

Joanne Cleland and James M. Scobbie

# Acquisition of new speech motor plans via articulatory visual biofeedback

**Abstract:** This chapter describes the concept of categorising persistent Speech Sound Disorder in children as a disorder characterised by erroneous motor plans. Different types of articulatory visual biofeedback are described, each of which is designed to allow children to view their articulators moving in real time and to use this information to establish more accurate motor plans (namely, electropalatography, electromagnetic articulography and ultrasound tongue imaging). An account of how these articulatory biofeedback techniques might lead to acquisition of new motor plans is given, followed by a case study of a child with persistent velar fronting who acquired a new motor plan for velar stops using ultrasound visual biofeedback.

**Keywords:** visual feedback, articulation, Speech Sound Disorders, electropalatography, ultrasound, electromagnetic articulography

## 1. Introduction

Children with Speech Sound Disorders (SSD) have difficulty acquiring the speech sounds of their native language in the course of normal development; producing certain sounds incorrectly, substituting them with other sounds or omitting them altogether. SSDs are the most common type of communication impairment; around 11.5 % of eight-year olds (Wren, Miller, Emond, & Roulstone, 2016) have SSDs ranging from common distortions such as lisps and /r/ distortions to speech that is unintelligible even to close family members.

For many children, the cause of their SSD is unknown (though SSDs are also associated with a range of conditions including hearing impairment and cleft palate) and is usually thought to arise from a difficulty acquiring the phonology of their ambient language. Indeed, most children with SSDs have “phonological” impairments (87.5 % in an analysis of caseload referrals by Broomfield & Dodd, 2004). It appears that a lesser number (12.5 % of caseload) have “articulation disorders”, in that they more clearly have a problem producing certain (normally late-acquired)

speech sounds. Overall, the problem is thought to be mainly cognitive, so that children have difficulty learning the patterns of their language which often leads them to display the simplification processes representative of an earlier age in typical development, for example by reducing clusters or replacing velars with alveolars, resulting in phonological merger.

In therapy, the resulting homophony motivates remediation in part by confronting children with their inability to signal contrast. There is good evidence that in young children these auditory-based phonological interventions, for example minimal pairs intervention (Law, Garrett & Nye, 2003) are very effective. However, in around half of children with SSDs the problem persists into the school years, and a smaller number still become “intractable”, beyond the age of eight. There is growing evidence that these children may not have a purely cognitive phonological disorder, but display (also) subtle motor problems. For example, Wren et al. (2016) found that weak sucking at six weeks of age is a risk factor for SSD at eight years of age. These types of potentially motoric speech impairments need interventions that capitalise on the principles of motor learning (see Maas et al., 2008 for a tutorial). Children with ingrained incorrect motor plans (for example, children who persistently misarticulate certain phonemes) need motor-based techniques for teaching and practicing new articulatory gestures.

In the motor-learning literature, the ontogeny of complex movements is studied by looking at an individual’s ability to imitate a novel movement (Paulus, 2014). This is problematic for children who haven’t acquired articulatory gestures via the normal auditory route because the main articulator, the tongue, is largely hidden from view. Researchers and clinicians have therefore sought to circumvent this problem by augmenting the acoustic (and tactile) information already available to the speaker through the use of instrumental imaging technologies conveying aspects of vocal tract articulation directly to the speaker, that is, by providing biofeedback.

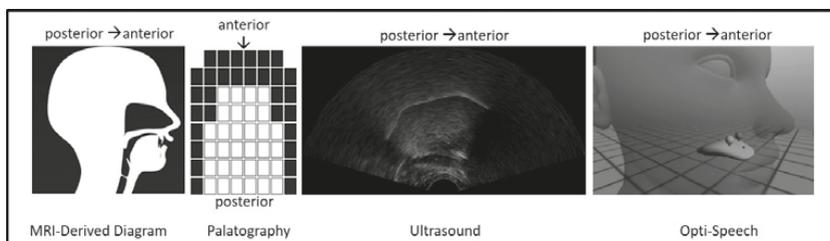
## 2. Articulatory feedback approaches

In phonetics the use of instrumental techniques to measure movement of the articulators has a longer history than of sound recordings being used to measure acoustics, beginning with static palatography in the late 18th

century through to cine-Magnetic Resonance Imaging (MRI) in recent years. Techniques like electropalatography (EPG) and electromagnetic articulography (EMA) are well established, with ultrasound and MRI gaining popularity thanks to methodological improvements and falling costs. All of these techniques give researchers data that can be used to create visual images of otherwise invisible articulators, especially the tongue. However, only a small number allow data to be visualised in real time in a way that is immediately meaningful to the viewer, namely EPG, EMA and Ultrasound Tongue Imaging (UTI). Since the 1980s (Dagenais, 1995) the potential for using visualisations of the articulators as a powerful speech therapy tool has been explored. Most of the research to date has focussed on EPG, with a large number of “small n” studies showing its potential as a visual biofeedback (VBF) device (Gibbon, 2013).

