Journal of Speech, Language, and Hearing Research. Accepted 20/09/2018.

Enabling New Articulatory Gestures in Children with Persistent Speech Sound Disorders using Ultrasound Visual Biofeedback.

Joanne Cleland[1], James M.Scobbie[2], Zoe Roxburgh[2], Cornelia Heyde[2], Alan Wrench[2&3]

[1]Psychological Sciences and Health, University of Strathclyde, Glasgow, UK, [2]Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, Edinburgh, UK, [3]Articulate Instruments Ltd, Edinburgh, UK

Corresponding Author: Joanne Cleland[1]

Address: Psychological Sciences and Health, 40 George Street, Glasgow, UK, G1 1QE.

Email: joanne.cleland@strath.ac.uk

Phone: +44 (0)141 548 3037

Conflict of Interest: There are no conflicts of interest.

Funding: This study was supported by a grant from the Chief Scientist Office of Scotland (*ETM/402)* awarded to James M. Scobbie and Joanne Cleland.

Abstract

Purpose: This study evaluated ultrasound visual biofeedback treatment for teaching new articulations to children with a wide variety of Speech Sound Disorders. It was hypothesized that motor-based intervention incorporating ultrasound would lead to rapid acquisition of a range of target lingual gestures with generalization to untreated words.

Method: Twenty children aged 6-15 with a range of mild to severe speech disorders affecting a variety of lingual targets enrolled in a case series with replication. Of these, fifteen children completed the intervention. All of the children presented with a variety of errors. We therefore employed a target selection strategy to treat the most frequent lingual error. These individual speech targets were treated using ultrasound visual biofeedback as part of 10 to 12 one hour intervention sessions. The primary outcome measure was percentage target segment correct in untreated wordlists.

Results: Six children were treated for velar fronting; three for post-alveolar fronting; two for backing alveolars to pharyngeal or glottal place; one for debuccalisation (production of all onsets as [h]); one for vowel merger; and two for lateralised sibilants. Ten achieved the new articulation in the first or second session of intervention despite no children being readily stimulable for their target articulation before intervention. In terms of generalization, effect sizes for percentage target segments correct ranged from no effect (five children); small effect (one child); medium effect (four children) and large effect (five children).

Conclusion: Ultrasound visual biofeedback can be used to treat a wide range of lingual errors in children with various speech sound disorders, from mild to severe. Visual feedback may be useful for establishing new articulations; however, generalization is more variable.

Introduction

Speech Sound Disorders (SSD) are the most common type of communication impairment, with recent figures suggesting 11.5% of eight year olds (Wren, Miller, Emond, & Roulstone, 2016) have SSDs ranging from common clinical distortions such as lisps and /r/ distortions to speech that is unintelligible even to close family members. For children in the preschool years there is good evidence that phonological delay/disorder can be remediated using auditory-based methods which focus on the likely root of the impairment, for example minimal pairs and core vocabulary therapy for the treatment of consistent and inconsistent phonological disorders respectively (Broomfield & Dodd, 2011; Law, Garett & Nye, 2003). However, SSDs can become persistent and intractable in older children if phonological-based interventions are unsuccessful. For these children, there is a growing body of evidence that visual biofeedback might be a promising way to re-program erroneous articulatory gestures by providing the client with a novel form of real-time articulatory feedback.

One such approach, ultrasound visual biofeedback (U-VBF), allows the client to view a real-time image of their own tongue. Though lending itself to motor-based therapeutic approaches, studies using U-VBF report success with various subtypes of SSDs (see below), not just those in which a motor impairment is a key causal factor, such as childhood apraxia of speech (CAS) and developmental dysarthria. In part, this broader success may be trivially due to difficulties surrounding accurate diagnosis of subtypes of speech disorders in children and differences between countries in use of classification systems (Waring & Knight, 2013). More interestingly, it may be that children diagnosed with articulation and phonological problems may in fact exhibit some difficulties with speech motor control even if they do not meet criteria for CAS or dysarthria. In fact, recent work by Shriberg (2017) suggests the category of “Speech Motor Delay” be added to classification systems. This speaks to a more gradient approach to the diagnosis of SSD subtypes, rather than the traditional dichotomy between motor/phonetic disorders and phonological disorders (for example, Dodd’s discrete classification system (1994) which is widely used in the UK). Given the theoretical and diagnostic uncertainty surrounding children with complex and persistent SSDs, in the current study we chose to include children with any diagnosis of SSD of unknown origin, for which previous intervention had failed to be completely successful, with the hypothesis that success with a motor-based approach would imply an underlying motoric deficit.

Ultrasound Visual Biofeedback: A Motor-Based Approach

U-VBF uses medical ultrasound scanners to image the tongue in real-time, allowing children to see their own tongue's profile moving in time with natural acoustic feedback, and thus use this additional source of information on speech production to modify erroneous tongue movements. In the motor-learning literature, viewing these types of movements is said to provide "knowledge of performance (KP)" (Maas et al., 2008; Preston, Brick & Landi, 2013 and Preston & Leaman, 2014), in addition to the "knowledge of results (KR)" provided in traditional articulation therapy approaches where children are able only to listen to their own productions and judge their correctness (the "result") auditorily.

While Preston et al. (2013) contrast KP and KR in terms of the external feedback given by the treating clinician, it should be noted that in terms of internal feedback, unlike other motor-learning tasks where the action may be dissociated from the result by a significant time delay, this dissociation is not possible in speech. The tongue movement is not separated (entirely) from the somatosensory feedback of articulation, nor from the acoustic consequence, and hence U-VBF adds visual information to what is already multi-modal learning with an articulatory underpinning. In addition to the more explicit and objective information on performance revealed to the speaker, U-VBF can provide the Speech and Language Therapist (SLT) with objective assessment detail that can reveal diagnostically-useful covert contrasts and covert errors if the data is recorded and analyzed (Cleland, Scobbie, Heyde, Roxburgh, & Wrench, 2017). Such sub-phonemic behaviors are also evidence of motoric rather than phonological difficulties- a clinically-relevant consideration which further motivates the retention of production data for post-hoc analysis. When viewed live, the instrumental images can be used by the SLT in comparison to targets and/or previous productions to provide more accurate feedback to the client (for example, avoiding inaccurate and/or inconsistent feedback from the SLT if the child's output is ambiguous or straddling the boundary between two auditory categories). U-VBF is therefore suited to visually-modulated multi-channel assessment and motor-based intervention approaches, in contrast to the auditorily assessed and mediated phonological approaches which have dominated the literature.

To date, over 30 small studies have been published in the literature investigating the efficacy of U-VBF (see Cleland and Isles [2018] for a comprehensive summary of client groups and speech targets). Of these studies, 25 were published in the last 10 years and 20 in the last four, suggesting this is an area of rapid growth. Most studies are single-case studies or single-case experimental designs. A notable recent exception to this is Furniss and Wenger (2018) who conducted a small randomized control trial to compare U-VBF to articulatory therapy of residual speech sound disorders. While both groups made good progress, the group who received ultrasound appeared to make much faster progress. It is therefore seems likely the success of U-VBF in leading to more rapid acquisition may be due to its underpinning as a motor-based intervention coupled with real-time visual biofeedback. Despite U-VBF normally being described as a motor-based intervention, only five studies have explicitly included speakers with motor-based SSDs (Childhood Apraxia of Speech in four studies, Preston et al., 2013, 2016a, 2016b, 2017 and one on acquired Apraxia of Speech in an adult Preston & Leaman, 2014). Even so, all studies show at least some improvement in outcomes following U-VBF, though some of the group studies report "non-responders" or difficulties with generalization. This highlights that patients are more likely to be selected for U-VBF on the basis of the surface-form of their SSD, i.e. a difficulty with lingual phonemes, rather than the underlying nature of the impairment. Moreover, since U-VBF is a visually-mediated intervention and may circumvent issues with inadequate speech perception, some studies have focused on children and young people with hearing impairment (Gick, Bacsfalvi, & Ashdown, 2003; Bacsfalvi, Bernhardt et al., 2005; Bacsfalvi, & Gick, 2007 [and related studies]) or with Down syndrome (where visual skills are thought to be a relative strength, Fawcett et al., Bacsfalvi & Bernhardt, 2008). However recent studies have moved towards focusing on children and young people with idiopathic SSDs of mixed subtypes. Normally these children have persistent or residual SSDs which are resistant to change, therefore the “Late 8” (Shriberg, 1993) consonants /s, z, l, r, ʃ, tʃ/ are unsurprisingly popular targets for English speakers (with the exception of the dental fricatives, presumably because these are not so well imaged with ultrasound as the tongue-tip tends to be in shadow and they are visible in a mirror). Within this set, rhotics are a particularly popular choice, targeted (though not necessarily exclusively) in around two thirds of studies. This is perhaps because correct production of English /r/ is socially important in North America where many of the studies are based, and also because /r/ is acquired particularly late. Additionally, it is important that because /r/ is arguably articulatorily more complex than other consonants (Gick, Bernhardt, Bacsfalvi & Wilson, 2008), this in turn would suggest that motor learning might be a particularly suitable technique. In this study we chose not to treat rhotic errors despite the children all having rhotic accents because mild clinical distortions on /r/ are not associated with the same risks to academic achievement as other types of errors and do not necessarily affect intelligibility (Shriberg, Paul, and Flipsen, 2009) (albeit they may have a negative impact on self-esteem due to the acceptability of them, see Hitchcock, Harel, and Byun, 2016). We also chose not to restrict our study to late acquired segments, instead accepting children on to the study with errors in any lingual consonant or vowel. In this way the children in our study represent a cross-section of hard-to-treat children who are eligible for Speech and Language Therapy services in the UK.

