Systematic Review of Ultrasound Visual Biofeedback in Intervention for Speech Sound Disorders

Eleanor Sugden[1]

Susan Lloyd[1,2]

Jenny Lam[3]

Joanne Cleland[1]

[1]School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

[2]Clinical Audiology, Speech & Language Research Centre, Queen Margaret University, Musselburgh, UK

[3]Paediatric Speech and Language Therapy, NHS Lothian, Edinburgh, UK

Keywords: speech sound disorders, biofeedback, intervention, ultrasound

Corresponding author:

Dr Eleanor Sugden

School of Psychological Sciences and Health

University of Strathclyde

Graham Hills Building

40 George Street

Glasgow, G1 1QE

Scotland

Ph: +44 141 548 4393

Email: eleanor.sugden@strath.ac.uk

Abstract

Background: As cost and access barriers to ultrasound technology have decreased, interest in using ultrasound visual biofeedback (U-VBF) as a tool for remediating speech sound disorders (SSD) has increased. A growing body of research has investigated U-VBF in intervention for developmental SSD; however, diversity in study design, participant characteristics, clinical methods, and outcomes complicates interpretation of this literature. There is thus a need for a synthesis and review of the evidence base for using U-VBF in intervention for SSD.

Aims: This study aimed to synthesise and evaluate the research evidence for U-VBF in intervention for developmental SSD.

Methods: A systematic review was conducted. Eight electronic databases were searched for peer-reviewed articles published prior to 2018. Details about study design, participants, intervention procedures, service delivery, intervention intensity, and outcomes were extracted from each study that met the inclusion criteria. The included studies were rated using both a critical appraisal tool and for their reporting of intervention detail.

Main Contributions: Twenty-eight papers, comprising 29 studies, met the inclusion criteria. The most common research design was single-case experimental design (44.8% of studies). The studies included between 1 and 13 participants (mean = 4.1) who had a mean age of approximately 11 years (range = 4;0 to 27 years).Within the research evidence, U-VBF intervention was typically provided as part of, or as an adjunct to, other articulatory-based therapy approaches. A range of lingual sounds were targeted in intervention, with 80.6% of participants across all reviewed studies receiving intervention targeting rhotics. Outcomes following therapy were generally positive with the majority of studies reporting that U-VBF facilitated acquisition of targets, with effect sizes ranging from no effect to a large effect. Difficulties with generalisation were observed for some participants. Most studies (79.3%)

were categorised as efficacy rather than effectiveness studies and represented lower levels of evidence. Overall, the reviewed studies scored more highly on measures of external validity than internal validity.

Conclusions: The evidence base for U-VBF is developing, however most studies used small sample sizes and lower strength designs. Current evidence indicates that U-VBF may be an effective adjunct to intervention for some individuals whose speech errors persist despite previous intervention. The results of this systematic review underscore the need for more high-quality and large-scale research exploring the use of this intervention in both controlled and community contexts.

Keywords: speech sound disorders, biofeedback, intervention, ultrasound

What this paper adds

What is already known on this subject? A growing body of research has investigated the use of U-VBF in intervention for SSD and has shown promising outcomes following intervention. However, to date no systematic review or synthesis of this research has been undertaken.

What this study adds. This paper provides speech and language therapists (SLTs) and researchers with a comprehensive review of the use of U-VBF in intervention for developmental SSD. The findings highlight a need for more large-scale research, representing a higher level of evidence, in both clinical and controlled contexts.

Clinical implications. There are a number of small-scale studies showing that U-VBF can be an effective component of intervention for some individuals with SSD, particularly in the initial stages of intervention when an individual is acquiring the target sound. However, limited detail about the pre-practice phase of intervention were reported in this literature, holding implications for the implementation and replication of the evidence into clinical practice. Although the reviewed studies represent lower levels of evidence, the results of this systematic review show that U-VBF can be effective for some individuals with a range of SSD subtypes whose errors have not responded to previous intervention.

Introduction

Speech sound disorders (SSD) are one of the most common childhood communication impairments with prevalence rates of between 2.3% and 24.6% (Law *et al.* 2000, Wren *et al.* 2016). While some children with SSD will acquire typical speech by the age of 8 years, other 8-year olds will continue to present with common clinical distortions (7.9% of children in a community sample) or a persistent SSD (3.6%; Wren *et al.* 2016) and approximately 1 to 2% of older children and young adults will demonstrate residual speech sound errors or persistent SSD (Flipsen 2015). Given the high prevalence of the disorder, it is unsurprising that children with SSD comprise a large proportion of speech and language therapy caseloads worldwide (Hegarty *et al.* 2018, McLeod and Baker 2014). Although there is no universally-agreed upon system for classifying the different presentations of SSD, it is largely established that children with the disorder form a heterogeneous group (Waring and Knight 2013). For some children, the cause of their SSD is known (e.g., associated with a hearing impairment, cleft palate, or genetic syndrome), however for the majority of children the cause of their SSD is unknown (Shriberg *et al.* 2010). For these children, the disorder is generally thought to stem from either a cognitive-linguistic difficulty in acquiring the phonological system of the ambient language or from a difficulty in acquiring the correct motor plan for particular sounds or sound sequences (Dodd *et al.* 2018, Waring and Knight 2013). These different causes may give rise to different subtypes or a combination of different subtypes of SSD, such as a phonological delay/disorder, an articulation disorder, or childhood apraxia of speech (CAS). Treatment for different subtypes of SSD can be effective (Law *et al.* 2004).

Worldwide, most of the interventions currently used by speech and language therapists (SLTs) to treat SSD rely on auditory input (Baker *et al.* 2018, Hegarty *et al.* 2018, Sugden *et al.* 2018, Brumbaugh and Smit 2013), with the child being provided with auditory information regarding the target sounds and their own productions. In these approaches to intervention, particularly for those aimed at remediating phonetic errors, the SLT typically provides auditory cues to children regarding how to move and where to place their articulators (e.g., Secord *et al.* 2007). However, such information can be difficult for many children to understand and apply to their production of target sounds as the articulators are largely invisible during speech. Since the 1980s, the prospect of using different imaging techniques as a tool for providing additional articulatory information in the form of visual biofeedback has been explored (see, for example, Ruscello 1995).