EPG is a technique for displaying the timing and location of tongue-palate contact (Hardcastle & Gibbon, 1997). The speaker sees an abstract representation (Figure 1) of linguo-palatal contact, which is very useful for conveying aspects of coronal (and dorsal) consonants (and some vowels) in real time, and is encouraged to use this to modify their own erroneous articulations. It is worth noting that the display in EPG is normalised. All speakers see the same display irrespective of the size and shape of their hard palate. This potentially makes the display easier for the Speech and Language Therapist (SLT) to interpret. Additionally, the anterior third of the EPG palate is displayed in the anterior half of the normalised computer display. This is because the tongue-tip (the part most often in contact with the anterior part of the palate) contains more nerve endings and achieves more fine-grained articulation. While the  $\frac{2}{3}$  to  $\frac{1}{2}$  ratio is arbitrary, the understanding of this visual display is thought to be relatively intuitive (Gibbon & Wood 2010), even for those with cognitive impairment (Cleland et al. 2009).

While EPG shows tongue-palate contact rather than visualising the articulators directly, EMA shows the movements of a small number of specific flesh-points. Sensors are directly attached (glued) to articulators such as the jaw, lips, and (crucially) the tongue, and can be visualised in real time on a computer screen (Figure 1). While EPG shows 62 points of contact on the hard palate, EMA normally tracks a much more limited number of points: usually three sensors attached near to the midsagittal



**Figure 1:** Instrumental articulatory technique displays (not recorded simultaneously). From left to right: MRI-derived animation (produced with permission from Eleanor Lawson), electropalatography, Ultrasound, Opti-Speech (electromagnetic articulography).

tongue tip, then two more on the front of the tongue, about 1.5cm and 3cm posterior (Katz & Mehta, 2015) which is about as far into the anterior oral cavity as can be reached easily. More recent systems, for example the Wave Electromagnetic Speech Research System (NDI, Waterloo, ON) allow three-dimensional tracking of five small sensors affixed to the client's tongue. Software such as "Opti-Speech" (Vick, Mental, Carey, & Lee, 2017) shows the sensors in the context of an avatar (see Figure 1).

EMA has been popular in articulatory phonetics studies because it is one of the few techniques which allows velocity and acceleration of movements to be calculated and interpreted easily, because of the flesh point tracking. However, it is not likely that speakers control speech production in terms of a small number of such points, nor that in experimental studies the most meaningful points are selected, nor studied in a replicable manner. In terms of biofeedback, EMA has not been particularly popular: the equipment is expensive, positioning the sensors on the articulators requires training, and it is potentially invasive, especially for children. However, a small number of studies have shown it to be potentially useful for VBF. Katz and Mehta (2015) evaluated the technique for teaching native speakers of American English to produce the non-English segment [d]. In this study, the Opti-Speech system was used to display the EMA sensors superimposed on an animated avatar showing the tongue in a mid-sagittal head context. Target areas for the sensors were also shown, and on-target articulations were highlighted by changing the sensor colour from red to

green. Results indicated a rapid gain in accuracy associated with visual feedback training. However, extrapolating from these results into the clinical domain should be interpreted with caution for three reasons: firstly, the speakers did not have SSDs; secondly, the speakers were not asked to integrate the new articulation into words; and lastly a similar experiment by Cleland, Scobbie, Nakai, and Wrench (2015) using ultrasound showed that retroflexes were just as easy to teach to English-speaking children using auditory methods as they were with VBF.

To date, just one study has used the Opti-Speech (EMA) system to treat residual speech errors in children and young people. Vick et al., (2017) treated residual /s/ (two children) and /r/ (two children) distortions. Early results showed that it is possible to use the technique to remediate these errors, and that generalisation can occur. However, further research is needed to determine the effectiveness of EMA for treatment of SSDs and also to determine whether clinicians in the field find this technique useable in the practical sense.

In contrast to these studies which use direct EMA displays of the real-time movements of sensors, more recent research has sought to gamify the articulatory information, again in (near) real time. Yunusova et al. (2017) used a single tongue tip sensor to drive a computer game in which the object was for a dragon character to breathe as much fire as possible. The size of the dragon's flames was directly related to the size of the speaker's articulatory working space (AWS). In this case, the augmented VBF was designed with a very specific population in mind: speakers with Parkinson's disease. This particular neurodegenerative condition causes a reduction in articulatory movements (causing dysarthric symptoms such as undershoot) and leads to reduced intelligibility. By providing a metaphor (the fire-breathing dragon) which visually produces more fire in correlation with increasing AWR, speakers with Parkinson's disease were able to use the feedback to increase their intelligibility. Increasing the strength and range of movements which already follow the correct articulatory trajectory is, however, quite different from establishing a correct gesture in replace of an erroneous one (for example, a central fricative produced laterally), or an absent one (for example, in someone who has no velars in their phonetic inventory). Therefore, any gamification of VBF designed for establishing new articulations is likely to need games which relate more

directly to the trajectory of a specific segmental gesture rather than to the global magnitude of change during the production of a word.

In contrast to EPG and EMA, which show a discrete number of points, U-VBF shows an anatomically accurate speaker-specific representation of the tongue. With this technique most of the surface of the tongue is visible in a mid-sagittal view (Figure 1), and interpretation of the images is thought to be relatively intuitive (Bernhardt et al. 2005). In contrast to EPG, the image is an anatomically correct representation of part of the tongue, however, other important anatomical information, such as the relation of the tongue to the hard palate, is not normally visible (Cleland et al., 2019). Moreover, this “raw” ultrasound suffers from artefacts, and the tip of the tongue is often in shadow from the mandible. However, ultrasound has practical advantages over EPG and EMA in that it does not require expensive individual artificial palates or expensive sensors. Moreover, since it involves no intra-oral equipment it is less physically invasive, potentially making it more suitable for children.