Recent studies looking at vowels and lingual consonants other than /r/ in children with SSDs have shown promising evidence that a range of errors can be remediated in school-aged children (over five years) with SSD (see for example, Cleland, Scobbie, & Wrench, 2015; Melo, Dias, Mota, & Mezzomo, 2016). In contrast, results for preschool children are mixed (Heng, McCabe, Clarke, & Preston, 2016;), perhaps suggesting that this approach is better suited to children in the school years, though more research is needed. Most studies specify one or two errors for treatment in each participant. When treating children with complex speech disorders with many errors, the problem of target selection becomes important. Cleland et al. (2015) took a pragmatic approach by prioritizing errors which affected the shape of the tongue, and then by treating the most errorful segment. It is possible too that this approach leads to the largest impact on intelligibility, though more work is needed on prioritizing (selecting and even sequencing) targets for therapy.

A concentration on the treatment of children (rather than adults) with ultrasound mirrors the clinical literature on electropalatography (EPG), the predominant articulatory VBF technique of the last 40 years, which traditionally has focused on older children for whom previous interventions have failed (Carter & Edwards, 2004). Focusing EPG towards older children with intractable problems (rather than using it for early intervention, as might be expected) is for three key reasons: the cost of individual palates; the need for stable dentition; and the notion that VBF is cognitively demanding. Clearly, for U-VBF the need for stable dentition does not apply; nor does the cost of any individual equipment. With regards to the cognitive challenges of using biofeedback, this would likely apply equally to EPG and U-VBF, though it has not been proven that a child needs a specific cognitive level to benefit from biofeedback, and indeed Cleland, Timmins, Wood, Hardcastle, and Wishart (2009) reported good outcomes in children with Down syndrome even though the participants had moderate to severe intellectual impairments. Further evidence is needed to show the efficacy for using U-VBF with a wider variety of ages and with a wider variety of SSDs

Stimulability, Acquisition, Retention and Generalization.

Gibbon and Wood (2010), discussing EPG, argue that visual biofeedback is most useful when it is used for establishing new articulations. In part this conclusion might be rather due to the selection bias of children referred for EPG therapy: normally older children who despite conventional therapy have been unable to achieve a particular speech sound for reasons outlined above. Furthermore, before visual biofeedback can be used to facilitate practice, it makes sense that it is used first to establish a new articulation. In the motor learning literature the ontogeny of complex movements (such as skilled fine motor control) is often studied by looking at an individual’s ability to imitate a novel movement (Paulus, 2014). In essence, when children begin U-VBF with a target which is not in their phonetic inventory, they begin by imitating what to them is a novel movement. The nature of this articulation is revealed in an accessible manner by watching an ultrasound movie (or live demonstration) of a typical speaker producing the gesture.

It therefore seems crucial to distinguish whether children begin therapy using VBF with either an absent or erroneous motor programme, or with one that has some correct aspects (albeit perhaps inconsistent) that need more practice. However, despite adopting a motoric therapeutic model, few studies report on whether children are stimulable for a particular speech sound before treatment begins, typically only reporting percentage target segments correct in real words (either read or imitated) pre-therapy. However, since only whole words are used it is not possible to infer that a child was completely unstimulable. A score of zero percent target correct (PTC), is not equivalent to a judgement of non-stimulability, since stimulability-testing typically incorporates imitation of speech sounds in isolation and/or in nonsense words/syllables (Powell & Miccio, 1996) rather than real words. Notable recent exceptions to this trend are Sjolie, Leece, and Preston (2016) and Preston, Leece, McNamara, and Maas (2017) which both report which of their participants were stimulable for /r/ and/or /s/ prior to intervention. In the Sjolie et al. study, out of four participants, the two who were not stimulable prior to intervention remained non-stimulable (in their stimulability probe) throughout intervention and thus in essence were non-responders to the U-VBF. Similarly, Preston et al. (2017) demonstrated that children who were stimulable prior to intervention achieved the best outcomes (2/6 children), although all children benefitted in some way from the intervention. It seems then, that in these studies there was a mixed picture of U-VBF not increasing stimulability (Sjolie et al., 2016) or only increasing it marginally (Preston, et al., 2017). However, in children who were already stimulable for a particular speech sound, U-VBF led to generalization by making the correct gesture explicit to speakers and allowing accurate practice of the appropriate motor program. This may mean these studies are primarily showing the value of U-VBF for practicing an articulation already in the child’s phonetic inventory rather than acquiring a new one.

If the literature includes more cases of children who began treatment already stimulable it does draw into question whether the children concerned would have benefitted equally well from a motor-based approach without the U-VBF. Moreover, Sjolie et al. (2016) note that the participants who were non-responders learned to accurately imitate /r/ in isolation during the intervention sessions, though there is no indication of how long it took to establish this stimulability. Preston and Leece (2017) report that in their study of intensive treatment for rhotic distortions participants were able to achieve correct productions of /r/ in sessions one to three. It was therefore a key objective of the current study to determine how quickly children become stimulable for other segments as it may be a predictor of ultimate success in intervention.

Following establishment of stimulability with U-VBF, the goal of treatment is a progression through acquisition, retention and then generalization. In Sjolie et al. (2016) acquisition is measured by comparing a treated word list at the beginning and end of the same session. Improvement within the session is said to indicate acquisition. This is somewhat of a misnomer, since it does not take into account any accurate productions by the child during the actual therapeutic intervention, only whether they are able to produce the new articulations within words at the end of the session. So it would have been more useful to record also whether children produced the target articulation in any context during the intervention session. Indeed, information on in-therapy success could be highly useful if it could be shown that it is a prognostic factor. This is a particularly important question for clinicians (rather than researchers) because in practice it may be difficult to justify persisting with a treatment when a child cannot be shown to have responded quickly to treatment. In Cleland et al.'s (2009) study using EPG with children with Down syndrome some children took up to 20 sessions to become stimulable for a new articulation. In contrast, some Speech and Language Therapy services in the United Kingdom offer only six hours of intervention (Law & Conti-Ramsden, 2000).

Retention is measured by Sjolie et al. (2016) as a child's ability to maintain progress in treated words across sessions and generalization is defined as the ability to transfer learning to new words (with for example, different vowel environments) not treated during intervention. None of these studies measure generalization outside the clinic environment, which is the ultimate goal of intervention and has been shown to be problematic in many studies with EPG (Gibbon & Paterson, 2006).

Aims

Our focus is the efficacy of U-VBF in a key client group, namely school-aged children (aged 6 to 15). Given (a) the heterogeneity of this population, (b) the phonological effects of phonetic errors in production, (c) the importance of phonetically-accurate feedback and (d) the direct embodiment of speech-production information in U-VBF; we hypothesize that the incorporation of U-VBF into therapy should deliver positive results across a range of speech sound disorder subtypes, whether or not a child's diagnosis explicitly includes a motor-speech component, for any segment in which lingual articulations are critical. A key aspect of this evaluation was to look at the stimulability of in-error segments, the establishment of new articulatory gestures, and the relation of these new gestural abilities to outcome measures based on generalization to percentage target segments correct (PTC) in untreated words more holistic. The research questions therefore were:

1. Stimulability/Acquisition: How quickly (if at all) does a non-stimulable target becoming stimulable during intervention? Do children who begin the intervention non-stimulable for the chosen target take longer to acquire the new articulation?
2. Lexical and phonotactic generalization: Does accuracy of the targeted phoneme(s) in words/pseudo-words and phrases not trained during the therapy (i.e. in untreated wordlists) improve post-intervention?
3. Functional Generalization: Does intelligibility outside of the clinic environment improve post-intervention?

Method

Participants: Twenty children with SSDs living within one health board in Scotland were invited to participate in the research. Children were recruited from community Speech and Language Therapists/Pathologists (SLTs) by writing to the SLTs and asking them to identify children on their caseload who met the following criteria: aged 6 to 15; SSDs of any subtype (typically in the UK clinicians use Dodd's 1994 classification system) with systemic errors on lingual target speech sounds; English used as a main language either at home or in school. Children with major physical disability or structural abnormality of the vocal tract were excluded, as were children with moderate to profound learning disability and/or moderate to profound hearing loss. We accepted to the study the first 20 children referred who met these criteria following referral by phone from community SLTs. Recruitment of the children was an ongoing process over the first year of the project. Of the 20 children recruited, five children were subsequently excluded following assessment: three had rising baselines which would make it hard to conclude that any improvement was due to therapy, one presented only with cluster reduction (we chose to treat systemic errors only) and one withdrew mid-way through treatment. Table 1 shows the remaining 15 children's characteristics including the SSD subtype diagnosis. Subtyping was performed by the referring clinician using their own clinical judgment or diagnostic procedures. We therefore provide the subtypes for information only. All children had had previous speech therapy but none had had any type of visual biofeedback therapy (e.g. using acoustics, ultrasound or electropalatography). None of the children except 10M received other speech therapy for the duration of the project. Therapy provided to 10M was on a different target and for less sessions that the research intervention.

Table 1: Participant demographic information including Speech Sound Disorder (SSD) subtype.