Visual biofeedback has been defined as “the use of instrumentation to make covert physiological processes more overt” (Huang *et al.* 2006: 1). In the field of SSD, this has generally involved providing visual information about the position, shape, movement, and/or placement of the articulators, most commonly the tongue. Such information is thought to be useful for remediating SSD as it allows for both the client and the clinician to access hitherto unavailable information about the tongue. This information can be beneficial in guiding diagnosis and treatment. Visual biofeedback, when used in intervention, is considered to provide immediate and concurrent Knowledge of Performance (KP) feedback to an individual regarding the nature of a target articulation. KP feedback is important when acquiring a new, or modifying an existing, motor plan (Maas *et al.* 2008).

Different approaches to providing real-time articulatory visual biofeedback in intervention for SSD have been explored, including electropalatography (EPG; e.g., Lee *et al.* 2009), ultrasound tongue imaging, and electromagnetic articulography (e.g., Katz *et al.* 2010). Most of the biofeedback intervention research to date has focussed on EPG (see Gibbon 2013 for a summary), which is a technique for displaying the timing and location of tongue-palate contact (Hardcastle and Gibbon 1997). While the approach can be effective at remediating SSD (e.g., Michi *et al.* 1993), purchasing the equipment and custom-made palates requires a large initial and continued ongoing costs. In contrast, in recent years the cost of ultrasound systems has been decreasing and portable machines are now readily available. This, coupled with improved frame rates and analysis methods, has led to growing research and clinical interest in the use of ultrasound tongue imaging in intervention for SSD.

When an ultrasound probe is placed under the chin in either mid-sagittal or coronal view an anatomically accurate image of the tongue surface is visible (see Figure 1). In intervention, this dynamic image can be used by both the treating clinician and the child to cue and provide feedback on the movement and position of the tongue. Although ultrasound visual biofeedback (U-VBF) is not among the most commonly used intervention approaches in everyday clinical practice (Hegarty *et al.* 2018, McLeod and Baker 2014), the combination of decreasing cost and increasing research evidence has led to a growing clinical interest in the approach. However, diversity in study design, participant characteristics, clinical methods, and outcomes present within the external evidence base make interpretation of the research literature difficult.

[INSERT FIGURE 1 ABOUT HERE]

One strategy to support health professionals, such as SLTs, in synthesising the external research evidence is to provide a systematic review of the relevant literature (Clarke 2011). To date no systematic review of the evidence for U-VBF in intervention for SSD has been conducted. Although a review of the role of intervention intensity in visual biofeedback intervention for treating SSD has recently been conducted (Hitchcock *et al.* 2019), this study did not include all of the literature on U-VBF nor did it consider the individual; studies in detail. A systematic review of all U-VBF would therefore provide clarity regarding the research evidence base and could be used by SLTs to advocate for funding and access to U-VBF as a therapy option for the children on their caseload. The findings of such a review could also support researchers in identifying avenues for future research. Thus, the purpose of this paper is to present the findings of a systematic review of the evidence for using U-VBF in intervention for developmental SSD. Given the prevalence of developmental (i.e. non-acquired) SSD in clinical caseloads worldwide (e.g., Broomfield and Dodd 2004) this review will focus on the use of U-VBF in intervention for developmental SSD of both known and unknown origins.

Method

A systematic review was conducted. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher *et al.* 2009) were used to inform our search strategy and reporting. This review was prospectively registered on PROSPERO (number: CRD42018088778).

Identification

We searched the following electronic databases: Scopus, MEDLINE, PROQUEST, LLBA, CINAHL, speechBITE, ASHA's online journal site, and the Cochrane Library. The following terms were used to search keywords, titles and abstracts with Boolean operators: intervention; therapy; treat*; ultrasound; biofeedback; speech; articulat*; phon*; apraxi*; dyspraxi*. A publication date limit of 2017 was included. The authors' personal literature databases were also searched to identify relevant articles. The final search was conducted on the 15th of February, 2018.

Screening and Eligibility

Title and abstract screening was conducted by the first two authors using Covidence software (Veritas Health Innovation). Conflicts regarding title and abstract screening were resolved through discussions between the first two authors.

Full-text copies of articles were sourced and assessed against the inclusion criteria for eligibility by the first two authors before being reviewed. The inclusion criteria were: (1) peer-reviewed articles published in or before 2017; (2) available in English (to allow for analysis by monolingual English speakers); (3) reporting on an investigation of the use of U-

VBF in intervention, and; (4) intervention delivered to children or adults identified as having a developmental (i.e., non-acquired) SSD. No exclusion criteria regarding study design were applied. Conflicts in full-text screening were resolved between the first, second and fourth authors.

Data Extraction and Analysis

The following data was extracted from each paper that met the inclusion criteria: publication details, study design, participant characteristics (including number, age, type of SSD, presence of concomitant disorders, and previous intervention history), intervention details (including procedures, other interventions provided within a session, ultrasound probe orientation, and provision of home practice), outcome measures used, analyses conducted, information on service delivery, and information on intervention intensity (following the conceptual framework proposed by Warren, Fey, & Yoder, 2007). Data were entered into a Microsoft Excel® spreadsheet. One paper included two studies, which were entered separately into the spreadsheet.

The following three rules were used to inform data extraction: (1) for location of studies, if not explicitly stated, the country of the first author's institutional affiliation was used; (2) only information explicitly reported in the paper was extracted (that is, even if information was available elsewhere, it was not extracted); (3) when information reported in each paper was unclear, this was conservatively coded as 'not reported'.

The first author completed data extraction for all included studies. We did not contact the authors of papers to clarify information as, because there was no limit to the start date of publication for our searches, this practice was considered to favour more recently published studies. Extracted data was analysed descriptively. A meta-analysis was not conducted due to the heterogeneity in research design and outcome measures used in the included studies.

Level of Evidence, Risk-of-Bias Assessment and Reporting of Intervention

The level of evidence of the included studies was determined according to the American Speech-Language-Hearing Association's level of evidence hierarchy (2004), which assigns a level (from a high of Ia to a low of IV) based on study design. The included studies were assessed for risk-of-bias using either the PEDro-P tool (Perdices *et al.* 2009) for group study designs or the RoBiN-T scale (Tate *et al.* 2015) for single-case research and case studies. All included studies were also rated using the Template for Intervention Description and Replication (TIDieR) checklist (Hoffmann *et al.* 2014). This is a 12-item checklist that considers the reporting of intervention protocols and procedures, with a view to considering the ease of replication of intervention studies. The first and fourth author completed the risk-of-bias ratings, with inter-rater reliability of 75.7%. The first and third author conducted ratings using the TIDieR checklist, with inter-rater reliability of 87.4%. Conflicts on both measures were resolved through discussion between the raters, and the consensus ratings are reported henceforth.