Given the practical limitations of EMA most of the clinical studies in the literature have used EPG and, more recently Ultrasound-VBF. Indeed, U-VBF is rapidly gaining popularity, probably because of its lower cost and because more portable high-speed ultrasound systems are now available. To date, 29 small studies have been published in the literature investigating the efficacy of U-VBF (see Sugden, Lloyd, Lam and Cleland, 2019 for a systematic review). Of these studies, 27 were published in the last 10 years and 17 in the last three. While larger clinical trials of both EPG and UTI are needed in the future, it is essential to know theoretically why and how these techniques work because identifying the agents of change (the “active ingredients”) in an intervention is essential for refining the intervention and establishing dosage.

None of these instrumental techniques are therapies in their own right (Bacsfalvi et al. 2007); most SLTs use them to supplement traditional techniques, such as articulation therapy (Van Riper & Emerick, 1984) or motor-based intervention (Preston et al., 2013). One key ingredient of articulatory VBF is that it can be used to demonstrate complex articulations that are normally difficult to describe. Describing articulatory movements is an essential part of traditional articulation therapy (Van

Riper and Emerick, 1984). Normally this is done with verbal descriptions, or perhaps diagrams, ranging from impromptu sketches to computer animations.

It is crucial, moreover, to unpick the visual model aspect of EPG/UTI from the biofeedback aspect. That is, we need to know the extent to which a speaker benefits from informative general visual models of articulation, and the extent to which real-time biofeedback of the learner's own tongue during speech production provides crucial additional information.

Considering first the model aspect on its own, studies which investigate the use of an articulatory model to teach new speech sounds are few. Massaro et al. (2008) used a "Talking Head" to teach native English speakers a new vowel [y] and consonant [q]. Talking Heads are artificial animations of speech usually based ultimately on instrumental (e.g. MRI or EMA) data. Some are 3D (e.g. Badin & Serrurier, 2006) and some are 2D (e.g. Kröger et al., 2013), but most attempt to model the movement of the tongue during speech with a cut-away profile or mid-sagittal view of the tongue.

The main application of Talking Heads is usually as a teaching tool for pronunciation training in second language learning (Cleland et al., 2013). However, there is little evidence that this is effective. In the Massaro et al. study (2008) a view of the lips was useful for teaching the high-front rounded vowel [y] but a mid-sagittal Talking Head did not improve learning of the distinction between [k] and the uvular stop [q]. There is a confound here, however, due to one study involving a segment where lip-rounding is the defining feature and one where it is uvular place: lip reading is not only a natural phenomenon but one known to improve perception of speech (see below). Similarly, a study by Fagel and Madany (2008) which used a Talking Head to teach [s] and [z] to children with interdental lips failed to show an effect. Thus, a visual model alone appears not to be the essential ingredient for success. However, since the above studies did not give the learners any information about closeness to target (e.g. from a human judge or automatic speech recognition), and since articulatory constriction is a key feature of production, further study is required to directly compare an articulatory model against VBF using the same type of display and mediation.

### 3. Theoretical explanations for the role of biofeedback in learning new articulations

Children who make inappropriate phonetic realisations of certain speech sounds do so because they have an inappropriate motor plan for that sound (Preston et al., 2014; Cleland et al., 2019). Cleland, Scobbie and Wrench (2015) suggest that these erroneous motor plans can be ascribed to one of three categories: 1. It is identical to that of another phoneme, resulting in perceived homophony (as in canonical velar fronting); 2. the motor plan is abnormal or underspecified resulting in something which is perceived as homophonous but is subtly different in some way (as in covert contrast, Gibbon & Scobbie, 1997), for example /t/= [t] and /k/= [t̚] or; 3. the motor plan is abnormal to the extent that it results in the realisation of an obviously non-native speech sound, for example a lateral lisp in English-speaking children. It is possible that different types of VBF are needed to overcome each of these erroneous motor plans. In the case of category 1, normally a phonological cause would be ascribed, however Cleland et al. (2017) present several cases of children with persistent velar fronting with identical tongue-shapes for /t/ and /k/ but awareness of the error and (initially) an inability to produce a velar articulation of any type. In these, and other cases, the inability to produce the correct articulatory gesture upon imitation is often coupled with a lack of understanding (despite previous intervention) of how the gesture is achieved at all, with one of the children in the Cleland, Scobbie and Wrench (2015) study stating that she thought producing a velar was “impossible” the first time she viewed an ultrasound movie of that segment, highlighting the lack of understanding she had as to the movements required to achieve a velar despite previous therapy targeting this very sound (Cleland et al., 2019).

In addition to a lack of explicit understanding about the movements required to achieve a particular sound, there may be some implicit learning involved in the viewing of tongue movements. In typical audio-visual speech perception, viewing the speaker’s lips enhances perception, particularly in noise (Benoît & Le Goff, 1998). Typical speakers integrate lip information into their perceptual system, as shown by the McGurk effect (McGurk & MacDonald, 1976). Clearly whilst lips are easily visible during interactions, the tongue is not. Even so, Badin, Tarabalka,