Participant	**Sex**	**Age**	**SSD Subtype**	**Co-occurring Diagnoses**	**Other Languages**
01F	F	8;8	inconsistent phonological disorder	Developmental Language Disorder	
03F	F	10;11	childhood apraxia of speech		
04M	M	7;2	phonological delay		
05M	M	6;5	phonological disorder	Developmental Language Disorder	
06M	M	6;4	phonological delay		
07M	M	8;11	childhood apraxia of speech		
08M	M	10;2	childhood apraxia of speech	Autism	
10M	M	13;4	childhood apraxia of speech		
11M	M	6;7	phonological delay		
15M	M	6;1	phonological delay		
16M	M	7;7	articulation disorder		Polish
17M	M	13;2	phonological delay	Attention Deficit Disorder/Autism	
18F	F	7;1	articulation disorder	Autism	
19M	M	10;0	phonological delay		
20M	M	9;2	articulation disorder		Basic Punjabi
MEAN	**M=12**	**8.77**			
SD		**2.37**			

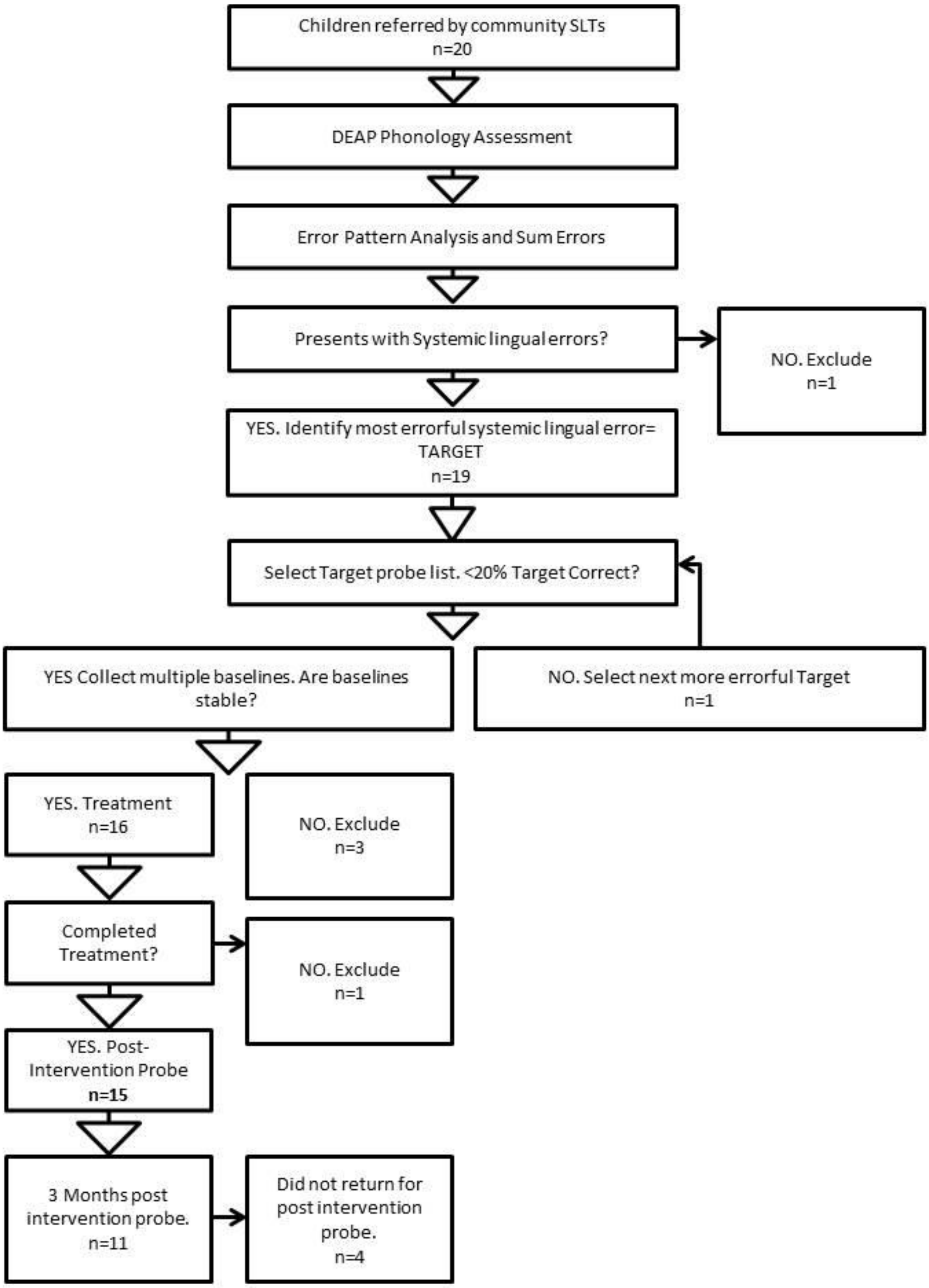
Ethical Approval: Approval was obtained from the South East Scotland Research Ethics Committee 01 (15/SS/0042). Informed consent was obtained from parents and children over 12, younger children gave assent prior to any recordings.

Design, Target Selection: A case-series with replication was employed. This allowed a degree of flexibility in a heterogeneous group and circumvented the problem of collecting many baselines and follow-up recordings which was impractical since many of the children had to travel substantial distances to the university clinic. Therapy targets were created individually for each child. This was necessary because children with SSDs are a heterogeneous group and we wished to explore a variety of child-appropriate targets. We did not treat any errors on /r/ alone because this is not typically an intervention target in the UK and was not a high priority for any of the participants, but most did have errors with this consonant. All of the children presented with more than one error, therefore we employed a target selection strategy based on Cleland et al. (2015). Firstly, each child completed the phonology subset of the Diagnostic Evaluation of Articulation and Phonology (Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002). This test (DEAP) is in common usage in UK clinics and we aimed to design a target selection method which would be useable by clinicians. From the DEAP a full pattern (phonological and phonetic error analysis) was undertaken following the manual instructions and the most errorful systemic pattern affecting a lingual segment (mild /r/ distortions excepted), was chosen for further probing.

A focused baseline probe was then collected. These wordlists (each specifically created or adapted to address the intervention target) contained around 100 words, and some sentences. The location of the target in the words was controlled to sample word position (initial, medial, final) and/or syllable position and a range of vowel contexts (or consonant contexts, for the child with vowel merger). The words exemplified a range of structural complexity, from a simple CV structure, through VC, CVC, CCVC monosyllabic words, up to multi-syllabic words. For errors which resulted in homophony in the child's phonological system (e.g. "cap" as [tap]), minimal pairs with the merged targets were included for diagnostic purposes (in this case, "tap") but not included in the calculation of PTC. Note that the items in these baseline probe wordlists were then excluded from use in intervention in order that the wordlists could also be used as mid and post-therapy probes: thus they are composed of "untreated words" and they measure both retention and lexical generalization.

From these baseline probes children were required to score less than 20% target segment correct. Where children scored over 20%, the next most errorful segment was then probed and so on until either a target was found or the child was excluded. Children who scored over 20% correct in all probes; who presented with rising baselines (across three baselines); or who did not present with errors on lingual segments were excluded. It should be noted that scoring at each point before and during therapy was performed live by the treating clinician (the third author) whereas outcomes (see below) are based on scoring by raters blind to the time point of the probe, and performed after all data had been collected. This means that there is a potential for discrepancy in the 20% threshold, and that children could score over 20% in the outcome measure analysis, especially if inter-rater reliability is low. Any children this applied to were not subsequently excluded from the analysis process, as their scores were close to 20%. Figure 1 shows the participant recruitment and target selection procedure.

Figure 1: Flowchart of study recruitment



Design, Baseline Measures: Each child underwent three baseline probes of untreated wordlists (probes as above), in weeks 1-3. Some children, for example 07M, attended for more than three assessment sessions as new probes were introduced to find a target which was <20% correct. As well as establishing the stability of the targeted speech error(s) in the untreated wordlist, these allowed an in-depth diagnostic analysis of their speech prior to intervention, see Cleland et al. (2017) for examples of errors documented during the assessment phase. The DEAP phonology subtest (Dodd et al., 2002) was completed at each baseline. To address the third research question on functional generalization, the Intelligibility in Context Scale (McLeod, Harrison & McCormack, 2012) was completed at baseline 2, post therapy and maintenance. Children received approximately once weekly sessions for ten weeks, with some gaps of longer than one week due to other commitments. Children who were not able to produce the target speech sound in multisyllabic words after 10 sessions were offered two extra sessions of intervention, because the dosage of U-VBF is unknown. We were unable to offer further sessions for practical reasons. 17M received only 9 intervention sessions due to an administrative error. Finally, there was a pair of post-therapy probes three months apart, first immediately post-therapy and again, after no contact, three months post-intervention to assess maintenance. See Figure 2 for a timeline for each participant.

Figure 2: Assessment and Intervention Schedule

Week No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	…	26
Probe	B1	B2	B3	10 Sessions of Therapy										M1		M2
								Mid								

Pre-Therapy Assessments: *Across Baselines 1-3*

The participants completed a battery of standardized speech, language, and nonverbal assessments across the three baseline sessions to establish the presence of concomitant difficulties: The British Picture Vocabulary Scales-3 (BPVS-3, Dunn, Dunn, Styles & Sewell, 2009); The Clinical Evaluation of Language Fundamentals- 4UK (CELF-4UK, Semel, Wiig & Secord, 2006) core language score; The Comprehensive Test of Phonological Processing 2nd edition (CTOPP-2; Wagner, Torgesen, Rashotte,

& Pearson, 2013) and the Robbins and Klee clinical assessment of oropharyngeal motor development in young children (RK, Robbins & Klee, 1987).

Stimulability Assessment

The articulation subtest of the DEAP (Dodd et al., 2002) was used to measure stimulability of speech sounds. Again, this test was chosen in preference to other tests used in research contexts as it is in common use in clinics in the UK. In this assessment, children first name 30 pictures containing all consonants in words with simple syllable structures (mainly CVC or CVCV though a few words contain clusters, and one item, “television”, which elicits /ʒ/, is atypically multisyllabic.) Any segments which are incorrect are then further probed to determine if children can imitate (three attempts) the segment in CV/VC and then in isolation (three attempts) if the CV/VC context is not successful.

Ultrasound Recording set-up

As in previous studies (see Cleland et al., 2015 for more details) we used an ultrasound system in which the probe is stabilized with a headset (Scobbie, Wrench & Van der Linden, 2006) to facilitate natural head movement by the client, which is particularly important both for ecological validity (given the extent of natural head movement during speech), which we hypothesize to facilitate generalization, and because of the frequency with which the client needs to look at the feedback screen then turn their head to the clinician, and back. The stability of the probe ensured by a headset within-session allows straightforward analysis of tongue shape, location and movements, both for real-time diagnostic purposes (see Cleland et al., 2017), and because it facilitates the post-hoc analysis of intervention-related changes to speech production objectively across sessions (Cleland et al., 2015).

The high-speed ultrasound data was acquired using an Ultrasonix SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced software^{TM} (Articulate Instruments Ltd, 2014) version 2.15 which internally synchronised the ultrasound and audio data (Wrench & Scobbie, 2016). The echo return data was recorded at ~121 frames per second (fps), i.e. ~8ms per frame with a 135 degree field of view (FOV) in a mid-sagittal plane. Pre-intervention ultrasound analysis for the children with t/k mergers is presented in Cleland et al. (2017) and results of the pre/post intervention ultrasound analyses will be presented elsewhere.