Results

Figure 2 shows the flow diagram for the systematic review, with reasons for exclusion identified. A total of 3529 articles were identified from database screening with three additional articles added from other search strategies. After duplicates were removed, 2128 articles underwent title and abstract screening. Of these, 66 full-texts were assessed for eligibility. Twenty-nine studies met the inclusion criteria, reported in 28 papers. Table 1 and 2 show summary information about each included study. Studies are presented chronologically and further details of each study is included in Appendix 1.

[INSERT FIGURE 2 ABOUT HERE]

Study Characteristics

The studies were conducted in the following countries: the United States (*n* = 14, 48.3%), Canada (*n* = 12, 41.4%), Scotland (*n* = 2, 6.9%), and Australia (*n* = 1, 3.4%). Studies were published in 14 different journals, most commonly in *Clinical Linguistics and Phonetics* (*n* = 10, 34.5%). Figure 3 shows the year of publication of the 29 studies. Single-case experimental design was the most frequently used research design (*n* = 13, 44.8%), with other studies using the following designs: case series (*n* = 8, 27.6%), case study (*n* = 6, 20.7%), randomised-controlled trial (*n* = 1, 3.4%), and quasi-experimental group design (*n* = 1, 3.4%). As shown in Table 1, most studies represent level III evidence (*n* = 15, 51.7%), with 13 studies (44.8%) representing level IIb evidence and one study categorised as level Ib evidence. Most studies could be categorised as efficacy studies (*n* = 23, 79.3%), with five studies examining the effectiveness of U-VBF (17.2%), and one study (3.4%) examining the long-term outcomes following intervention (Bacsfalvi and Bernhardt 2011). Studies included between 1 (*n* = 6, 20.7%) and 13 (*n* = 1, 3.4%) participants, with a mean of 4.1 and a median of 3 participants.

[INSERT TABLE 1 ABOUT HERE]

[INSERT FIGURE 3 ABOUT HERE]

Participant Characteristics

Combined, the studies reported data from 118 participants. However, several of the included studies reported secondary analyses of data collected from the same intervention study or included the same participants in subsequent studies (e.g., data from the participants Peran, Purdy, Palmer and Pamela were reported in the following papers Bernhardt *et al.* 2003, Bernhardt *et al.* 2005, Bacsfalvi *et al.* 2007, Bacsfalvi and Bernhardt 2011). Another participant (“Lilianne”) participated in two studies (Hitchcock and McAllister Byun 2015, study 2 from McAllister Byun *et al.* 2014). With these duplicates removed, intervention incorporating U-VBF has been provided to 103 unique individuals across the external evidence base.

Participants ranged in age from 4;0 years (Participant P1 from Heng *et al.* 2016) to 27 years (Participant 3 from Fawcett *et al.* 2008), with a mean age of approximately 11 years (median = approximately 10).

The included studies investigated the use of U-VBF in intervention for participants with a range of reported SSD subtypes or presentations, including: residual speech sound errors (*n* = 7; e.g., Preston *et al.* 2014); childhood apraxia of speech (*n* = 5; e.g., Preston *et al.* 2013), and; dysarthria (*n* = 2; e.g., one participant with concomitant CAS and dysarthria from Preston *et al.* 2016b). Other studies included participants with "persistent primary SSD" (e.g., Cleland *et al.* 2015: 579), "residual speech impairment" (e.g., Bernhardt *et al.* 2008: 153), and a "single persistent articulatory defect" (Shawker and Sonies 1985: 90). Other studies did not identify the subtype of SSD (e.g., Modha *et al.* 2008), although some studies identified that participants had a concomitant disorder that may have been related to the SSD such as hearing impairment (e.g., Bernhardt *et al.* 2003), repaired submucous cleft palate (e.g., Roxburgh *et al.* 2016), or Down syndrome (e.g., Fawcett *et al.* 2008).

Participants in most studies had received previous therapy for their speech errors (*n* = 25 studies, 86.2%), with one study reporting that participants had received no prior speech therapy (e.g., Cavin 2015) and one study reporting that this information about participants was unavailable to the researchers (e.g., Bressmann *et al.* 2016). One or more participants in 3 of the 25 studies (10.3% of all studies) had previously received U-VBF intervention; for the remaining studies, authors either explicitly reported that participants had not previously received U-VBF intervention (*n* = 9 studies, 31.0%), or did not report this aspect of participants' therapy history (*n* = 15, 55.2%).

Details of Intervention

Within the included studies, U-VBF was typically provided as one component of intervention. Nineteen studies (65.5%) reported including articulation-based production practice without the ultrasound, such as traditional articulation intervention or non-U-VBF intervention based on the principles of motor learning (e.g., Cleland *et al.* 2015, Sjolie *et al.* 2016). In these 19 studies, U-VBF was provided in a variety of schedules: in the pre-practice phase of intervention only (e.g., Heng *et al.* 2016), in alternating periods within an intervention session (e.g., Preston *et al.* 2014), on a gradually decreasing basis determined by the participant's progress (e.g., Hitchcock and McAllister Byun 2015), or it may have been used in some but not all sessions (e.g., Foss *et al.* 1990). Other studies (*n* = 5, 17.2%) reported providing other types of biofeedback in addition to U-VBF, such as EPG (e.g., Bacsfalvi *et al.* 2007). Five studies (17.2%) reported including auditory or perceptual training in addition to intervention focussed on production (e.g., Preston *et al.* 2017b).

Intervention procedures.

As shown in Table 2, a variety of speech sounds were targeted in the intervention. Rhotics were the most common therapy target, included for 83 participants (80.6% of the 103 unique participants). Sibilants were an intervention target for 17 participants (16.5%), velars were a target for 7 participants (6.8%), "sequences involving lingual sounds" were targeted for 6 participants (5.8%; Preston *et al.* 2013: 627), vowels were a target for 4 participants (3.9%) and /l/ was a target for 4 participants (3.9%). Other sounds that were targeted (for one participant each) were /n/, /t/, and a "general awareness of articulatory setting" (Lipetz and Bernhardt 2013: 5). One study (Foss *et al.* 1990) did not report specific intervention targets. Fifteen participants (14.6%) received U-VBF intervention for more than one target sound.

