demonstrated that using ultrasound to aid phonetic transcription leads to improved inter-rater reliability and reveals errors such as double articulations that were missed during traditional phonetic transcription. These covert errors are subtle unexpected articulations which suggest underlying motor constraints but do not lead to loss of contrast or pattern phonologically. For example, Cleland et al. (2017) describe a case study of a child called Rachel with persistent velar fronting who produced oral stops using a variety of different tongue shapes (retroflexes, undifferentiated gestures, and typical alveolar gestures) which appeared random, rather than phonologically patterned. These error types are important because they can provide useful diagnostic information, in Rachel's case a potential motor speech disorder. Similarly, Bressmann et al., (2011) describe Sandra, a nine year old with cleft lip and palate, who alternated inaudible alveolar and velar gestures during attempts at /k/ which were transcribed as [ʔ], again this information is important because it shows that Sandra is attempting appropriate lingual movements.

Ultrasound can also be used to confirm perception based transcription in a more objective way, for example, backing errors show as clear raising of the tongue dorsum. Table 3 summarises

Ultrasound Tongue Imaging

both types of ultrasound errors as reported so far in studies of children in the Ultrasuite corpus (Eshky et al., 2019). Errors are classified broadly in line with Gibbon’s 2004 taxonomy of error types identified using electropalatography in the speech of children with cleft lip and palate. Since ultrasound shows tongue shape rather than tongue palate contact, “open pattern”, an EPG pattern where there is no tongue-palate contact, can be sub-divided into retraction to uvular or pharyngeal, or articulatory undershoot. Retroflex productions can also be identified using ultrasound. It is important to note that not all the error types are mutually exclusive, for example, it is possible for one segment to be retracted, variable, and have unusual timing. The error types are described in more detail below.

Error Type	Description	Ultrasound Example	Expected IPA Transcription
1. Increased Contact	Raising of tongue body and tip/blade towards or in contact with the hard palate		Simultaneous alveolar + postalveolar + palatal
2. Retraction to velar or palatal	Anterior target retracted to velar or palatal		Velar or palatal
3. Retraction to uvular*	Any velar or pre-velar target retracted to uvular		Uvular
4. Retraction to pharyngeal*	Any uvular or pre-uvular target retracted to pharyngeal		Pharyngeal
5. Fronted	Posterior target is fronted to palatal, post-alveolar, or alveolar		Alveolar, post-alveolar, or palatal
6. Complete Closure	No visible groove in the coronal view.		Any lateral sibilant
7. Undershoot*	No contact with palate		“lowered” diacritic, or approximant
8. Double Articulation	Simultaneous production of two consonants: normally alveolar-velar		Any double articulation e.g. [k̠t̠] or [p̠t̠]

Ultrasound Tongue Imaging

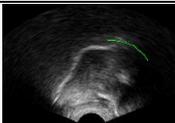
9. Increased Variability	Different tongue-shapes per repetition	<i>(dynamic analysis required)</i>	Different transcriptions across repetitions
10. Abnormal Timing	Mis-directed articulatory gestures or release of articulations with abnormal timing	<i>(dynamic analysis required)</i>	Any diacritic denoting timing such as lengthening marks
11. Retroflexion	Tongue tip retroflexion during any non-retroflex target		Any retroflex consonant

Table 3: Ultrasound Error Types. *An earlier version of these error types appears in Cleland et al., 2019 without retraction to uvular and pharyngeal and where **undershoot is called “open pattern”.

Increased Contact

Undifferentiated lingual gestures are an error type found in children with persistent SSD of unknown origin (Gibbon, 1999) where the child fails to adequately differentiate the tongue tip/blade and tongue dorsum, instead moving the whole tongue in a single undifferentiated gesture. This type of error is indicative of a problem with speech motor control and thus is important from a diagnostic and therapeutic point of view. In UTI undifferentiated gestures show as a raising of the whole tongue body, usually during an alveolar or velar target. If a hard palate has been superimposed on the image then increased or complete contact can be seen from the alveolar to velar region.

Retraction Errors

Retraction errors involve production of an alveolar (or sometimes labial or labiodental) target at a more posterior place of articulation. These errors are often found in the speech of children with cleft lip and palate where compensatory backing articulations are very common (see for example Bressmann et al., 2011). These errors are typically not covert in nature, i.e. the phonetician is able to accurately transcribe an error as palatal, velar, uvular or pharyngeal. However, ultrasound analysis can lead to more accurate transcription and confirmation of place of articulation by viewing the image.