Elisei, and Bailly (2010) suggest that it is possible to “tongue-read” in the same way as it is possible to lip-read. That is, viewing a Talking Head of tongue movements leads to better discrimination of speech in noise and potentially could be used for learning new articulations. Badin et al. (2010) hypothesise that this is due to a natural, intuitive ability for listeners/viewers to tongue-read, suggesting that this provides support for a perception/production link which could relate to the theory of mirror neurons (Cleland et al., 2019). Mirror neurons are thought to underlie the imitation system, because they are neurons that fire when a person both sees an action being performed (or hears it being performed, in which case they may be called echo neurons) and performs that action themselves. So, in theory, when a person hears a speech sound, the neurons in the motor area required for articulating that speech sound fire. In fact, even passive listening to speech sounds evokes a pattern of motor synergies mirroring those occurring during speech production (D’Ausilio, Bartoli, Maffongelli, Berry & Fadiga 2014). There is emerging evidence that this does not just apply to hearing a speech sound, but also to seeing it. Treille, Vilain, Hueber, Schwartz, Lamalle and Sato (2014) showed activation in the premotor and somatosensory cortices when observing lingual movements from ultrasound, suggesting that demonstration of correct articulatory movements may be a crucial aspect of visual biofeedback. Moreover, using delayed U-VBF might evoke the same process. In this type of feedback, the child (as well as watching the live visual biofeedback) watches their own production replayed after a delay (once they have finished speaking, not to be confused with delayed auditory feedback, which has very short delay times). The SLT then encourages the child to reflect on the correctness of their production. While viewing their own *incorrect* production could potentially have an adverse effect, viewing their own correct production gives a speaker-specific representation of the required articulatory gesture.

Whilst it would be unethical and ethically dubious to compare U-VBF without demonstration to U-VBF with it, it would be feasible to conduct a randomised control trial where one arm of the trial involved the use of an ultrasound-based visual articulatory model, without biofeedback (Cleland et al., 2019). Indeed, a small study of speakers with cleft palate (Roxburgh, 2018) found that the children did just as well with a visual

articulatory model to learn new articulations as they subsequently did with U-VBF. However, this study was limited by a small sample size of just two participants, and that neither had had previous therapy to address the relevant speech problem (i.e. they were not ‘intractable’, Cleland et al., 2019).

The question remains as to how VBF, or indeed a visual model alone, could lead to acquisition of *new* articulations, especially when, in the case of intractable SSDs, the speakers have been exposed to extensive models of the correct articulation from other speakers, albeit only in auditory form. It seems in this case that the auditory imitation system has failed somehow, perhaps enabling the visual modality to offer useful new information. Indeed, evidence exists that the observation of completely novel behaviour (in this case a previously unseen articulatory movement) generates mirroring activity in the premotor cortex (Cross, Hamilton and Grafton, p. 11, 2006). Moreover, Mattar and Gribble (2005) show that complex motor behaviours, which speech undoubtedly is, are greatly assisted by first observing another engage in the activity. Via this mechanism, models of the new activity are formed in the premotor cortex via the mirror neurons and presumably intensity of neuronal firing increases with practice/exposure. It is not enough to simply watch the new movement repeatedly and expect acquisition of a new motor plan: practice is required by the speaker. (Imagine trying to learn the piano only by watching videos of a pianist’s fingers!) Del Giudice, Manera and Keyzers (2009, p. 352) explain the mechanisms by which practice of movements leads to acquisition, by looking at grasping: “activity in the premotor cortex leads to a grasping movement. The movement is *seen by the acting individual*, causing activity in neurons in the temporal cortex. This activity is sent to the parietal and premotor cortex, where it finds neurons that are active because the subject is currently performing the action. This leads to Hebbian enhancement of the congruent connections from temporal to parietal and from parietal to premotor neurons representing the same action; incongruent connections do not undergo such enhancement”. It is therefore conceivable that seeing a novel speech motor movement leads to development, or otogeny, of the mirror neuron whilst actually *doing* the novel tongue movement yourself leads to Hebbian enhancement, which in turn is enhanced by lingual visual biofeedback. Repeated association of the sound (knowledge of results) with the movement (knowledge of performance) leads to enhancement in

acquisition of the new skill. Of course, this ought to be entirely possible with only the articulatory model, provided the speaker is able to practice accurately, and biofeedback may not be required. However, it is likely that some individuals are unable to make the leap between seeing the new articulation and beginning to practice it themselves, that is, no matter how many times they see it they cannot perform it, or even approximate a performance of it. In this case the speech and language therapist too benefits from the visual feedback as s/he is able to use shaping techniques (Bleile, 2004) to explicitly demonstrate to the speaker that similar motor programmes are already within their grasp.

Evidence for the *biofeedback* aspect of U-VBF comes from experiments on experiential canalised learning. Canalisation is the means by which a developmental process is buffered against perturbations. It ensures that important features of the organism emerge reliably despite great variation between individuals in environmental conditions and genotypic makeup. The classic example is that of ducklings raised in incubators which still spontaneously exhibit the 'correct' preference for their own species' maternal calls, despite never hearing a mother duck. However, if the ducklings are prevented from hearing their own vocalizations, they fail to exhibit selective responses to maternal calls (Gottlieb, 1991) suggesting a key factor is self-produced vocalizations. That is, the speaker must *make* the articulatory movements themselves and evaluate the acoustic output in order to acquire them. Visual biofeedback offers a new modality for learners who have failed to acquire speech sounds via the normal routes. Moreover, in live *bio-feedback* the speaker is able to bootstrap the new visual modality not only onto the auditory modality but also onto the haptic modality to make small adjustments to their articulatory gestures in real time. In the speech therapy clinic this is evidenced by articulatory groping towards the target in the early stages of intervention.