It is worth noting certain effects on the intervention protocol which arise from this set-up, since they result in practical differences from other studies (for example Bernhardt et al., 2005; Bacsfalvi, 2007; Sjolie et al., 2016; Preston et al., 2015). First, the biofeedback was provided using a bespoke version of AAA 2.16 (Articulate Instruments Ltd, 2016) rather than a hardware manufacturer's default ultrasound machine display. This, added to the stability of the images within session due to the use of a headset, allowed us to add accurate speaker-specific overlays of hard palate traces and to indicate target locations for articulations. By adding this useful context to the ultrasound display we think we were more accurate in matching the tactile feedback from tongue-palate contact experienced by the client to the visual feedback, and thus to capitalize on this additional aspect of articulatory feedback which is thought to be beneficial (Cleland, McCron, & Scobbie, 2013). The software also allowed us to use simple markers, such as crosses on the screen, for the participants to use to help them reach their target tongue shape and to quickly record and play-back the child's speech during therapy. However, not using a headset, (as in intervention studies by other research groups), and instead hand-holding of the probe by either the speaker or the clinician does have the benefit that it may be more comfortable for the participant. It also allows the clinician to quickly take charge of the probe to demonstrate target articulations. To circumvent this we use ultrasound videos of age-matched typically developing children producing the target articulation as standardized visual articulatory models for the children to imitate at the stimulability phase of treatment, during sessions. Recordings and therapy took place in a sound-treated studio with the SLT sitting alongside the participant (on their right side). Simultaneous acoustic and lip-camera recordings (~60fps) were also made, using an audio technica 803D clip-on microphone sampling at 22050Hz and an NTSC micro-camera synchronised to the audio. Synchronised audio and ultrasound was used for providing delayed feedback to the children (see Cleland et al., 2018) and the lip camera data was used for articulatory analysis (not reported here).

Intervention

The therapy comprised 10 (or 12) one-hour weekly mixed-content therapy sessions from week 4 to week 13 (or 15). Therapy followed a motor-based therapy approach using ultrasound visual biofeedback, similar to that in Preston et al. (2014) and Cleland, et al. (2015) and described in Cleland et al. (2018). Each participant received a single block of 10 individualised therapy sessions (12 for those who had not reached level 3 of the protocol by week 10). Sessions lasted around one-hour, with around 30 minutes spent using ultrasound and around 30 minutes doing table-top activities and discussing progress with parents/carers. Ultrasound recordings and intervention took place in a university laboratory in a sound deadened ultrasound recording room; table-top activities took place in a custom-built clinic room, also within the university laboratory. Therapy began with elicitation/stimulability of the target segment. This is arguably the hardest part of the therapy since the children enrolled in the study were not reliably stimulable for the target (see below).

The stimulability phase of therapy is described in detail in Cleland et al., 2015. To summarize, it began by demonstrating the target consonant or vowel to the child with an ultrasound video of a typically developing child producing the segment. These videos were played in both real-time and slow motion (4 times slower) to allow the treating clinician to explain the salient features of the segment to the child. For example, when demonstrating a /k/ the clinician would point out dorsal raising combined with keeping the anterior tongue low in the mouth. The clinician would then attempt to elicit (level 0, Table 2) the target consonant or vowel in CV or VC context likely to facilitate production, for example when treating /k/, a high back vowel, the GOAT vowel /o/ ([o] for Scottish English), was used to facilitate production of a velar plosive.

Once children were able to achieve an acceptable production of the target articulation a video recording was made, and this provided both video and stills of their own production which were subsequently used as a target. Synchronized audio-ultrasound recordings were also made to provide resources for more detailed phonetic analysis in the future, or used for self-scoring for level progression by the client or clinician. There was an 80% pass criterion for stepping up the levels in the protocol (table 2, adapted from Preston et al., 2013). This was measured by noting and scoring (on play-back immediately after the recording by the treating clinician) three words/nonwords containing the target in combination with corner vowels or consonants 10 times at the beginning of each therapy session. If participants scored 8/10 or more correctly in all three words/nonwords, they moved on to the next level. Children were allowed to move through more than one level in any given session, the highest level achieved was therefore noted to quantify progress. At each level the clinician used a range of treated words (see Cleland et al., 2018 for examples) to drill production of around 100 attempts at words or phrases (depending on level) containing the target segment. Games depending on the child's preferences such as dice throws, sticker charts, etc were used to encourage high numbers of repetitions. Feedback on attempts was given to participants in the form of KP (for example, "good, you moved the back of your tongue towards the cross") and later KR (for example, "that was a good /k/"). Correct attempts of the target in any context at any time during the intervention were also noted by the ultrasound clinician to determine the session in which there was establishment of stimulability of the target.

Table 2: Intervention phonotactic hierarchy

Step up with 80% pass criterion	
Level 0	CV or VC facilitative vowel, non-words
Level 1	CV
Level 1	VC
Level 2	CVC WI
Level 2	CVC WF
Level 3	Multisyllables
Level 4	Phrase repetition WI
Level 4	Phrase repetition WF
Level 5	Cloze (sentence completion)
Level 6	Clusters
Level 7	Complex sentences repetition and invention

Note. C=Consonant; V=Vowel; WI= target segment in word initial position; WF=target segment in word final position. Adapted from Preston et al. (2013).

Data Analysis: Our key outcome measure was PTC correct in untreated probes. All probe wordlists were transcribed using symbols of the IPA and ExtIPA independently by two certified SLTs blind to the probe time-point. Time points, but not individual items were randomised, that is, the rater listened to the recording for one randomly selected session before listening to another. The raters were not involved in the research project and were generalist-SLTs with recent experience of transcribing disordered speech. Ratings were then scored for % segment on target. For example, if velars were targeted, the number of correct (i.e. phonetically accurate velar stops) were counted and expressed as a percentage, then displayed graphically for each participant. Reliability of these correct/incorrect judgements on a point-by-point basis was calculated between the two raters on all of the data using Cohen's Kappa and was Kappa = .572 (p <.0005), 95% CI (.548, .596) which is "moderate". While previous studies have reported higher reliability, these have often included children only with minor distortions on sibilants or rhotics. Our children represented a range of complex and severe speech disorders which are known to be vulnerable to transcription difficulties. Standardised effect sizes, d_2

were also calculated. We present also the session number in which children were first stimulable for their new articulation.

To determine whether any intervention effect led to improved intelligibility outside of the clinic context, at baseline 2, post-therapy and maintenance we asked the children's parent/carer to complete the "Intelligibility in Context Scale". This questionnaire is a rating of how well understood children are to both familiar and unfamiliar listeners and was scored according to published instructions.

Procedural Fidelity

All intervention sessions were conducted by a certified SLT (the third author) trained in the use of ultrasound biofeedback by the first author. Initial sessions of intervention were supervised by the first author. Sessions took place in the same clinic room each time. Fidelity of the ultrasound image was ensured by using a headset to stabilise the probe and ensuring that the mandible and hyoid shadows were visible in all recordings. To determine whether the SLT adhered to the step-up criteria one randomly selected session per child was reviewed by the first author to determine agreement on level achieved in each session (see table 2). The rater reviewed productions at the beginning of every session and judged the accuracy of the 10 productions and then assigned it a level. For example, if the recording of the child was of ten productions of CVC with the target segment in WI position, the rater calculated the PTC and assigned "level 2" as achieved if the score was ≥80%. Agreement was 87%, disagreements varied by one level only and always in favour of a higher level, i.e. the clinician moved to the next level too quickly.

Results

Results: Baseline

Language and Cognitive Measures

The speech, language, and cognitive profile of the participants were in line with a diagnosis of primary SSD. Table 3 shows the results of the Raven's Matrices (RM, non-verbal ability), British Picture Vocabulary Scale 3 (BPVS, receptive vocabulary), Clinical Evaluation of Language Fundamentals 4UK (CELF, receptive and expressive language) and the Children's Test of Phonological Processing. Results are presented as standard scores. While only one child (10M) presented with nonverbal ability outwith the normal range, five children presented with measurable language impairment, consistent with a diagnosis of developmental language disorder.

Table 3: Participant results from standardized assessments.

CHILD	RM	BPVS	CELF	CTOPP
01F	100	80	69	78
03F	81	77	50	58
04M	100	108	100	97
05M	119	106	73	94
06M	110	97	120	118
07M	100	75	67	74
08M	81	104	99	69
10M	75	100	87	79
11M	124	101	111	106
15M	124	96	87	91
16M	110	81	81	101
17M	90	79	73	83
18F	119	93	84	90
19M	110	93	93	82
20M	100	75	82	90
MEAN	102.87	91.00	85.07	87.33
STDEV	15.81	11.97	17.95	15.08

Note. RM= Raven's Progressive Matrices; BPVS= British Picture Vocabulary Scale 3; CELF= Clinical Evaluation of Language Fundamentals, Fourth UK Edition; CTOPP= Comprehensive Test of Phonological Processing, second edition. All scores are standard scores.

Oral Structure and Function

Table 4 shows the Robbins-Klee structure and function raw scores. For structure, Robbins and Klee (1997) report a range of 18-24 for children aged 6;0 to 6;11 (the upper age-bound of the sample). 01F and 07M both scored lower than this, 01F due to missing teeth/malocclusion, a high narrow palate and a deviated uvula. 07M due to missing and misaligned teeth. 04M was noted to have a bifid uvula but further investigation did not indicate a submucous cleft palate. For function, Robbins and Klee (1997) report a range of 108-112 for age 6;0 to 6;11. Despite being older than this our children all scored below 107. This is not surprising since many of the items require accurate imitation of articulatory gestures or words.