Fronted Placement

Speakers who produce a target velar at an alveolar place of articulation can be seen to have fronted placement on the UTI. These errors are often phonological in nature, for example velar fronting, a common developmental phonological error, where /k/ is produced as [t] (see Cleland et al., 2017). Fronting of /k/ to a midpalatal plosive can also be observed in speakers with cleft

lip and palate (Bressmann et al., 2011). Again, these errors are likely to be overt and therefore the ultrasound offers confirmatory, rather than additional information.

Complete Closure

This error type is viewed in the coronal plane. In this view raising and lowering of the sides of the tongue are visible. During production of a grooved fricative such as /s/ the sides of the tongue are raised and the centre grooved, producing a wing-shaped UTI. In a complete closure error both the sides and centre of the tongue are raised in a dome-shaped UTI. For /s/ and /ʃ/ productions these errors will typically sound like lateral or palatal fricatives. These errors are a common residual (articulatory) speech sound error and UTI confirms phonetic transcription, rather than adds to it. A similar error, loss of concavity, is described by Bressmann et al. (2015) in speakers with lingual hemiparesis. It is not straightforward to determine from coronal ultrasound which place of articulation the complete closure occurs at because, unlike mid-sagittal views, there is no standard method of obtaining a particular coronal slice. The probe can be angled (usually by hand) to point more towards the tip or root to give some indication of the coronal slice being viewed. Alternatively, 3D ultrasound could be used to determine both place of articulation and degree of grooving.

Undershoot

In an undershoot error the tongue approaches the hard palate but fails to make contact. These types of errors are commonly described in the literature on dysarthria using other articulatory techniques, however, despite UTI being applied to patients with dysarthria in one of the first studies of disordered speech (Shawker & Sonies, 1984) it has been under-used with this population. Errors may be transcribed as weak or lowered articulations, with the ultrasound providing important information about the magnitude of the undershoot. This type of error is also seen in children with Childhood Apraxia of Speech or during the therapeutic process where a child moves their tongue in the direction of an articulatory target but fails to make contact.

Double Articulation

Double articulations are normally covert errors or covert contrasts. They involve two simultaneous primary places of articulation such as velar + alveolar or bilabial + velar. In the latter example the phonetician usually transcribes a bilabial because it is visually salient, UTI provides extra information which is diagnostically important. Gibbon (2004) suggests these labio-lingual double articulations may occur as an intermediate step between an incorrect

Ultrasound Tongue Imaging

backing pattern and a correct bilabial production in children with cleft lip and palate, suggesting that presence of this error may indicate positive change in a child's speech system. Bressmann et al., (2011) also report one speaker with cleft lip and palate who used a glottal+pharyngeal double articulation for /k/. Alveo-velar double articulations are harder to identify using ultrasound because they can look very similar to increased contact. Identifying a small pocket of air between the alveolar and velar region is challenging because it relies on very good probe stabilisation and spatial resolution.

Increased Variability

Increased sub-phonemic variability may be indicative of a subtle motor speech disorder. In this error type the speaker produces multiple repetitions of the same speech sound with greater variability than a typical speaker. This is essentially a covert error, because typically the phonetician will transcribe each repetition using the same phonetic symbol but dynamic ultrasound analysis reveals subtle articulatory variability. Figure 5 shows variable productions of /k/ from one child produced during a diadochokinesis task. Single spline tracings from the burst of each production shows variation in the place of the tongue shape, despite all productions being transcribed as [k].

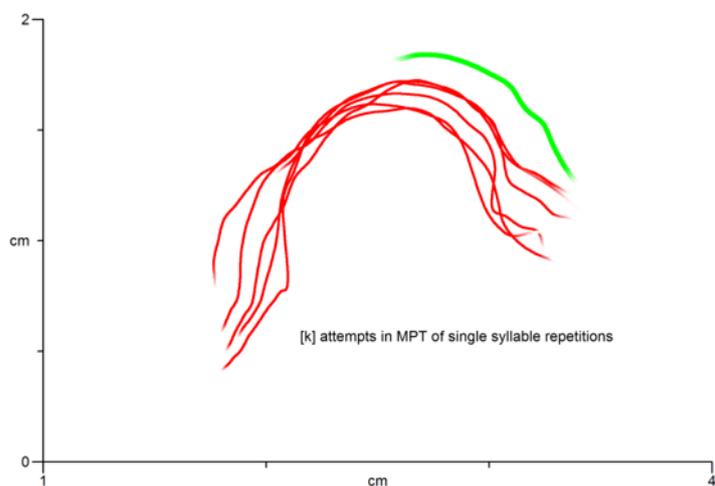


Figure 5: Example variability in tongue shape for six consecutive productions of [k]

Abnormal Timing

Ultrasound can reveal timing errors such as abnormally long closures or misdirected gestures which may indicate difficulties with motor planning or programming. Dynamic analysis is required to identify these errors. They may be transcribed as lengthened stop closures (for

example), however, even visual inspection of the ultrasound can reveal steady tongue-state during silences or long stop closures. These may also be observed in speakers with disfluencies.