In sum, U-VBF works by first showing the learner what is to them a novel movement, then performance of the new movement leads to Hebbian learning, which is boosted by the visual knowledge of performance provided by U-VBF, this leads to increasing activation of the mirror neuron, laying down of a new general motor programme and hence eventually mastery of the new sound. If the mastery of the new sound is a gradual process then we might expect to detect various types of phonetic gradience in the

short-term longitudinal change, potentially in addition to rapid categorical change. Some evidence of incomplete generalisation of a new articulation is shown in U-VBF studies where post-intervention scores for target segments are lower than 100 % correct. For example, Cleland, Scobbie, Roxburgh, Heyde and Wrench (2019) show that after intervention children with a wide variety of lingual errors show improvements in accuracy of targeted gestures, but no child achieved perfect percent target consonants correct in all phonotactic contexts. However, the approach of categorising segments within words as correct or incorrect obscures the potential subtlety of the process. More important for understanding the pathway to acquisition is the fine detail necessary for a full evaluation of new articulations produced by children as the result of clinical intervention.

For example, consider the two children reported by Cleland et al. (2019) who made progress towards the target, changing posterior (pharyngeal fricatives for sibilants) to anterior articulation, but with incorrect lateral airflow. For these children, the updated motor plan is more accurate, since in it contains more of the correct features of the target, even though the output is still wrong linguistically. The motor plan has therefore changed in a gradient manner, as both children also show progress towards achieving the correct airflow. However, gradient acquisition of targets may manifest differently in each of the three erroneous motor plans 1. Motor plans identical to another sound; 2. Motor plans which are covertly different but perceived as a different sound and 3. Motor plans which result in a non-native sounding phone. Type one is particularly interesting, because in a traditional model these children would be said to have classic substitution errors, thought to be phonological in nature. If this were the case, we would not expect these children to acquire a new articulation in a phonetically gradient manner (though they may acquire it in some phonotactic conditions before others as is the case in typical acquisition of a segment).

What follows is a case study of a child who presented with a classic substitution error who nevertheless shows gradient change during remediation. Rather than presenting only binary information on the correctness of her new articulations, which would obscure more subtle changes, we explore the process in more articulatory detail *during* the therapeutic process.

#### 4. An illustration of gradient acquisition of a new articulation

While typically developing children are usually able to produce velars correctly by the age of three and a half years (Dodd, 2013), those with SSDs may not be able to produce velars till much later. A lack of velars in a child's phonetic inventory has been recognised as a prognostic indicator for a phonological disorder (Grunwell, 1987). Children who persistently fail to differentiate coronal and dorsal articulations may therefore have an underlying motoric deficit. Gibbon (1999) suggests that this may manifest as an "Undifferentiated Lingual Gesture" (ULG), where the tongue moves as a whole, rather than, as expected, by executing gestures using independent parts. Children with UGs show abnormally extensive tongue-palate contact patterns in EPG studies (Gibbon 1999) and (in just one study to date) abnormal dorsal raising in ultrasound (Cleland et al., 2017). This error pattern is motoric, rather than phonological.

While there are studies showing these abnormal articulations, there are no studies showing how articulations change as children initiate a coronal/dorsal differentiation or achieve mastery of it. In several of our previous studies (Cleland et al., 2015b, 2017, 2019) we reported on children who persistently front velars to alveolars, despite being over six years of age. Velar fronting is readily remediated using U-VBF, with some children showing a categorical shift from 0 % velars correct pre-therapy to 100 % post-therapy. Speaker "07F\_Ultrax" is reported in Cleland et al. 2015 and 2017. At the time of the U-VBF intervention she was aged 7;6 and presented with velar fronting in the absence of a history of any other errors. Pre-intervention, she produced no correct velars, half-way through intervention she was not perceived to produce any correct velars, but 6 weeks later, at the end of the intervention period, she produced 100 % correct velars in a word list designed to probe this segment in multiple phonotactic positions. She maintained that gain three months later. Prior to intervention she produced both /t/ and /k/ with identical tongue shapes, in other words, a classic merger (see Cleland et al., 2017) appears to have been almost instantly fixed. We turn our attention now to an ultrasound analysis of 07F's productions of alveolars and velars at various time-points in the intervention process.

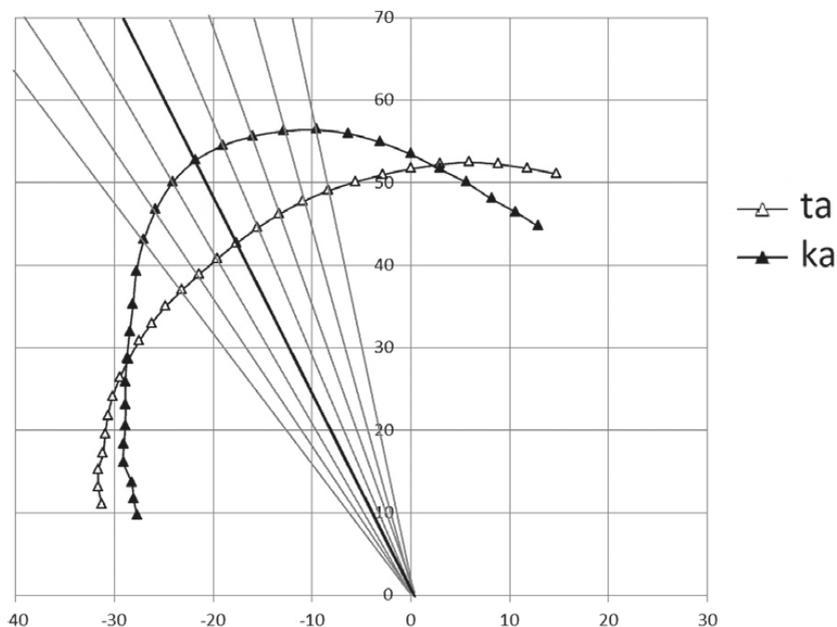
07F\_Ultrax was recorded with simultaneous high-speed ultrasound and audio. The ultrasound was probe-stabilised with a headset (Scobbie, Wrench & Van der Linden, 2008) to allow us to compare tongue shape for /t/ and /k/ directly. Materials were a wordlist containing velars in a wide range of vowel environments and word positions.