[INSERT TABLE 2 ABOUT HERE]

Fifteen studies (51.7%) reported using both midsagittal and coronal images in intervention, 4 studies reported using only a midsagittal view (13.8%), no studies reported using only a coronal view, one study reported using a transverse view (3.4%) and 10 studies (34.5%) did not report on the orientation of images used in intervention. One study reported using a headset to stabilise the ultrasound transducer (Cleland *et al.* 2015), with 8 studies (27.6%) reporting that the transducer was held by hand (and thus not stabilised), 2 studies (6.9%) reporting that the participant rested their head against a stabilised transducer, and 18 studies (62.1%) not reporting on this aspect of ultrasound tongue imaging.

Service delivery and intervention intensity.

Intervention was most commonly delivered individually (*n* = 16, 55.2% of studies). Two studies (6.9%) reported delivering intervention in a combination of individual and group contexts, and 11 studies (37.9%) did not report on this aspect of service delivery. Across the evidence base, intervention was delivered in a range of locations, including: in a university clinic (*n* = 9, 31.0%), at the participant's school or college (*n* = 3, 10.3%), at the participant's home (*n* = 2, 6.9%), and at a research centre (*n* = 2, 6.9%). Four studies (13.8%) reported that intervention was delivered in more than one location, and eighteen studies (62.1%) did not report where the intervention was delivered. Intervention was most commonly delivered by an SLT (in 21 studies, 72.4%), or a student SLT (4 studies, 13.8%). One study reported that either an SLT or a student delivered the intervention, and three studies (10.3%) did not report who delivered the intervention.

Eight studies (27.6%) reported providing home practice to participants. Home practice was completed without the ultrasound, and was generally completed with the support of parents. Two studies explicitly reported that home practice was not provided, and 19 studies (65.5%) did not report on this aspect of service delivery.

A range of intensities of intervention were reported within the included studies (see Appendix 1). Regarding dose frequency, intervention was most commonly delivered 1 × weekly (in 11 studies, 37.9%) or 2 × weekly (in 9 studies, 31.0%). Four studies (13.8%) did not report on dose frequency. Sessions were most commonly 60 minutes in duration (14 studies, 48.3%), but ranged in duration from 20 minutes (e.g., the individual sessions provided in Fawcett *et al.* 2008) to 2 hours (e.g., Preston and Leece 2017). Three studies (10.3%) did not report on session duration. For those studies that reported dose, between 60 (e.g., McAllister Byun *et al.* 2014) and an average of 366 production trials (e.g., Preston and Leece 2017) were provided per session in the practice phase of intervention. Eighteen studies (62.1%) did not report on the number of production trials provided in intervention. Only one study (Heng *et al.* 2016) reported on the number of production trials provided in the pre-practice phase of intervention; in this study, the two participants produced a mean of 65.8 and 52.8 pre-practice trials in each session respectively. Regarding the total number of sessions provided, participants received between one (e.g., the children from the south-central communities in Bernhardt *et al.* 2008, who received one U-VBF consultation in between two blocks of non-U-VBF intervention delivered by their regular SLT) and 18 sessions (Preston *et al.* 2013) of intervention, most commonly 14 sessions (in 31.0% of studies). Intervention was delivered for a period of between 1 and 22 weeks, with 12 studies (41.4%) not reporting on the total duration over which intervention was delivered. Cumulative intervention intensity reported in the included studies ranged from an average of 2335 to 6219 trials in structured practice.

Outcomes following intervention.

Table 2 presents a summary of intervention outcomes. More details are provided in Appendix 1. Nineteen of the studies (65.5%) reported positive results for all participants following intervention using U-VBF (e.g., Hitchcock and McAllister Byun 2015, Adler-Bock *et al.* 2007, Cleland *et al.* 2015), with ten studies (34.5%) reporting mixed results for the participants in that some individuals responded to the intervention but others did not. For example, the studies by Preston *et al.* (2016b) and Heng *et al.* (2016) reported that some participants responded to the intervention whereas others demonstrated no effect of treatment to target sounds or words. Several studies reported low levels of generalisation to untreated words (e.g., Preston *et al.* 2016b), with other studies reporting extensive generalisation for some participants but not for others (e.g., Sjolie *et al.* 2016).

Regarding measurement of outcomes following intervention, most studies (*n* = 19, 65.5%) included a perceptual judgement of correctness of the target by either expert or everyday listeners. Eleven studies (37.9%) analysed production accuracy based on phonetic transcription, five studies (17.2%) included acoustic analysis measures, three (10.3%) used structured qualitative descriptions of ultrasound images from before and after intervention, two (6.9%) used quantitative analysis of ultrasound images, one study (3.4%) reported participants' knowledge of the lingual components of the target articulation, and one study (3.4%) reported a general description of articulatory changes following intervention. No studies examined changes in participants' activity and participation following intervention, and only one study (Bacsfalvi & Bernhardt, 2011) considered the long-term outcomes following U-VBF intervention.

As shown in Table 2, the early studies exploring U-VBF did not commonly report statistical effect sizes from the intervention. More recent studies have tended to report effect sizes, most commonly the modified Cohen's *d* (d_2; Beeson and Robey, 2006). However, effect sizes for different aspects of intervention have been reported (for example, per target, per participant, or per phase of intervention) which limits comparison between studies. Overall, however, the results and conclusions of the included studies suggest that intervention incorporating U-VBF can be effective in the acquisition stages of motor learning for some individuals with SSD.

Quality Ratings

Critical appraisal (either with the PEDro-P or RoBiNT scale) and TIDieR ratings for each study are shown in Table 1. Overall, the included studies rated higher on assessment of external validity than of internal validity. For example, most studies included some information about generalisation measures, the dependent and independent variables, and baseline measures (including participant characteristics). Regarding internal validity, the studies generally scored poorly on the RoBiNT items of design, randomisation, and sampling. There was a general trend for more recently published studies to score more highly on the RoBiNT than earlier published studies. Ratings using the TIDieR checklist showed that studies did not often report sufficient detail about who provided intervention (only reported in 20.7% of studies), when and how much intervention was provided (20.7% of studies), where intervention was provided (34.5%), how intervention was delivered (34.5%), or what materials were used in treatment (58.6%). Most studies (65.5%) did not report on the fidelity of intervention.

Discussion

This paper synthesises the research evidence for the use of U-VBF in intervention for developmental SSD. Through a comprehensive search, we identified 29 studies included in 28 papers that reported on the use of U-VBF in intervention with just over 100 unique individuals with SSD. When the results of all studies are considered together, it appears that U-VBF can facilitate acquisition of a range of lingual speech targets for some individuals with SSD but that it does not always lead to generalisation. Many of the included studies represented lower levels of evidence and reported insufficient information to allow for implementation and replication with high levels of fidelity. Combined, these issues have implications for future research and for SLTs wishing to implement this emerging intervention into their clinical practice. These issues will now be considered.