Retroflex Productions

Retroflexion is visually salient on ultrasound. In Cleland et al.'s (2017) case study of Rachel retroflex stops for both alveolar and velar targets were identified from the UTI but these productions had not been transcribed as retroflex, suggesting a covert error. It is unclear why children might produce retroflexes in error in a language such as English where retroflexes only occur in the late acquired articulatory complex /r/ target. Another study, Cleland et al., (2015) suggested that one child with persistent velar fronting produced a retroflex stop for /k/ only in a stimulability task where he was instructed to try to produce [k], in this case the explanation for the error could be a conscious effort from the child to produce a more posterior articulation.

Covert Contrast

The term covert contrast is used to describe differences in perceptually neutralised contrasts which can be identified instrumentally. For example, a Speech-Language Pathologist may identify that a child is unable to make a contrast between /t/ and /k/, producing both as [t], however acoustic or articulatory analysis may show small differences between these target phonemes. This is important diagnostically because a child with a covert contrast may have a motor speech problem rather than a phonological one. Two studies have attempted to identify covert contrasts using UTI (Cleland et al., 2017 and McAllister Byun et al., 2016) and a further paper (Zharkova, Gibbon & Lee, 2017) gives suggestions of metrics for identifying covert contrast and a case study of a typically developing child with covert contrast between /s/ and /ʃ/. Identifying covert contrasts in children with speech sound disorders first requires recording the child saying minimal pair words containing the target phonemes (or at least words where the vowel environment is controlled) preferably multiple times. Figure 6 shows average tongue shapes for /s/ and /ʃ/ in a child with cleft lip and palate, both were transcribed as [ʃ]. By comparing the tongue shapes visually we can see that although the tongue shapes overlap in the anterior region, there is an area in the root where the shapes are differentiated. The t-test function in AAA confirms a statistical difference, this child therefore has a covert contrast between these phonemes.

Ultrasound Tongue Imaging

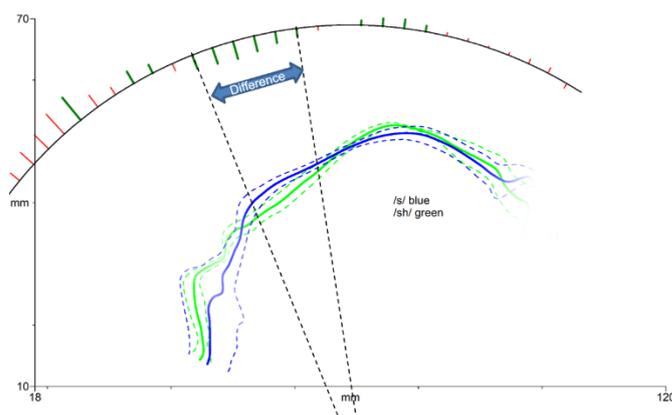


Figure 6: Averaged /s/ (blue) and /ʃ/ (green) productions showing difference in the tongue root despite both being transcribed as [ʃ].

The Cleland et al. (2017) and McAllister Byun et al. (2016) papers both attempted to find covert contrast in children with velar fronting. This particular phonological process is common the speech of typically developing children, but can persist. In both studies children with velar fronting were recorded saying multiple words with velar and alveolar targets and these were then compared to determine if there were any differences in tongue shape between the two target phonemes. The McAllister Byun study also employed acoustic analysis. By comparing tongue shapes for /t/ and /k/ using the in-built t-test function in AAA software, Cleland et al., (2017) showed that the children in their study showed no evidence of covert contrast. McAllister Byun et al., (2016) employed the ratio metric, DEI, in their study of two children with perceptually neutralised velars and found evidence of covert contrast in one child. These studies exemplify the two different approaches to analysing ultrasound data: Cleland et al. (2017) using headset stabilisation to compare sets of tongue curves directly and McAllister Byun et al. (2016) using a scalar metric which can be calculated independent of translation and rotation.