Table 4: Test of Oromotor Function Results

CHILD	STRUCTURE (RAW)	FUNCTION (RAW)
01F	16	78
03F	22	84
04M	22	97
05M	20	105
06M	19	106
07M	17	88
08M	24	67
10M	19	92
11M	21	89
15M	20	88
16M	22	97
17M	16	93
18F	20	87
19M	23	93
20M	20	81
MEAN	20.13	91.00
STDEV	2.33	11.06

Selecting Therapy Targets: Error Analysis

The DEAP phonology subtest facilitated the identification of 18 definable processes. After eliminating the structural processes, and processes not related to changes in place of articulation, nine error types characterised the participants: gliding, velar fronting, post-alveolar fronting, alveolar fronting, backing to velar, backing to pharyngeal/glottal, vowel merger, debuccalisation (sound preference for [h]) and lateralisation. The process with the highest number of errors, excluding gliding, was then probed further to ensure that the child had <20% segments correct at baseline. This was true for 14/15 children. One child, 07M scored >20% correct for the initial target choice of velars, further probing of post-alveolar fricatives was then undertaken and found to be <20% correct, and hence chosen for intervention.

In summary, six children were treated for velar fronting; three for post-alveolar fronting; two for the unusual pattern of backing to pharyngeal or glottal; one for the similar process of debuccalisation (production of all syllable onsets as [h]); one for vowel merger; and two for lateralised sibilants. This is a larger range of error types than reported in previous studies and comprises both typical processes and unusual ones. Table 5 shows the analysis of the DEAP errors and which segment was chosen for therapy, denoted by shading.

Table 5: Error pattern analysis from the Diagnostic Evaluation of Articulation and Phonology, Phonology Subtest.

ERROR PROCESS		CHILD														
		01M	**03F**	**04M**	**05M**	**06M**	**07M**	**08M**	**10M**	**11M**	**15M**	**16M**	**17M**	**18F**	**19M**	**20M**
Lingual shape errors	Velar Fronting	20*	14*	20*	3	2	5	1	3	20*	19*		17*		6	
	Post alveolar fronting	5	1	6	5	7*	3*	5		7	4				7*	
	Alveolar fronting											2				
	Backing to velar		1		1		2	9		1	1	2		7		
	Backing (to pharyngeal/glottal)								22*					43*		
	Lateralisation											22*				32*
	Debuccalisation in SI							52*								
	Vowel Merger				17*											
	Gliding	8	7	14	12	2	17		14	8	18	3		1	10	13
Non-lingual shape errors	Stopping	1	5							2	2	2				2
	Deaffrication	1	3		1	2	3	2	2		2	1	2		1	
	Voicing errors		4		1		11	5		6	1				3	
	Weak syllable deletion						2				1					
	FCD		7		1		2			1	1	1		1		
	ICD							2								
	MCD	1	2		4		1	2		1	2				1	
	Cluster Reduction	6	11		3	1	18	2	4	20	2	3	1	1	11	3
	Assimilation														1	
Unclassified	Other		6		6		13	8	7	3		5	1	1	5	1
	TOTAL ERRORS	42	61	40	54	14	77	88	52	69	53	41	21	54	45	51
	% Consonants Correct	70	57	72	73	89	44	38	64	51	64	69	83	62	68	64

Note. *denotes the treated segment/process.

Stimulability of Chosen Targets

We were interested in whether children who began the intervention non-stimulable for the chosen target took longer to acquire the new articulation. Table 6 shows the results of the DEAP articulation stimulability assessment. Importantly, stimulability was confirmed with ultrasound, to check for covert errors on attempts perceived to be correct. One child, 01F was perceived to produce 1 out of 3 attempts at [ki] correctly but visual inspection of the ultrasound revealed a retroflex articulation on the attempt transcribed as [ki], a more accurate transcription would therefore be [t̠]. Twelve out of 15 children were not able to imitate the target articulation, or its voiced cognate (where applicable) in either a CV or VC context, or in isolation, at baseline. Three children 04M, 05M, 08M were, however, able to imitate the target in isolation. Therefore, none of the children were reliably (in context) stimulable prior to intervention.

Table 6: DEAP Articulation stimulability results. Numbers are the number of correct attempts at the target out of three, achieved in either CV or VC context, and in isolation. Numbers marked * show correct attempts.

CHILD	ERROR PROCESS	DEAP ARTICULATION STIMULABILITY			
		Target	CV/VC	Isolation	Other non-stimulable segments
01F	velar fronting	k	0	0	tʃ, dʒ
03F	velar fronting	k	0	0	ʃ,ʒ,tʃ
04M	velar fronting	k	0	3*	
05M	vowel merger	ɛ	0	3*	ʃ,tʃ
06M	post-alveolar fronting	ʃ	0	0	tʃ,dʒ,ʒ
07M	post-alveolar fronting	ʃ	0	0	ŋ,θ,dʒ,ʒ,l
08M	debuccalisation	s	0	3*	b,d,g,ŋ, ʃ,ʒ, tʃ, dʒ, l,ɹ,w
10M	pharyngeal /s/	s	0	0	ŋ,z
11M	velar fronting	k	0	0	θ,ð, ʃ,ʒ, tʃ, dʒ,ɹ
15M	velar fronting	k	0	0	l
16M	lateral /s/	s	0	0	tʃ, dʒ
17M	velar fronting	k	0	0	
18F	pharyngeal alveolars	**s**	0	0	t,d,n, ʃ,ʒ,tʃ,dʒ
19M	post-alveolar fronting	ʃ	0	0	tʃ,dʒ,ʒ,ɹ
20M	lateral /s/	s	0	0	θ,ð, ʃ,ʒ, tʃ, dʒ,ɹ

Intervention- Stimulability and Acquisition

Since none of the children were readily stimulable for the target in a CV or VC context (i.e. were unable to imitate it all 3 attempts), intervention began with elicitation (level 0). Table 7 shows the session number in which each child first achieved the target articulation at least once during the 10 recorded attempts at the beginning of the session (judged from audio and ultrasound on immediate play-back), and the phonotactic context in which this was achieved. Most children (10/15) achieved the new articulation quickly, in the first or second session. The three children who had already been stimulable in isolation (04M, 05M, 08M in Table 6) all achieved the target articulation in the first session. Four children took until a much later session (6th to 9th) to achieve the new articulation, and one never did (20M, who was treated for lateral fricatives). Of those children who took longer to acquire the new sound, two were receiving intervention for velar fronting (and were also two of our younger participants, both aged six) and three for disordered sibilants.

Table 7: Session number in which each participant was first stimulable

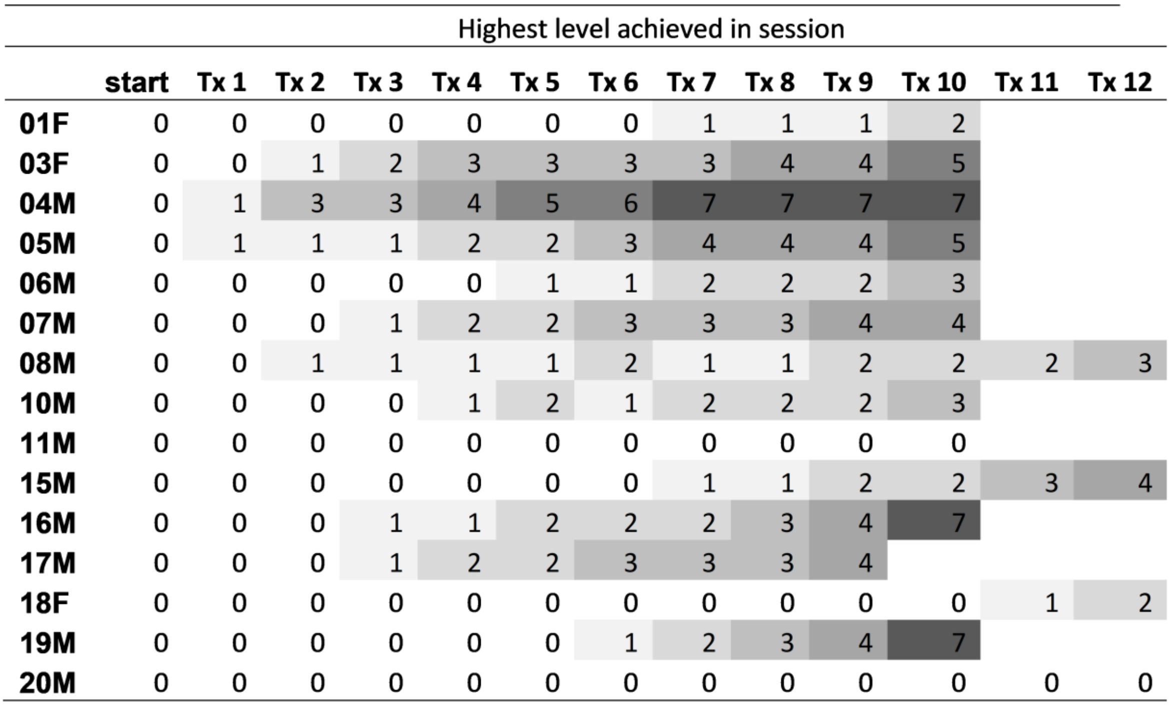
Participant	Target	Session	Realisation
01F	/k/	2	[xk]
03F	/k/	1	[ok]
04M	/k/	1	[ok]
05M	/ɛ/	1	[ɛ]
06M	/ʃ/	2	[iʃ]
07M	/ʃ/	1	[iʃ]
08M	/s/	1	[si]
10M	/s/	1	[ts]
11M	/k/	7	[xk]
15M	/k/	7	[ŋk]
16M	/s/	1	[ts:]
17M	/k/	2	[ko]
18F	/s/	9	[s:]
19M	/ʃ/	6	[ʃi]
20M	/s/	Never	
MEDIAN		2	
MODE		1	

Note. Target is the goal segment the clinician was attempting to elicit from the child. The realization column is a broad phonetic transcription of the child's first production containing the target segment

While table 7 shows only the first session of acquisition, figure 3 shows how quickly the children moved through the protocol, as they achieved and retained their new articulation in more complex contexts. One of the children, 11M, was able to achieve the target articulation at least once (Table 7) but was not able to progress beyond level 0, as he was unable to achieve [k] consistently at 80% correct (the step-up criteria). Figure 3 shows that 10/15 children stayed at level 4 (phrase repetition) or below, whereas 5/15 children progressed to level 5 (cloze sentences) or better. Of the three children who began treatment stimulable in isolation, two made rapid progress (04M and 05M, treated for velars and vowel merger respectively), but 08M who was treated for onset plosives made slower progress.