U-VBF may Facilitate Acquisition but not Generalisation of Targets

The results of this review show that U-VBF has been used in intervention for remediating a range of SSD subtypes including those associated with hearing impairment, cleft palate or other causative conditions. Although most studies (65.5%) reported positive outcomes following intervention, just over half of the studies (51.7%) represented level III evidence, indicating that the design of the study—and thus the positive results—may have been susceptible to bias. The more recently published studies, however, tended to use a single-case experimental design which controls more strongly for threats to internal and external validity. When considering this more recent and higher quality evidence alone, it appears that intervention incorporating U-VBF may be effective for facilitating the acquisition of target sounds for some, but not for all, individuals with a persistent or residual SSD. This intervention, however, was delivered to a range of individuals with a range of different speech errors.

The included studies show that a variety of phones are candidate intervention targets. Most of the included studies reported targeting rhotics in intervention, either exclusively (e.g., Preston and Leece 2017, Bressmann *et al.* 2016, Modha *et al.* 2008) or as one of several targets (e.g., Bacsfalvi and Bernhardt 2011, Preston *et al.* 2014). The results of these studies indicate that U-VBF can be an effective adjunct to more traditional intervention approaches targeting rhotics in both prevocalic and vocalic position for those individuals who have not responded to other treatment approaches. The focus on /ɹ/ as an intervention target is potentially due to several factors: first, many of the studies originated from North America, where correct production of rhotics is considered to be socially important (Hitchcock *et al.* 2015). Second, as many SLTs report difficulty targeting /ɹ/ with traditional intervention approaches—possibly due to its articulatory complexity and the wide range of acceptable articulatory variations (Boyce 2015)—the error may not have responded to previous intervention. Finally, ultrasound tongue imaging is suitable for visualising correct productions of /ɹ/ as it allows for the dual anterior and posterior constrictions to be imaged simultaneously. Despite the focus on /ɹ/, other research studies have shown that U-VBF can be an effective intervention approach for a range of other lingual targets including velars, sibilants, and vowels (e.g., Cleland *et al*. 2015). Such a finding regarding the diverse applications of U-VBF is promising, given that other articulatory visual biofeedback techniques (such as EPG) may be less appropriate to use when treating sounds with limited or reduced palatal contact (i.e., vowels) or more posterior articulations (i.e., velars). Although there is research evidence supporting the use of U-VBF in intervention for a range of lingual targets, the included studies typically incorporated U-VBF as one component of an intervention program.

The use of U-VBF as an adjunct to other intervention approaches was common across the studies included in this systematic review. As such, it is difficult to conclude if the positive outcomes are due to U-VBF exclusively, to the effects of other interventions, or to the combination. Some of the reviewed studies conducted small-scale comparisons of intervention incorporating U-VBF with other motor-based approaches or non-U-VBF intervention (e.g., Preston *et al.* 2017a, Roxburgh *et al.* 2016), but diversity in study design and quality, varied responses to intervention, and low participant numbers hamper interpretation of these studies. Unfortunately no large randomised-controlled trials comparing approaches met the inclusion criteria for this systematic review. Since the final search to identify papers for this review was conducted, however, a randomised-controlled trial comparing U-VBF with traditional intervention has been published (Furniss and Wenger 2018). Seventeen children aged 5;4 to 11;8 years participated in this small-scale effectiveness study, which was conducted in a community-health context in Australia. The results showed a clear benefit for U-VBF in the initial stages of intervention compared to the traditional approach (with a median of 80% of target words produced correctly by children in the treatment group at the mid-point assessment compared to a median of 25% for the control group, $p = 0.04$). Despite the positive initial response to U-VBF, both approaches appeared to result in equivalent gains by the end of the treatment period (a maximum of 10 intervention sessions), with both groups producing a median of 90% of target words correctly ($p = 0.37$).

While many studies included in this systematic review (65.5%) reported positive results for all participants, ten studies (34.5%) reported that there were responders and non-responders to intervention incorporating U-VBF. For example, the single-case study by Preston *et al.* (2017a) examined the effects of motor-based intervention with U-VBF and motor-based intervention without U-VBF on the speech of 12 children with residual speech sound errors affecting /ɹ/. All children received both interventions, the order of which were counterbalanced across participants. The results of this study indicated varied responses to intervention, with "some children showing evidence of learning in only one condition, some showing evidence of learning in both conditions, and some failing to reveal evidence of learning with either approach" (p. 93). Other studies have reported similar findings, with some participants demonstrating minimal stimulability or acquisition of the target (e.g., Heng *et al.* 2016, Sjolie *et al.* 2016). Despite the presence of responders and non-responders to intervention, however, many studies have reported that U-VBF can facilitate acquisition—or at least support the elicitation—of new sounds in the speech of children with persistent or intractable errors (e.g., Bernhardt *et al.* 2003, Modha *et al.* 2008, Preston *et al.* 2016a). Combined with the results of the Furniss and Wegner (2018) study, it thus appears that U-VBF may be optimally effective in the initial stages of intervention. In contrast—and importantly for motor learning—several studies have reported that limited generalisation or transfer of these skills to untreated words, word positions, or sounds occurred (Preston *et al.* 2016b, e.g., study 1 from McAllister Byun *et al.* 2014, Sjolie *et al.* 2016). Limited generalisation following U-VBF intervention is consistent with findings about the effects of other visual biofeedback approaches for treating SSD (Gibbon and Paterson 2006).

It is also worth noting that generalisation in the context of SSD intervention involves both response and stimulus generalisation. The studies included in this systematic review overwhelmingly focussed on response generalisation, which is typically a measure of production accuracy on untrained items (such as to untrained words or other word positions; Baker and McLeod 2004). Stimulus generalisation, on the other hand, is a measure of generalisation of skills to new people or settings (Baker and McLeod 2004). Only one study included in the systematic review explicitly considered stimulus generalisation, which involved reporting qualitative comments from the participants and their parents about changes in how the participants were understood by family and friends (e.g., Bacsfalvi 2010). The limited consideration of stimulus generalisation within the literature limits the social validity of the evidence base, and belies the ultimate goal of intervention for SSD, which is to improve the communicative abilities and participation of people with speech disorders. Likewise, long-term follow-up is important if we are to ensure that any gains made during intervention are maintained.