Using AAA v2.16 software (Articulate Instruments, 2012) /t/ and /k/ segments were annotated at the beginning of the burst, the nearest ultrasound frame was then selected and a spline indicating the tongue surface fitted to the image using the semi-automatic edge-detection function in AAA software. Splines were then averaged by target segment and compared.

In this case, we are interested in the degree of separation between /t/ and /k/. If 07F presents with merged productions of /t/ and /k/, then we would expect to see no degree of separation between /t/ and /k/ and if she presents with ULGs for both, then we might expect a reduced degree of separation between /t/ and /k/ compared to typically developing children. The difference between /t/ and /k/ can be characterised as maximum radial dorsal difference between these two segments (Figure 2).

Scobbie and Cleland (2017) report the average maximum width of the radial difference between /t/ and /k/ at mid-closure for 30 typically developing children as 11.9mm, 7.5mm and 12.1mm for symmetrical /a/, /i/ and /o/ contexts respectively.

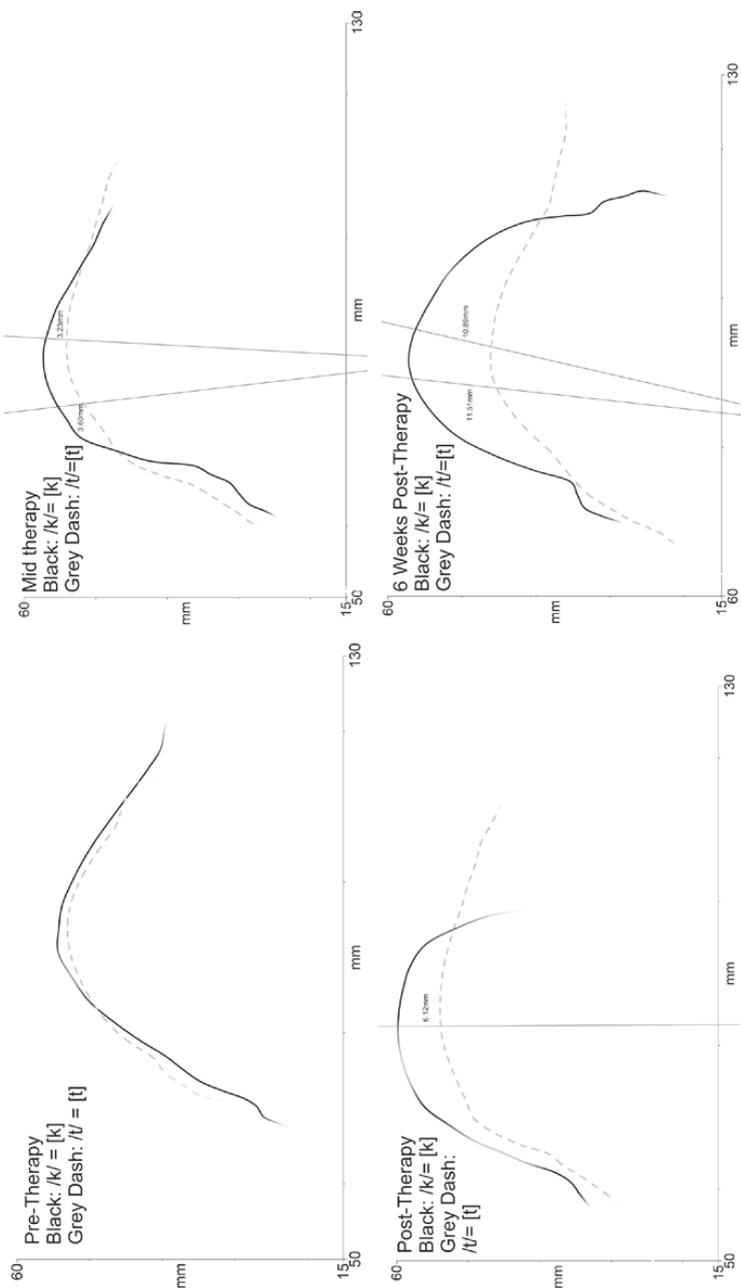
By applying the same measurements (Figure 3) to all the time-points from 07F's data, we can quantify the gradient increase in the degree of separation between /t/ and /k/ at each time point (Figure 4). What is interesting, is that by looking only at percent target consonants correct, 07F appears to make a categorical shift from 0 % to 100 % correct between mid-therapy and post-therapy sessions, whereas in fact she was already beginning to change the production by the mid-therapy session (panel 2) while in the post-therapy session (panel 3) her coronal/dorsal differentiation (6.12mm) actually remained abnormally small. Presumably with practice, as is consistent with the motor learning literature, over time her articulations become more phonetically accurate, until the point where /t/ and /k/ are perceived by a listener as occupying different perceptual categories.