This particular child had an autism diagnosis, suspected CAS and a particularly severe SSD with only 38% consonants correct in the DEAP at baseline, it is therefore not surprising that he made slow progress integrating a new articulation.

Figure 3: Highest level achieved in therapy protocol in each session. Darker shading denotes more advanced levels where 0(white cells)= unable to achieve the target articulation on at least 8/10 attempts and 7=able to achieve the target articulation at sentence level in at least 8/10 attempts.



	Highest level achieved in session												
	start	**Tx 1**	**Tx 2**	**Tx 3**	**Tx 4**	**Tx 5**	**Tx 6**	**Tx 7**	**Tx 8**	**Tx 9**	**Tx 10**	**Tx 11**	**Tx 12**
01F	0	0	0	0	0	0	0	1	1	1	2		
03F	0	0	1	2	3	3	3	3	4	4	5		
04M	0	1	3	3	4	5	6	7	7	7	7		
05M	0	1	1	1	2	2	3	4	4	4	5		
06M	0	0	0	0	0	1	1	2	2	2	3		
07M	0	0	0	1	2	2	3	3	3	4	4		
08M	0	0	1	1	1	1	2	1	1	2	2	2	3
10M	0	0	0	0	1	2	1	2	2	2	3		
11M	0	0	0	0	0	0	0	0	0	0	0		
15M	0	0	0	0	0	0	0	1	1	2	2	3	4
16M	0	0	0	1	1	2	2	2	3	4	7		
17M	0	0	0	1	2	2	3	3	3	4			
18F	0	0	0	0	0	0	0	0	0	0	0	1	2
19M	0	0	0	0	0	0	1	2	3	4	7		
20M	0	0	0	0	0	0	0	0	0	0	0	0	0

Lexical-phonotactic Generalization: Untreated Probes.

Figure 4 shows individual results for each child in the untreated wordlists designed to probe their specific error and used also for baselines. Comments are then made by grouping children into those with effect sizes: no-effect, small, medium or large. Effect sizes (d_2) were calculated using standard mean difference which is the difference between pre and post intervention means divided by the pooled standard deviation (see Gierut, Morrisette, & Dickinson, 2015 for further details). The effect size thresholds were designated as small (>1.4), medium (>3.6) or large (>10.1) in line the benchmarks reported by Gierut et al. (2015) for children with “functional phonological disorders”. Although we do not use that terminology here, and some of our children may have motor-based disorders, the inclusion of this categorization serves the purposes of giving some guidance for clinicians as to the clinical effectiveness of U-VBF. Three Children (01F, 07M, 20M) failed to attend for maintenance probe three months after intervention and there are therefore missing data-points.

No Generalization Effect

Five children showed no effect of U-VBF on generalization into words, as evidenced by the scores in the untreated probes. None of these children were stimulable prior to intervention. Two, 01F and 11M, were treated for velar fronting; both achieved a velar in intervention but 11M was unable to produce it in even CV sequences and 01F did not progress beyond production in CVC. 11M had no concomitant diagnoses which might explain slow progress but was observed to have attentional difficulties which may have precluded his engagement with the intervention. 01F had a concomitant language disorder which may have impacted on her progress. She also presented with many covert errors suggestive of a motor-speech disorder and is reported in detail in Cleland et al. (2017). Children who backed alveolars to pharyngeal articulations, 10M and 18F, failed to generalize despite showing signs of improvement. They had changed articulations from pharyngeal to alveolar, with 10M realizing /s/ as [ʃ] post-intervention and 18F realizing it as [ɬ]. Both were able to produce [s] in limited contexts during intervention. Arguably in both cases this represents an improvement in the acceptability of their speech, as their post-intervention errors are more common and less severe misarticulations than the pre-intervention ones. This speaks to a need for future studies to move away from right/wrong judgments and to adopt a more gradient approach to correctness. It also shows an effect of intervention, albeit not in our designated outcome measure. 20M is the only child who failed to acquire or generalize a new articulation. He presented with particularly severe lateralized sibilants along with inconsistent lateral release of alveolar plosives, and his failure to progress may be indicative of the lack of highly relevant visual biofeedback information given the mid-sagittal section used during most of the intervention.

Small Generalization Effect

17M showed a small effect of intervention on untreated probes for velars despite not being stimulable for /k/ prior to intervention. At maintenance, his % velars correct had increased from near 0% to 17% correct (and he was working at level 4). It is possible that his diagnosis of autism, along with deeply entrenched incorrect articulations (he was aged 13 yet presenting with velar fronting) made it difficult to achieve the initial new articulation. Attention to the ultrasound display was also hampered by an attention deficit diagnosis. While the small increase in PTC is likely to not be clinically significant, the degree of improvement by session 9 suggests that there was potential for further improvement with further intervention.

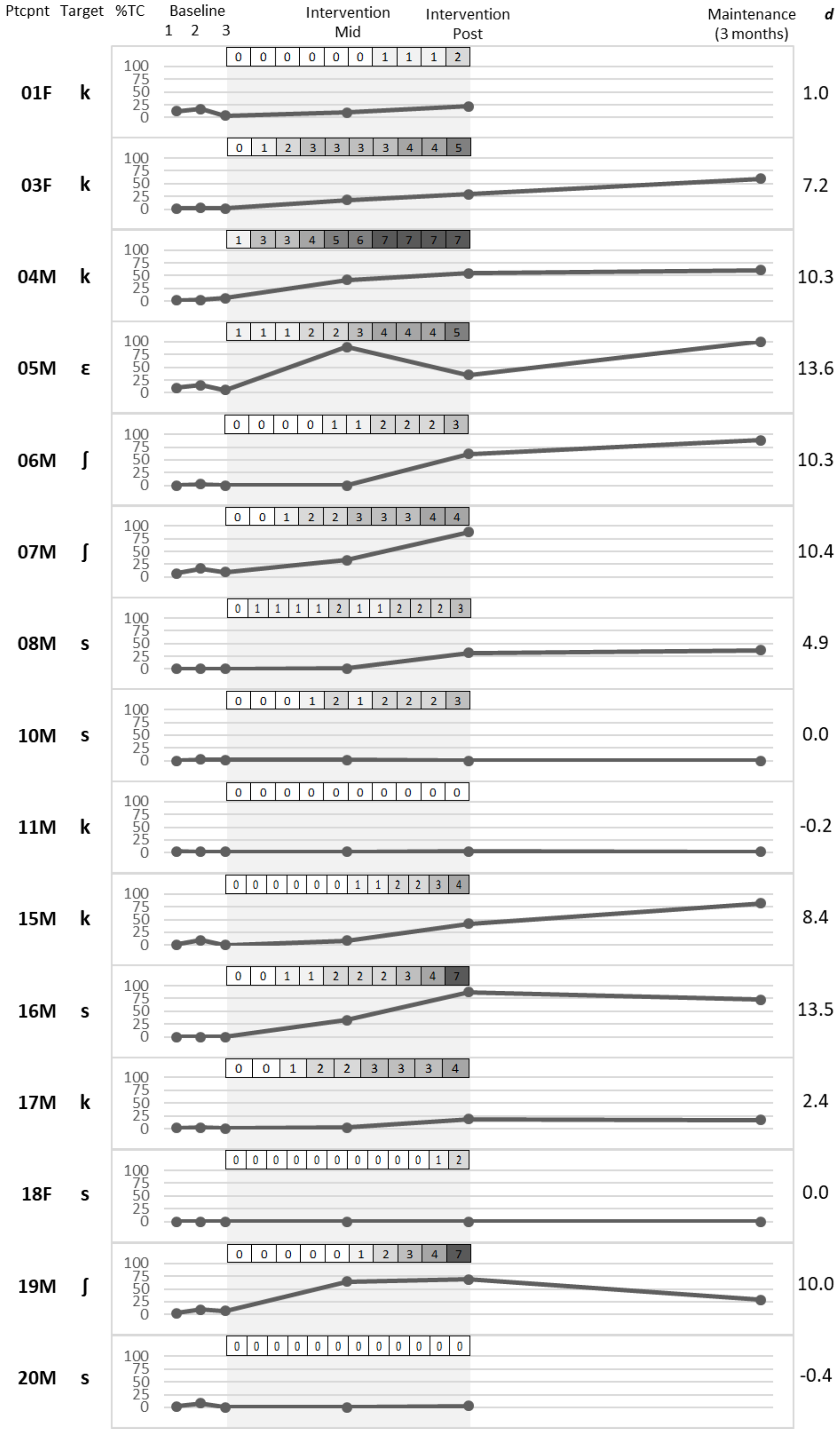
Medium Generalization Effect

Four children showed a medium effect. 03F and 15M were treated for velar fronting; 19M for post-alveolar fronting and 08M for the unusual pattern of replacement of onsets with [h]. Given the nature of his errors it should be noted that 08M PTC includes all onset obstruents although intervention began with just [s]. Of these four children, only 08M was stimulable prior to intervention.

Large Generalization Intervention Effect

Five children showed a large effect of the intervention. 04F was treated for velar fronting; 06M and 07M for post-alveolar fronting; 05M for vowel merger; and 16M for lateralized sibilants. Of these five children, only two, 04M and 05M were stimulable prior to intervention. It is worth noting that with the exception of 05M who had a mild delay in language (standard score of 73 in the CELF-4UK) none of these children presented with any co-occurring diagnoses, it is therefore probable that children without additional difficulties will respond better to U-VBF and in our study this was potentially a better predictor of success than initial stimulability.

Figure 4: Untreated probe data for all children. Shaded area indicates intervention period. B1,2,3- Pre-treatment; Mid= mid treatment; Post= immediately post treatment; Maintenance: 3 months post treatment. *d*= Standardized effect size. Shaded chart across intervention period are duplicated from table 8 for ease.