Implementation and Replication

All of the studies included within the systematic review were rated using the TIDieR checklist (Hoffmann *et al.* 2014), which considers the reporting and descriptions of interventions in published studies. The majority of studies (82.8%) provided sufficient detail about how intervention is tailored to suit individual participants, a finding which contrasts with the reporting of intervention in RCTs in other areas of speech and language therapy (Ludemann *et al.* 2017). This likely reflects the predominant study designs used in U-VBF research—single-case experimental designs, case series, and case studies—which allow for individualised and tailored intervention (Byiers *et al.* 2012). Overall, however, the results show that the U-VBF intervention studies included in this systematic review were generally poor at including sufficient information about service delivery and intervention intensity, particularly about where intervention was delivered, and when and how much intervention was provided. Given the importance of intensity for effective intervention for SSD (Kaipa and Peterson 2016), the limited details about intensity provided in the evidence base make it difficult to adequately examine the contribution of intensity to the responses to intervention demonstrated by participants in the reviewed studies (Hitchcock *et al.* 2019). This limited reporting also has implications for the implementation of effective U-VBF interventions into clinical practice and replication of studies within research. Delivering intervention as described in the research evidence, with the same service delivery and intervention intensity, is an important component of fidelity and evidence-based practice (Kaderavek and Justice 2010). It is thus important that this information is included in future research studies reporting on U-VBF intervention.

When reviewing the studies included in this review it became clear that, while many studies reported detailed information about the practice phase of intervention (e.g., Hitchcock and McAllister Byun 2015, Preston and Leece 2017), very few provided specific details about pre-practice. For example, Preston *et al.* (2017b: 845) stated that "pre-practice included verbal and visual instruction to help the participant understand what was required for a correct production of the target movements". No specific details about the verbal and visual instruction were provided. Other studies reported that one or more participants spent most, if not all, of the intervention in the pre-practice phase (e.g., Heng *et al.* 2016, Preston *et al.* 2016b), speaking to the importance of pre-practice for the acquisition phase of motor learning and in intervention for motor speech disorders more generally. In light of the fact that many children on clinical caseloads may also struggle with acquisition of speech targets—combined with anecdotal evidence from SLTs that eliciting, cueing and shaping accurate productions can be challenging—the limited details reported in the literature regarding the type of feedback, cues, and prompts provided in pre-practice has implications for implementation. More information about the optimal strategies for cueing accurate productions is needed in published papers. A recent tutorial paper goes someway in addressing this by providing some information about the pre-practice and elicitation phases of U-VBF intervention (e.g., Preston *et al.* 2017c); although useful, this tutorial focussed on production of rhotics rather than on the range of lingual phonemes able to be treated with U-VBF and may thus not be immediately useful to many SLTs working clinically. Pre-practice is important to ensure stimulability of intervention targets (Maas *et al.* 2008), and—given that many of the U-VBF intervention studies reported that the approach can be beneficial in the initial acquisition stages of therapy—adequate reporting about this stage of intervention appears crucial for successful implementation and replication.

A final consideration for implementation and replication concerns who conducted the research included in this systematic review. Almost all of the reviewed studies were conducted by a small number of research groups from North America, with most recent research conducted by just one group (e.g., Preston *et al.* 2013, Preston *et al.* 2017a, Sjolie *et al*, 2016). Although the single-case experimental design studies conducted by this group generally scored highly on the quality assessment tool (the RoBiN-T, which measures the risk of bias of single case research) and the TIDieR checklist, it is unknown whether similar findings would be replicated by other research teams or by practising clinicians who may use slightly different research or clinical methods.

Limitations

This systematic review considered the effects of U-VBF intervention for individuals with developmental SSD. Although other studies have explored the use of U-VBF for other populations, such as adults with acquired apraxia of speech (Preston and Leaman 2014) or adults who have had a partial glossectomy (Blyth *et al.* 2016), acquired speech disorders were excluded from this review. This decision was made as intervention for acquired SSD may involve a different mechanism of action to U-VBF for developmental SSD, given that the individuals have previously developed intact and likely accurate motor plans.

Only peer-reviewed articles were considered for this systematic review. This decision meant that other evidence—such as that presented in conference presentations or theses—was not included in the review, which may have added a publication bias to the findings of this review. This decision was pragmatic, made in light of the acknowledged difficulties of searching for, acquiring, managing, and synthesising grey literature (Adams *et al.* 2017).

Finally, the intervention in the included studies was conducted in English with mono- or bi-lingual English-speaking participants. Although such a finding may be due to the inclusion criteria of the systematic review, it is unknown whether the results of U-VBF would be replicated in languages other than English.

Directions for Future Research

In addition to the future research needs identified above, this systematic review has revealed other directions that the field can pursue. For example, the majority of studies included in this systematic review used either single-case study experimental or case study designs and were conducted in tightly controlled contexts that do not reflect typical everyday clinical practice. Future large-scale clinical research examining the effectiveness of U-VBF is needed. As mentioned above, a small-scale RCT in everyday clinical practice has been conducted (Furniss and Wenger 2018), but larger studies are needed to explore the applicability of U-VBF in clinical practice. In addition, well-designed studies representing a higher level of evidence are needed.

Such high-quality, large-scale research would also facilitate exploration and identification of individual characteristics that are associated with positive responses to U-VBF intervention. As shown in this systematic review, there were responders and non-responders to intervention, but it is currently unknown what individual characteristics may predict response to treatment. It may be that one or more factors such as pre-treatment stimulability, motivation, age, cognitive ability, type and severity of SSD are related to treatment outcomes (Preston *et al.* 2017b), and identifying these characteristics could facilitate clinical decision-making about which individuals are best suited to intervention incorporating biofeedback.

It is essential that future research considers why and how intervention incorporating U-VBF can be effective in teaching individuals new motor plans and in facilitating motor learning. Identifying the active ingredients of the intervention (of both U-VBF specifically and biofeedback more generally) would allow for a deeper understanding of these issues and allow future research studies to incorporate theory-driven ingredients, or combination of ingredients, that are more likely to lead to successful outcomes. It is also important to understand how procedural differences in the delivery of U-VBF may impact outcomes and motor learning (Preston *et al.* 2018) as well as the optimal conditions of feedback and practice. In doing this, it would also be useful to consider the similarities and differences between U-VBF and other biofeedback approaches, particularly acoustic biofeedback which is cheaper and more readily available for clinicians. Acoustic biofeedback may be particularly useful in intervention targeting /ɹ/ as it provides information about the third and fourth formants, but to date no studies comparing the efficacy of U-VBF and acoustic biofeedback—nor other biofeedback approaches—have been conducted.