Quantifying Change Post-Intervention

The majority of studies using ultrasound as a biofeedback tool use perceptual measures to quantify change post-intervention (Sugden et al., 2019). As well as being easier to undertake, perceptual measures have ecological validity. A small number of studies have used ultrasound metrics to measure change post-intervention. This has the advantage of being free from listener-bias and has the potential to be automated. Cleland et al., (2015) use the t-test function within AAA to show quantifiable differences in tongue shape after intervention for seven children with a variety of speech sound errors. This approach is suitable for measuring changes in any intervention target which involves changing the shape or position of the tongue, which

is likely to be the case for all ultrasound visual biofeedback therapy. However, it requires collecting assessment data with a probe stabilising headset and aligning sessions using a hard palate trace and therefore may not be suitable for all intervention studies, particularly with younger children.

An alternative approach is to use scalar metrics to measure changes. A number of those shown in table 2 would be appropriate to use with specific targets. For example, DEI would be a good choice for measuring changes post-intervention for children with fronting or backing patterns. The NINFL metric has been used to measure changes in tongue complexity post intervention for children with rhotic distortions (Preston et al., 2019), showing that after intervention children adopt more complex tongue shapes and that this correlates well with perceptual accuracy.

Summary and Conclusions

Ultrasound Tongue Imaging is a safe and useful technique for viewing and quantifying tongue shape and movement. When tongue movements are viewed in real time it is a useful biofeedback tool in the speech pathology clinic, showing promise of being effective in remediating persistent or residual speech sound disorders (Sugden et al., 2019). In the clinical phonetics lab ultrasound is a relatively new tool, having been hampered in the past by low frame rates; difficulties synchronised ultrasound to audio; and difficulties analysing the resultant data. Technological advances have led to several different ultrasound systems and software now being available at affordable prices. This coupled with the ease of acquisition of ultrasound data, even in small children, suggests that this technique is set to become standard equipment for clinical phoneticians.

Ultrasound can be used to identify all of the errors described in the electropalatography literature (Cleland et al., 2019) and advances in metrics for describing ultrasound data illustrate the potential to use this technique to measure tongue movement in a wide variety of clinical populations, either for answering theoretical questions or measuring progress during biofeedback intervention.

References

- Articulate Instruments Ltd 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.
- Aubin, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. In H. C. Yehia, D. Demolin, & R. Laboissière (Eds.), *Proceedings of the 7th international seminar on speech production* (pp. 209–216), Ubatuba, Brazil.
- Belmont, A. (2015). *Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter across the lifespan: An ultrasound study* (Master of Science). University of South Florida.
- Blyth, K. M., McCabe, P., Madill, C., & Ballard, K. J. (2016). Ultrasound visual feedback in articulation therapy following partial glossectomy. *Journal of communication disorders, 61*, 1-15.
- Boyce, S., Tiede, M., Espy-Wilson, C. Y., & Groves-Wright, K. (2015). Diversity of tongue shapes for the American English rhotic liquid. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 847.1-4 retrieved from <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0847.pdf>
- Bressmann, T., Koch, S., Ratner, A., Seigel, J., & Binkofski, F. (2015). An ultrasound investigation of tongue shape in stroke patients with lingual hemiparalysis. *Journal of Stroke and Cerebrovascular Diseases, 24*(4), 834-839.
- Bressmann, T., Harper, S., Zhylich, I., & Kulkarni, G. V. (2016). Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors. *Clinical linguistics & phonetics, 30*(3-5), 345-362.
- 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), 1028-1033.
- Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data

Ultrasound Tongue Imaging

from 12 normal speakers and a partial glossectomee. *Clinical Linguistics & Phonetics*, 19(6-7), 573-588.

Bressmann, T., Uy, C., & Irish, J. C. (2005). Analysing normal and partial glossectomee tongues using ultrasound. *Clinical linguistics & phonetics*, 19(1), 35-52.

Carl, M. (2018). *Vowel Production in Down Syndrome: An Ultrasound Study*. (Unpublished doctoral dissertation). City University New York, New York.

Cleland, J., Lloyd, S., Campbell, L., Crampin, L., Palo, J. P., Sugden, E., Wrench, A. & Zharkova, N. (2019). The impact of real-time articulatory information on phonetic transcription: ultrasound-aided transcription in cleft lip and palate speech. *Folia Phoniatica et Logopaedica*, 1-11.

Cleland, J., & Preston, J. (In Press). Biofeedback Interventions. In A. L. Williams, S. McLeod, R. J. McCauley, (Eds.) *Interventions for Speech Sound Disorders in Children*, 2nd Edition (pp. 1-50). Baltimore: Paul H Brookes Pub Co.