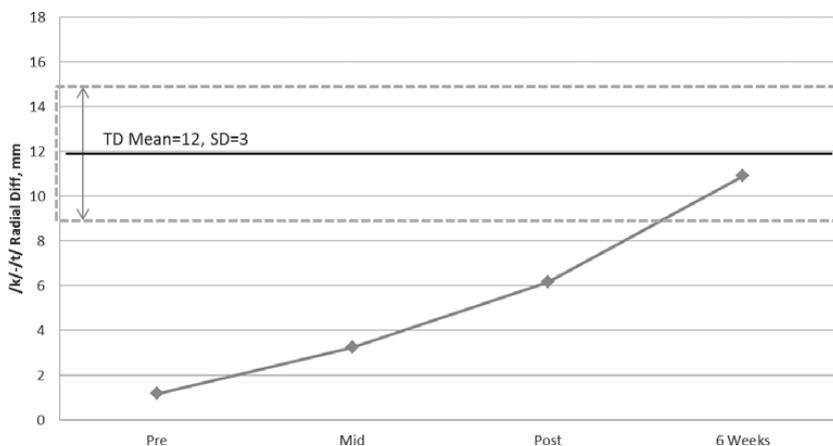
**Figure 2:** Average /t/ and /k/ from 30 typical children at mid-closure. The diagonal spokes are some of the radial fanlines (emanating from the probe's virtual centre) used for measurement. For each individual child the maximum distance /k/-/t/ along some fanline (in this case, the 4th diagonal line from the left) within the anterior and posterior crossing points of the splines for each individual child is taken as the degree of coronal-dorsal differentiation.

## 5. Conclusion

Since the 1980s instrumental phonetic techniques have increasingly been applied as biofeedback for learning new articulations in children who have failed to acquire particular phones through the normal route. While EPG has dominated the literature as the technique of choice, and has been shown to be successful for a large number of children, recent studies have focussed on ultrasound visual biofeedback. For the most part VBF is described as a motor-learning approach, though it is often used with children who present with errors described as “phonologically delayed”. The case study above shows that even in these cases, evidence of subtle



**Figure 3:** /k/ (black) and /t/ (grey dashes) attempts over time (L-R): pre, mid, post, 6 weeks post intervention. Increased separation between /k/ and /t/ can be seen, but is only at 6 weeks post intervention that /k/ is perceived as distinct from /t/.



**Figure 4:** Max radial difference of /k/-/t/ for 07F over time. Y-axis, radial difference between /k/ and /t/, x-axis intervention time point. Grey dashed box: expected radial difference between /k/ and /t/ for typically developing children.

motor-impairments can exist. This calls into question the underlying impairment these children have. However, we wish to caution the reader from drawing the conclusion that all children with “phonological delay” in fact have motor-based problems. Evidence from a large study by Wren et al. (2016) shows that early signs of subtle motor impairment such as weak sucking at six weeks of age, predicts *persistent* SSDs, and not SSDs which remediate in the preschool years. It therefore seems plausible that children with persistent disorders, as exemplified here, are a different subgroup from the outset.

The agents of change in VBF remain underexplored. There are at least four different potential “active ingredients” in VBF therapy that do not exist in traditional approaches: 1. Improved diagnostic information provided by articulatory analysis prior to intervention; 2. An accurate visual articulatory model provided by target patterns/tongue movements; 3. Increased accuracy of positive feedback from the treating SLT made possible by viewing movements; 4. Biofeedback. In reality a combination of all these factors likely impacts on the ability of children to achieve, practice, and ultimately generalise new articulations following biofeedback interventions.

## References

- Articulate Instruments Ltd 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.
- Bacsfalvi, P., Bernhardt, B.M. & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech-Language Pathology*, 9(1), 36–45.
- Badin, P. & Serrurier, A. (2006). Three-dimensional linear modelling of tongue: Articulatory data and models. In H.C. Yehia, D. Demolin, R. Laboissière (Eds.), *Seventh International Seminar on Speech Production, ISSP7*. Ubatuba, SP, Brazil, UFMG, Belo Horizonte, Brazil, 395–402.
- 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*, 52(6), 493–503.
- Benoît, C. & Le Goff, B. (1998). Audio-visual speech synthesis from French test: Eight years of models, designs and evaluation at the ICP. *Speech Communication*, 26(1–2), 117–129.
- Bernhardt, B., Gick, B., Bacsfalvi, P. & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics and Phonetics*, 19(6–7), 605–617.
- Broomfield, J. & Dodd, B. (2004). Children with speech and language disability: Caseload characteristics. *International Journal of Language and Communication Disorders*, 39, 303–324.
- Beile, K.M. (2004). *Manual of articulation and phonological disorders: Infancy through adulthood*. Cengage Learning.
- Cleland, J., McCron, C., & Scobbie, J.M. (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics and Phonetics*, 27(4), 299–311.
- Cleland, J., Scobbie, J.M., Heyde, C., Roxburgh, Z., & Wrench, A.A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics and Phonetics*, 31(1), 35–55.
- Cleland, J., Scobbie, J.M., Nakai, S., & Wrench, A.A. (2015a). Helping children learn non-native articulations: the implications for ultrasound-based clinical intervention. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. Paper number 0698, Susanne Fuchs, Joanne Cleland and Amélie Rochet-Capellan - 9783631797884  
Downloaded from PubFactory at 01/21/2021 10:48:04AM  
via free access