Conclusion

This paper presented a systematic review of the evidence for U-VBF in intervention for developmental SSD. In total, 29 studies published in 28 papers were identified that reported on an investigation of U-VBF in intervention. The studies were generally of lower-quality evidence; however, more recent research has been conducted with more robust study designs. The results of the included studies have shown that U-VBF can be used as part of intervention for a range of SSD subtypes and for a range of lingual targets. In particular, the results indicate that U-VBF may be effective for some individuals when used in the initial stages of motor learning, but may be less effective for promoting generalisation to untreated sounds or words. Future high-level research should explore the active ingredients of the intervention and the effectiveness of the approach within everyday clinical contexts, so that SLTs can deliver optimal intervention to individuals with SSD.

References

ADAMS, R. J., SMART, P. and HUFF, A. S. 2017, Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews,* 19**,** 432-454.

ADLER-BOCK, M., BERNHARDT, B. M., GICK, B. and BACSFALVI, P. 2007, The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology,* 16**,** 128-139.

AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION (ASHA). 2004, *Evidence-Based Practice (EBP): Step 3: Assessing the Evidence* (available at https://www.asha.org/Research/EBP/Assessing-the-Evidence/).

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**,** 206-217.

BACSFALVI, P. and BERNHARDT, B. M. 2011, Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, 25**,** 1034-1043.

BACSFALVI, P., BERNHARDT, B. M. and GICK, B. 2007, Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology***,** 9(1): 36-45.

BAKER, E. and MCLEOD, S. 2004, Evidence-based management of phonological impairment in children. *Child Language Teaching and Therapy,* 20**,** 261-285.

BAKER, E., WILLIAMS, A. L., MCLEOD, S. and MCCAULEY, R. 2018, Elements of phonological interventions for children with speech sound disorders: The development of a taxonomy. *American Journal of Speech-Language Pathology,* 27**,** 906-935.

BEESON, P.M. and ROBEY, R.R. 2006. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychological Review,* 16(4), 161-169.

BERNHARDT, B. M., BACSFALVI, P., ADLER-BOCK, M., SHIMIZU, R., CHENEY, A., GIESBRECHT, N., O'CONNELL, M., SIRIANNI, J. and RADANOV, B. 2008, Ultrasound as visual feedback in speech habilitation: exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics,* 22**,** 149-162.

BERNHARDT, B. M., BACSFALVI, P., GICK, B., RADANOV, B. and WILLIAMS, R. 2005, Exploring the use of electropalatography and ultrasound in speech habilitation. *Journal of Speech-Language Pathology & Audiology,* 29**,** 169-182.

BERNHARDT, B. M., GICK, B., BACSFALVI, P. and ASHDOWN, J. 2003, Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as elevated by trained listeners. *Clinical Linguistics & Phonetics,* 17**,** 199-216.

BLYTH, K. M., MCCABE, P., MADILL, C. and BALLARD, K. J. 2016, Ultrasound visual feedback in articulation therapy following partial glossectomy. *Journal of communication disorders,* 61**,** 1-15.

BOYCE, S. E. 2015, The Articulatory Phonetics of /r/ for Residual Speech Errors. *Seminars in Speech & Language,* 36**,** 257-270.

BRESSMANN, T., HARPER, S., ZHYLICH, I. and KULKARNI, G. V. 2016, Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors. *Clinical Linguistics & Phonetics,* 30**,** 345-362.

BROOMFIELD, J. and DODD, B. 2004, Children with speech and language disability: caseload characteristics. *International Journal of Language and Communication Disorders,* 39**,** 303-324.

BRUMBAUGH, K. M. and SMIT, A. B. 2013, Treating children ages 3-6 who have speech sound disorder: A survey. *Language, Speech, and Hearing Services in Schools,* 44**,** 306-319.

BYIERS, B. J., REICHLE, J. and SYMONS, F. J. 2012, Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, 21**,** 397-414.

CAVIN, M. 2015, The use of ultrasound biofeedback for improving English /r/. Victoria: University of Victoria, Department of Linguistics.

CLARKE, J. 2011, What is a systematic review? *Evidence Based Nursing,* 14**,** 64.

CLELAND, J., SCOBBIE, J. M., HEYDE, C., ROXBURGH, Z. and WRENCH, A. A. 2017, Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics,* 31**,** 35-55.

CLELAND, J., SCOBBIE, J. M. and WRENCH, A. A. 2015, Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, 29**,** 575-597.

DODD, B., REILLY, S., TTOFARI EECEN, K. and MORGAN, A. T. 2018, Articulation or phonology? Evidence from longitudinal error data. *Clinical Linguistics & Phonetics***,** 1-15.

FAWCETT, S., BACSFALVI, P. and BERNHARDT, B. M. 2008, Ultrasound as visual feedback in speech therapy for /r/ with adults with Down sydnrome. *Down Syndrome Quarterly*, 10**,** 4-12.

FLIPSEN, P. 2015, Emergence and prevalence of persistent and residual speech errors. *Seminars in Speech & Language,* 36**,** 217-223.

FOSS, M. C., WHITEHEAD, B., PATERSON, M. and WHITEHEAD, R. 1990, Ultrasound as a visual feedback aid for the hearing-impaired. *Journal of Diagnostic Medical Sonography,* 6**,** 80-86.

FURNISS, R. and 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**,** 76-82.

GIBBON, F. E. 2013, *Bibliography of Electropalatographic (EPG) Studies in English (1957-2013)* [Online]. Available: http://www.articulateinstruments.com/EPGrefs.pdf [Accessed].

GIBBON, F. E. and PATERSON, L. 2006, A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. *Child Language Teaching and Therapy,* 22**,** 275-292.

HARDCASTLE, W. J. and GIBBON, F. 1997. Electropalatography and its clinical applications. In*:* BALL, M. J. & CODE, C. (eds.) *Instrumental Clinical Phonetics.* London: Croom Helm.

HEGARTY, N., TITTERINGTON, J., MCLEOD, S. and TAGGART, L. 2018, Intervention for children with phonological impairment: Knowledge, practices and intervention intensity in the UK. *International Journal of Language & Communication Disorders***,** 1-12.

HENG, Q., MCCABE, P., CLARKE, J. and PRESTON, J. L. 2016, Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical Linguistics and Phonetics***,** 30(3-5): 382-397.

HITCHCOCK, E. R., HAREL, D. and MCALLISTER BYUN, T. 2015, Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in speech and language,* 36**,** 283-294.