Cleland, J., & Scobbie, J. M. (2018). Acquisition of new speech motor plans via articulatory visual biofeedback. *Fuchs, S., Cleland, J., and Rochet-Cappelan, A.(Eds.). Speech Perception and Production: Learning and Memory*.

Cleland, J., Scobbie, J. M., Heyde, C., Roxburgh, Z., & Wrench, A. A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical linguistics & phonetics*, 31(1), 35-55.

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical linguistics & phonetics*, 29(8-10), 575-597.

Davidson, L. (2005). Smoothing spline analysis of variance (ANOVA) for tongue curve comparison. *The Journal of the Acoustical Society of America*, 118(3), 2023-2024.

Dawson, K. M., Tiede, M. K., & Whalen, D. H. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical linguistics & phonetics*, 30(3-5), 328-344.

Eshky, A., Ribeiro, M. S., Cleland, J., Richmond, K., Roxburgh, Z., Scobbie, J., & Wrench, A. (2019). UltraSuite: a repository of ultrasound and acoustic data from child speech therapy sessions. *arXiv preprint arXiv:1907.00835*.

Ultrasound Tongue Imaging

- Girod-Roux, M., Hueber, T., Fabre, D., Gerber, S., Canault, M., Bedoin, N., ... & Badin, P. (2020). Rehabilitation of speech disorders following glossectomy, based on ultrasound visual illustration and feedback. *Clinical Linguistics & Phonetics*, 1-18.
- Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2), 382-397.
- Gibbon, F. E. (2004). Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. *Clinical linguistics & phonetics*, 18(4-5), 285-311.
- Gu, J., Bressmann, T., Cannons, K., and Wong, W. (2004). *The Ultrasonographic Contour Analyzer for Tongue Surfaces (Ultra-CATS)*, University of Toronto, Toronto, ON, Canada.
- Heyde, C. J., Scobbie, J. M., Lickley, R., & Drake, E. K. (2016). How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. *Clinical linguistics & phonetics*, 30(3-5), 292-312.
- Kelsey, C. A., Minifie, F. D., & Hixon, T. J. (1969). Applications of ultrasound in speech research. *Journal of speech and hearing research*, 12(3), 564-575.
- Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic/r/in Scottish English¹. *Journal of Sociolinguistics*, 15(2), 256-268.
- Lee, S. A. S., Wrench, A., & Sancibrian, S. (2015). How to get started with ultrasound technology for treatment of speech sound disorders. *Perspectives on Speech Science and Orofacial Disorders*, 25(2), 66-80.
- Ménard, L., Toupin, C., Baum, S. R., Drouin, S., Aubin, J., & Tiede, M. (2013). Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, 134(4), 2975-2987.
- McAllister Byun, T., Buchwald, A., & Mizoguchi, A. (2016). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical linguistics & phonetics*, 30(3-5), 249-276.
- Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research*, 61(6), 1355-1368.

Ultrasound Tongue Imaging

Preston, J. L., McCabe, P., Tiede, M., & Whalen, D. H. (2019). Tongue shapes for rhotics in school-age children with and without residual speech errors. *Clinical linguistics & phonetics*, 33(4), 334-348.

Preston, J. L., & Leaman, M. (2014). Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology*, 28(3), 278-295.

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 doctoral dissertation). Queen Margaret University, Edinburgh.

Scobbie, J. M., & Cleland, J. (2017). Area and radius based mid-sagittal measurements of comparative velarity. *Ultrafest VIII, Potsdam, Germany*.

Shawker, T. H., & Sonies Phd, B. C. (1984). Tongue Movement During Speech: A Real-Time Ultrasound Evaluation. *Journal of Clinical Ultrasound*, 12(3), 125-133.

Sugden, E., Lloyd, S., Lam, J., & Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International journal of language & communication disorders*, 54(5), 705-728.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543-553.

Wrench, A. A., & Scobbie, J. M. (2011). Very high frame rate ultrasound tongue imaging. In *Proceedings of the 9th International Seminar on Speech Production (ISSP)* (pp. 155–162), Montreal, Canada

Zharkova, N. (2013). A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes. *Clinical Linguistics & Phonetics*, 27(6-7), 484-496.

Zharkova, N. (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*, 50(1), 76-81.

Zharkova, N., Gibbon, F. E., & Lee, A. (2017). Using ultrasound tongue imaging to identify covert contrasts in children's speech. *Clinical linguistics & phonetics*, 31(1), 21-34

Ultrasound Tongue Imaging

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics*, 37(2), 248-256.