- Retrieved February, 2, 2017 from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0698.pdf>
- Cleland, J., Scobbie, J.M., Roxburgh, Z., Heyde, C., & Wrench, A.A. (2019). Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback. *Journal of Speech, Language, and Hearing Research* 62(2), 229–246.
- Cleland, J., Scobbie, J.M., & Wrench, A.A. (2015b). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics and Phonetics*, 29(8–10), 575–597.
- Cleland, J., Timmins, C., Wood, S.E., Hardcastle, W.J. & Wishart, J.G. (2009). Electropalatographic therapy for children and young people with Down's syndrome. *Clinical Linguistics and Phonetics*, 23(12), 926–939.
- Cross, E.S., Hamilton, A.F.D.C., & Grafton, S.T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage*, 31(3), 1257–1267.
- Dagenais, P. (1995). Electropalatography in the treatment of articulation/phonological disorders. *Journal of Communication Disorders*, 28(4), 303–329.
- D'Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J., & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philosophical Transactions of the Royal Society B*, 369(1644), 20130418.
- Dodd, B. (2013). *Differential diagnosis and treatment of children with speech disorder*. Chichester: John Wiley & Sons.
- Fagel, S. & Madany, K. (2008). A 3-D virtual head as a tool for speech therapy for children. In *Ninth Annual Conference of the International Speech Communication Association*. Brisbane, Australia, 2643–2646.
- 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 (2013). *Bibliography of electropalatographic (EPG) Studies in English (1957–2013)*. Retrieved November, 12, 2018 from <http://www.articulateinstruments.com/EPGrefs.pdf>
- Gibbon, F., & Scobbie, J.M. (1997). Covert contrasts in children with phonological disorder. *Australian Communication Quarterly*. (Autumn), 13–16.

- Gibbon, F.E. & Wood, S.E. (2010). Visual feedback therapy with electropalatography. In: Williams, A. L., McLeod, S. and McCauley, R.J. (Eds.) *Interventions for speech sound disorders in children*. Baltimore: Paul H. Brookes Pub, pp. 509–532.
- del Giudice, M.D., Manera, V., & Keyzers, C. (2009). Programmed to learn? The ontogeny of mirror neurons. *Developmental Science*, 12(2), 350–363.
- Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. *Developmental Psychology*, 27(1), 4–13.
- Hardcastle, W., & Gibbon, F. (1997). Electropalatography and its clinical applications. In M. Ball, & C. Code (Eds.), *Instrumental Clinical Phonetics* London: Whurr, pp. 149–193.
- Katz, W.F., & Mehta, S. (2015). Visual feedback of tongue movement for novel speech sound learning. *Frontiers in Human Neuroscience*, 9, 612.
- Kröger, B.J., Gotto, J., Albert, S., & Neuschaefer-Rube, C. (2013). *A visual articulatory model and its application to therapy of speech disorders: a pilot study*. Universitätsbibliothek Johann Christian Senckenberg.
- Law, J., Garrett, Z. & Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database of Systematic Reviews*, 3. Art. No.: CD004110.
- Maas, E., Robin, D.A., Hula, S.N.A., Freedman, S.E., Wulf, G., Ballard, K.J., & Schmidt, R.A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, 17(3), 277–298.
- Massaro, D., Bigler, S., Chen, T., Perlman, M., & Ouni, S. (2008). Pronunciation training: The role of ear and eye. *Ninth Annual Conference of the International Speech Communication Association*, 22–26 September, Brisbane, Australia, 2623–2626.
- Mattar, A.A., & Gribble, P.L. (2005). Motor learning by observing. *Neuron*, 46(1), 153–160.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264 (5588), 746–748.
- Paulus, M. (2014). How and why do infants imitate? An ideomotor approach to social and imitative learning in infancy (and beyond). *Psychonomic Bulletin & Review*. 21(5), 1139–1156.

- Preston, J.L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 22(4), 627–643.
- 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, QMU Edinburgh, UK.
- Scobbie, J.M. & Cleland, J. (2017). Dorsal crescents: Area and radius-based mid-sagittal measurements of comparative velarity. Paper presented at Ultrafest VIII, Potsdam, 4th–6th October 2017. Potsdam: University of Potsdam.
- Scobbie, J.M., Wrench, A.A., & van der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, 373–376.
- 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 and Communication Disorders*. <https://doi.org/10.1111/1460-6984.12478>
- Treille, A., Vilain, C., Hueber, T., Lamalle, L., & Sato, M. (2017). Inside speech: Multisensory and modality-specific processing of tongue and lip speech actions. *Journal of Cognitive Neuroscience*, 29(3), 448–466.
- Van Riper, C., & Emerick, L.L. (1984). *Speech correction: An introduction to speech pathology and audiology*. Englewood Cliffs, NJ: Prentice-Hall.
- Vick, J., Mental, R., Carey, H., & Lee, G.S. (2017). Seeing is treating: 3D electromagnetic midsagittal articulography (EMA) visual biofeedback for the remediation of residual speech errors. *The Journal of the Acoustical Society of America*, 141(5), 3647–3647.
- Wren, Y., Miller, L.L., Peters, T.J., Emond, A., & Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research*, 59(4), 647–673.
- Yunusova, Y., Kearney, E., Kulkarni, M., Haworth, B., Baljko, M., & Faloutsos, P. (2017). Game-based augmented visual feedback for enlarging speech movements in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, 60, 1818–1825.