Functional Generalization outside of the clinic environment:

The Intelligibility in Context Scale (McLeod, Harrison, & McCormack, 2012) was used as a global rating of how well children were understood in their daily lives by listeners ranging from parents, through friends to strangers. Children were rated on a 5 point scale from “always” (5 points) to “never” (1 point) intelligible to seven different types of listeners. Results were then averaged to give a score out of 5. Table 9 shows the ICS results. While some children show no improvement in ICS scores (or small negative results) as a group there was a significant increase in scores from baseline 1 to maintenance (paired samples t-test t(11)=-3.185; *p*=0.009). This represents a change in the average score from 3 (sometimes understood) to 4 (usually understood).

Table 8: Intelligibility in Context Scale results and improvements in raw score.

	BL1	**Post**	**Maint**	**Improvement Post-BL1**	**Improvement Main-BL2**
01F	3.6	3.3		-0.3	
03F	3.1	2.9	3.4	-0.3	0.3
04M	4	5	4	1	0
05M	3.6	4	4.1	0.4	0.6
06M	3.7	4.9	4.9	1.1	1.1
07M	2.6	3.4		0.9	-2.6
08M	3.4	3.3	3	-0.1	-0.4
10M	3.4	3.6	3.6	0.1	0.1
11M	3.1	3.4	3.4	0.3	0.3
15M	3.1	4	4.6	0.9	1.4
16M	3.6	3.7	4.6	0.1	1
17M	3.4	3.3	3.1	-0.1	-0.3
18F	2.7	3.4	3.4	0.7	0.7
19M	3.1	3.9	3.9	0.7	0.7
20M	2.3	2.7		0.4	
MEAN	3.25	3.65	3.83	0.39	0.22
SD	0.46	0.64	0.62	0.48	1.00

Discussion

This study evaluated the effectiveness of U-VBF as a treatment for stimulating and establishing new articulations in school aged children with a variety of primary speech sound disorders of different subtypes affecting different lingual phonemes. In general, our results are in line with previous studies which show a mixed pattern of some children responding well to the intervention, and generalizing to untreated words (e.g Bernhardt et al., 2003; Cleland et al., 2015), and other children either not responding at all or showing limited generalization (e.g. Bressman et al., 2016; Hitchcock & McAllister, 2014; Preston et al., 2017). This study adds information about some more unusual and arguably more severe errors, such as backing of sibilants to a pharyngeal place of articulation and debucccalisation of onsets.

Despite U-VBF being a motor-based treatment we chose to treat children with mixed SSD subtypes. Firstly, we need to consider the professional context in the UK, where most clinicians are diagnosing SSD according to Dodd's classification system (Dodd, 1994) which leads overwhelmingly to children being classified as having "phonological" impairments (87.5% in an analysis of caseload by Broomfield & Dodd, 2004) and to a lesser extent (12.5% of caseload) an articulation disorder. This practice means it is impractical to find children who have prior diagnoses of CAS in the UK (less than 1% in this UK-based study). Secondly, there is a good theoretical reason for our approach. Previous instrumental studies (for example Cleland et al., 2017) show that children with diagnoses of phonological impairments may, in fact, have subtle motor problems evident in covert errors in speech production. U-VBF, as a motor-based approach, is therefore an appropriate intervention and success with this intervention provides support for the view that the root cause of the disorder is at least partially motoric.

In the current study we focused on a typical profile of children in the UK prioritized for speech therapy services, that is those with multiple (persistent) error types. We assume these children display a greater level of unintelligibility than children in previous studies with residual speech sound errors characterized by only rhotic distortions. This shows the potential applicability of U-VBF to a wider range of more severe disorders, and to a wider range of languages. Likewise, we did not excluded children with language impairment and autism. It is possible that these co-occurring diagnoses may have impacted success in treatment; the three children with autism (08M, 17M, 18F) made only small improvements. 08M and 18F are particularly interesting cases as both presented with very unusual errors: deletion of onsets and pharyngeal productions of /s/ respectively. While previous studies have shown success with U-VBF in establishing production of /s/ (for example, Lipetz & Bernhardt (2013); Preston et al., 2014) these studies have treated more common distortions, such as lateralization, this study shows the potential for changing more unusual productions.

Children without co-occurring diagnoses showed larger effect sizes in lexical generalization. We conclude from this that rather than excluding children with concomitant disorders from U-VBF that they need a higher dosage of intervention, evidenced by the fact small improvements were made in this group of children. Previous studies of EPG and ultrasound with children with Down syndrome (Cleland et al., 2009; Fawcett et al., 2008) have shown improvements, although both studies offered participants more sessions of intervention (24 and 16 respectively) than the current study. While our children did not have intellectual impairments, language impairments and autism arguably require similarly increased intervention dosage.

In line with previous articulatory feedback studies we found that some children (5/15) did not respond to the treatment in terms of lexical generalization to untreated words, although all children bar one made some progress towards achieving the target. Gibbon and Wood (2010) suggest that real-time visual biofeedback of articulation (in their paper on EPG) is most useful for establishing *new* articulations and indeed our work and others (e.g. Preston & Leece, 2017; Bacsfalvi, 2010) show that U-VBF is also useful for achieving stimulability of new articulations. However, integration into words and/or generalization to untreated words is more problematic, because of the variation and complexity in phonetic articulation and planning, phonological generalizability and the likely effects of exemplar frequency. Previous work by Sjolie et al. (2016) and Preston et al. (2017) has shown that children who are not stimulable prior to intervention may show the worst outcomes in terms of generalization. While we showed some evidence that lack of stimulability is a problem (all five children who failed to generalize were not stimulable before intervention) we also showed that many children who began intervention unable to achieve the target articulation in CV or VC context were able to make progress to varying levels. In fact, *none* of the children in this study showed true stimulability prior to intervention (Miccio et al., 1999), with no child able to imitate the target articulation in CV or VC contexts.

Another important issue is the discrepancy between low stimulability scores and, for some children, scores above zero in words at baseline. How can the target be correct in a number of words but not be stimulable? One possibility is that words scored as correct were, in fact, misperceived by the transcriber for holistic reasons. Alternatively, if the transcription does however indicate an in-situ production of the target that would imply an ability on the child's part to produce a stand-alone correct-sounding version of the target, it may be that a more in-depth or ecologically valid stimulability assessment is required. A detailed articulatory analysis could address the former hypothesis, by revealing and explaining aspects of articulation and co-articulation from the high-speed audio-ultrasound data collected, in relation to lexical factors (neighbourhood density, minimal pair confusability and lexical frequency), but is outwith the scope of the current paper. However, we did find that one child, 01F who was transcribed as producing one attempt at [ki] correctly in the stimulability assessment, did in fact produce this sequence as [t̠i]. This highlights the diagnostic power of ultrasound tongue imaging since it can provide objective articulatory evidence of speech production and help clarify the relationship between these lexical and non-lexical production tasks and the underlying abilities.

Another important issues is that some children who were not stimulable, and scored at zero across baselines, were in fact able to achieve new articulations in therapy. This is in line with most previous studies who reported some progress even in children who are unable to generalize (e.g. Preston et al., 2016; Sjolie et al., 2016 Bacsfalvi 2010). U-VBF is therefore a good way of acquiring (if not rapidly learning and generalizing) new articulations. In fact, most of the children (10/15) achieved the new articulation very fast, in the first or second session of U-VBF. Future studies should focus on stabilizing and generalizing new productions, perhaps by increasing dosage. Again, it seems that the relationship of stimulability to lexical-phonotactic generalization is a complex one. Lack of success in the first or second session of intervention does not indicate that the treatment will not be successful. Four children only achieved the new articulation on or after the sixth of 10 intervention sessions, which means that we do not have enough participants, for a statistical analysis, or enough remaining sessions for us to be certain about the level of progress up the articulatory levels or in lexical generalization which might have been achieved by these late starters had there been more time. Visual inspection of the data reveals, however, that once children were able to produce their new articulation in CVC contexts, progression tended to be rapid and to generalize to untreated words. This is in contrast to Preston et al. (2016) where two children with CAS were able to achieve /r/ in words (and to a limited extent in phrases) during treatment sessions, but did not generalize to untreated words. Preston et al. interpret this finding as an indication of limited acquisition of the target and suggest that more practice is necessary. While we agree with this view, it is also possible that we saw rapid generalization following acquisition because we treated articulations that were less articulatory complex and/or perceptually more salient and/or functionally more important in a phonological sense. Many of our children presented with errors that led either to homophony (velar fronting) or to high levels of unintelligibility. Some of the unintelligibility arose because of the pervasive phonological neutralization effects of common errors like velar fronting, but other cases were due to the phonetic severity of a more unusual error (the use of pharyngeal fricatives). In these latter cases, phonetic reorganization which impacts heavily on intelligibility due to elimination of a phonological merger is likely to lead to rapid generalization if the child has mastered the motor program sufficiently to allow significant practice, perhaps outside the clinic environment. Such self-initiated practice may be motivated by either the conversational breakdown that likely occurs with such severe errors or by a desire to sound more like peers, i.e. the speech attunement framework (Shriberg, 1994) . A complementary explanation is that persistent speech errors, like rhotic and sibilant distortions, are in general more difficult to treat because they have a "*minimal impact on intelligibility,* *(and) might have been subject to longer periods of positive reinforcement and could thus represent more strongly habituated patterns*" (Flipsen, 2015, p218). Such accidental positive reinforcement is unlikely to be the case with errors which result in merging of phonemes or which are highly unusual to the point they lead to communication breakdown.

Preston et al. (2016) also points out that lack of generalization may be due to selection of inappropriate candidates: children who are complex cases with comorbidities which describes six of our participants. However, three of our children had diagnoses of autism yet did respond to intervention. While the five children who did not respond to intervention had moderate to severe speech disorders, one child with a large intervention effect had a severe speech disorder. On the other hand, one child with a mild speech disorder had only a small intervention effect (17M). In this particular child's case the clinician observed qualitatively throughout the intervention that the child's motivation and attention were low, and suggested as causal factors explaining the lack of success.

Limitations.

The children in this study were a representative sample of UK children with hard-to-treat SSDs who referring clinicians thought might benefit from U-VBF. Recruiting children in this way therefore led to a heterogeneous group of different subtypes of SSDs, different co-morbidities, and different treated segments. While this does provide interesting information about a range of children, it does make it difficult to determine (with only 15 participants) whether different subtypes or segments are more amenable to U-VBF. Clearly multiple larger studies with tightly defined groups would be useful, though in terms of the children with autism it would be difficult to construct a trial where children had this diagnosis and all required treatment for the same target segment.

Due to the heterogeneity of the participants, a case series design was utilised. However, for practical reasons (scheduling children who had to travel long distances and often had to be taken out of school) it was not possible to provide more or staggered baselines. This limitation makes it difficult to conclude that any post-intervention improvement is not due simply to maturation. Likewise, a follow-up recording of the children six months to a year post intervention would have been useful (see Cleland et al., 2015) but was prohibited by time constraints. It is also likely that at least some of the improvement may be due to employing an articulatory approach rather than to the ultrasound per se. A recent study by Furniss and Wenger (2018) showed no between group difference in post-intervention outcomes for children treated with articulation therapy versus those treated with U-VBF. However, it should be noted that their study included only participants with rhotic and sibilant distortions and that the ultrasound group showed some evidence of a quicker intervention response.

In terms of measuring progress during the intervention, while the untreated probe-lists were very useful, further measuring of transfer to conversational speech, outside the clinic environment and the impact of the intervention on the children's quality of life would have been more ecologically valid. While the Intelligibility in Context Scale provides some information, parental report may have been biased by the desire to see a positive intervention effect. Moreover, more objective measures of change in tongue shape and position (see Cleland et al., 2015) would complement phonetic transcription results and have the potential to be automated in the future (see Cleland et al., 2018).

Conclusions.

In line with previous studies using ultrasound we found U-VBF to be particularly useful for establishing new articulations, with most children becoming stimulable for a new sound in the first or second session. Lexical generalization is a more complex process, and remains harder and is slower, or cannot be demonstrated in the timescales investigated, perhaps due to individual factors not controlled for in the study (like motivation) but more likely due to inadequate dosage. Future studies should consider novel methods of measuring attention during sessions and consider the impact of client motivation on ability to change behaviour (Michie, van Stralen, & West, 2011). Children with concomitant disorders were slower to progress in intervention and therefore in particular may need greater dosage and intensity of intervention. There was some evidence to suggest that children with severe phonetic distortions (in this case, two children with pharyngeal productions of /s/ and one child with lateral productions of /s/ and, crucially, other obstruents) responded less well to intervention, perhaps due to differences in underlying aetiology of these disorders. It may be that non-neutralising distortions present differently to plain mergers (such as velar fronting) and distorting mergers (such as pharyngealisation of multiple targets) and that the additional phonological ramifications of merger help the children to be more responsive to the sorts of intervention described here. In sum, ultrasound visual biofeedback shows promise as a technique that can contribute positively to the treatment of a wider range of lingual errors in children with various subtypes of speech sound disorders and varying levels of severity.

Acknowledgments

Thank you to the children and parents who gave up their time to travel to the university to take part in the project. Thank you to the referring clinicians who identified the participants for us. Thank you to Steve Cowen for technical assistance and to Louise McKeever, Aisling Egan, Anna Womack, and Caitlin Adair for help with the transcriptions.

References

Articulate Instruments Ltd 2014. *Articulate assistant user guide: Version 1.18*. Edinburgh, UK: Articulate Instruments Ltd.

Bacsfalvi, P. (2010). Attaining the lingual components of/r/with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech-Language Pathology & Audiology*, *34*(3).

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.

Bernhardt, B., Bacsfalvi, P., Gick, B., Radanov, B., & Williams, R. (2005). Exploring the Use of Electropalatography and Ultrasound in Speech Habilitation Explorer l'électropalatographie et l'échographie pour l'éducation de la parole. *Revue D'orthophonie et D'audiologie*, *29*(4), 169.

Bernhardt, B. M., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics, 17*(3), 199–216.

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.

Broomfield, J., & Dodd, B. (2004). The nature of referred subtypes of primary speech disability. *Child Language Teaching and Therapy*, *20*(2), 135-151.

Broomfield, J. & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomized control trial. *International Journal of Language & Communication Disorders, 46*, 628–640

Carter, P., & Edwards, S. (2004). EPG therapy for children with long-standing speech disorders: predictions and outcomes. *Clinical linguistics & phonetics*, *18*(6-8), 359-372.

Cleland, J., & Isles, J. (17, July 1). Evidence of Efficacy: Ultrasound Speech Therapy Studies. Retrieved July 28, 2017, from http://www.articulateinstruments.com/US_Treat_Refs_0717.pdf

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 & 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 and Phonetics*, *29*(8–10).

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 & phonetics*, *23*(12), 926-939.

Cleland, J., Wrench, A., Lloyd, S., & Sugden, E. (2018). *ULTRAX2020: Ultrasound technology for optimising the treatment of speech disorders: Clinicians' Resource Manual*. Glasgow: University of Strathclyde. http://dx.doi.org/10.15129/63372

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale III – 3rd Edition*. London: GL Assessment.

Dodd, B. (2013). *Differential diagnosis and treatment of children with speech disorder*. John Wiley & Sons.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology (DEAP)*. Psychology Corporation.

Fawcett, S., Bacsfalvi, P., & Bernhardt, B. M. (2008). Ultrasound as visual feedback in speech therapy for/r/with adults with Down Syndrome. *Down Syndrome Quarterly*, *10*(1), 4–12.

Furniss, R., & Wenger, T. (2018). Seeing the Big Picture. The use of ultrasound in treating functional speech disorders in school-aged children in a community health setting. *Journal of Clinical Practice in Speech-Language Pathology, 20*(2), 76-82.

Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I., & Zampini, M. L. (2008). Ultrasound imaging applications in second language acquisition. *Phonology and Second Language Acquisition*, *36*, 315-328.

Gibbon, F. E., & Paterson, L. (2006). A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. *Child Language Teaching and Therapy*, *22*(3), 275-292.

Gibbon, F. & Wood, S. (2010) *Visual Feedback Therapy with Electropalatography.* In: Interventions for Speech Sound Disorders in Children. Paul H Brookes Pub Co., Baltimore, pp. 509-536.

Gierut, J. A., Morrisette, M. L., & Dickinson, S. L. (2015). Effect size for single-subject design in phonological treatment. *Journal of Speech, Language, and Hearing Research*, *58*(5), 1464-1481.

Heng, Q., McCabe, P., Clarke, J., & Preston, J. L. (2016). Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical linguistics & phonetics*, *30*(3-5), 382-397.

Hitchcock, E. R., & Mcallister Byun, T. (2015). Enhancing generalisation in biofeedback intervention using the challenge point framework: A case study. *Clinical Linguistics & Phonetics*, *29*(1), 59–75.

Hitchcock, E. R., Harel, D., & Byun, T. M. (2015, November). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. In *Seminars in speech and language* (Vol. 36, No. 04, pp. 283-294). Thieme Medical Publishers.

Law, J., & Conti-Ramsden, G. (2000). Treating children with speech and language impairments: Six hours of therapy is not enough. *British Medical Journal*, *321*(7266), 908.

Law, J., Garrett, Z. & Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database Systematic Review, 3,* CD004110.

Lipetz, H. M., & Bernhardt, B. M. (2013). A multi-modal approach to intervention for one adolescent's frontal lisp. *Clinical Linguistics & Phonetics*, *27*(1), 1–17.

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.

Melo, R. M., Dias, R. F., Mota, H. B., & Mezzomo, C. L. (2016). Imagens de ultrasonografia de língua pré e pós terapia de fala. *Revista CEFAC*, *18*(1), 286–297.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The intelligibility in context scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research*, *55*(2), 648-656.

Miccio, A. W., Elbert, M., & Forrest, K. (1999). The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *American Journal of Speech-Language Pathology*, *8*(4), 347-363.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, *6*(1), 42.

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

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., Leece, M. C., & Maas, E. (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience*, *10*(August), 1–9.

Preston, J. L., Leece, M. C., McNamara, K., & Maas, E. (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology*, *26*(3), 840-852.

Preston, J. L., Maas, E., Whittle, J., Leece, M. C., & Mccabe, P. (2016). Limited acquisition and generalization of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, *30*(3–5), 363–381.

Powell, T. W., & Miccio, A. W. (1996). Stimulability: A useful clinical tool. *Journal of Communication Disorders*, *29*(4), 237-253.

Raven, J., Raven, J.C., & Court, J.H. (2003, updated 2004). Manual for Raven's progressive matrices and vocabulary Scales. San Antonio, TX: Harcourt Assessment.

Robbins, J. & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52(3),*271-7.

Semel, E., Wiig, E., & Secord, W. (2006). *Clinical evaluation of language fundamentals, fourth UK Edition.* San Antonio, TX: Harcourt Assessment.

Scobbie, J.M., Wrench, A. & van der Linden, M. (2008). Head-Probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. *Proceedings of the 8th International Seminar on Speech Production*, 373-376.

Shriberg, L. (1993). Four new speech and voice-prosody measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research, 36,* 105–140.

Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders*, *4*(1), 38-53.

Shriberg, L. (2017) Motor speech disorder-not otherwise specified (MSD-NOS): Prevalence and phenotype. *Proceedings of the 7th International Conference on Speech Motor Control Groningen. Stem-, Spraak-en Taalpathologie,* 22, 26.

Sjolie, G. M., Leece, M. C., & Preston, J. L. (2016). Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of communication disorders*, *64*, 62-77.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing: CTOPP2*. Pro-ed.

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders*, *48*(1), 25-40.

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.

Wrench, A., Scobbie, J. (2016) *Queen Margaret University ultrasound, audio and video multichannel recording facility (2008-2016).*Technical Report. CASL Research Centre.