HITCHCOCK, E. R. and MCALLISTER BYUN, T. 2015, Enhancing generalisation in biofeedback intervention using the challenge point framework: A case study. *Clinical Linguistics and Phonetics***,** 29(1): 59-75.

HITCHCOCK, E.R., SWARTZ, M.T. and LOPEZ, M. 2019, Speech sound disorder and visual biofeedback intervention: A preliminary investigation of treatment intensity. *Seminars in Speech and Language,* 40, 124-137.

HOFFMANN, T. C., GLASZIOU, P. P., BOUTRON, I., MILNE, R., PERERA, R., MOHER, D., ALTMAN, D. G., BARBOUR, V., MACDONALD, H., JOHNSTON, M., LAMB, S. E., DIXON-WOODS, M., MCCULLOCH, P., WYATT, J. C., CHAN, A.-W. and MICHIE, S. 2014, Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *British Medical Journal,* 348.

HUANG, H., WOLF, S. L. and HE, J. 2006, Recent developments in biofeedback for neuromotor rehabilitation. *Journal of NeuroEngineering and Rehabilitation,* 3**,** 11.

KADERAVEK, J. N. and JUSTICE, L. M. 2010, Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology,* 19**,** 369-379.

KAIPA, R. and PETERSON, A. M. 2016, A systematic review of treatment intensity in speech disorders. *International Journal of Speech-Language Pathology,* 18**,** 507-520.

KATZ, W. F., MCNEIL, M. R. and GARST, D. M. 2010, Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology,* 24**,** 826-837.

LAW, J., BOYLE, J., HARRIS, F., HARKNESS, A. and NYE, C. 2000, Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders,* 35**,** 165-188.

LAW, J., GARRETT, Z. and NYE, C. 2004, The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research,* 47**,** 924-943.

LEE, A. S. Y., LAW, J. and GIBBON, F. E. 2009, Electropalatography for articulation disorders associated with cleft palate. *Cochrane Database of Systematic Reviews*.

LIMA, F. L. D. C. N., SILVA, C. E. E., VASSOLER, A. M. D. O., FABBRON, E. M. G. and BERTI, L. C. 2018, Ultrasonographic analysis of lateral liquids and coronal fricatives: judgment of experienced and non-experienced judges. *Revista CEFAC*, 20**,** 422-431.

LIPETZ, H. M. and BERNHARDT, B. M. 2013, A multi-modal approach to intervention for one adolescent's frontal lisp. *Clinical linguistics & phonetics*, 27**,** 1-17.

LUDEMANN, A., POWER, E. and HOFFMANN, T. C. 2017, Investigating the adequacy of intervention descriptions in recent speech-language pathology literature: Is evidence from randomized trials useable? *American Journal of Speech-Language Pathology***,** 1-13.

MAAS, E., ROBIN, D. A., AUSTERMANN HULA, S. N., FREEDMAN, S. E., WULF, G., BALLARD, K. J. and SCHMIDT, R. A. 2008, Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology,* 17**,** 277-298.

MCALLISTER BYUN, T., BUCHWALD, A. and MIZOGUCHI, A. 2016, Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics,* 30**,** 249-276.

MCALLISTER BYUN, T., HITCHCOCK, E. R. and SWARTZ, M. T. 2014, Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research***,** 57(6): 2116-2130.

MCLEOD, S. and BAKER, E. 2014, Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics and Phonetics,* 28**,** 508-531.

MICHI, K.-I., YAMASHITA, Y., IMAI, S., SUZUKI, N. and YOSHIDA, H. 1993, Role of visual feedback treatment for defective /s/ sounds in patients with cleft palate. *Journal of Speech, Language, and Hearing Research,* 36**,** 277-285.

MODHA, G., BERNHARDT, M. B., CHURCH, R. and BACSFALVI, P. 2008, Case study using ultrasound to treat /ɹ/. *International Journal of Language & Communication Disorders,* 43**,** 323-329.

MOHER, D., LIBERATI, A., TETZLAFF, J. and ALTMAN, D. G. 2009, Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Journal of Clinical Epidemiology,* 62**,** 1006-1012.

PERDICES, M., SAVAGE, S., TATE, R., MCDONALD, S. and TOGHER, L. 2009, *Rater's Manual for Between-Group Studies (RCTs and NonRCTs). Introduction to the Physiotherapy Evidence Database (PEDro) Scale for Rating Methodological Quality, Adapted for PsycBITE (PEDro-P)*, Sydney, NSW, Australia, University of Sydney.

PRESTON, J. L., BRICK, N. and LANDI, N. 2013, Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology,* 22**,** 627-643.

PRESTON, J. L. and LEAMAN, M. 2014, Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology,* 28**,** 278-295.

PRESTON, J. L. and LEECE, M. C. 2017, Intensive treatment for persisting rhotic distortions: A case series. *American Journal of Speech Language Pathology***,** 26(4): 1066-1079.

PRESTON, J. L., LEECE, M. C. and MAAS, E. 2016a, Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in human neuroscience,* 10**,** 440.

PRESTON, J. L., LEECE, M. C. and MAAS, E. 2017a, Motor-based treatment with and without ultrasound feedback for residual speech-sound errors. *International Journal of Language and Communication Disorders,* 52**,** 80-94.

PRESTON, J. L., LEECE, M. C., MCNAMARA, K. and MAAS, E. 2017b, 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**,** 840-852.

PRESTON, J. L., MAAS, E., WHITTLE, J., LEECE, M. C. and MCCABE, P. 2016b, Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics,* 30**,** 363-381.

PRESTON, J. L., MCALLISTER BYUN, T., BOYCE, S. E., HAMILTON, S., TIEDE, M., PHILLIPS, E., RIVERA-CAMPOS, A. and WHALEN, D. H. 2017c, Ultrasound Images of the Tongue: A Tutorial for Assessment and Remediation of Speech Sound Errors. *Journal of Visualized Experiments***,** e55123.

PRESTON, J. L., MCALLISTER, T., PHILLIPS, E., BOYCE, S., TIEDE, M., KIM, J. S. and WHALEN, D. H. 2018, Treatment for residual rhotic errors with high- and low-frequency ultrasound visual feedback: A single-case experimental design. *Journal of Speech, Language, and Hearing Research,* 61**,** 1875-1892.

PRESTON, J. L., MCCABE, P., RIVERA-CAMPOS, A., WHITTLE, J. L., LANDRY, E. and MAAS, E. 2014, Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech Language & Hearing Research,* 57**,** 2102-15.

ROXBURGH, Z., CLELAND, J. and SCOBBIE, J. M. 2016, Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate. *Clinical Linguistics & Phonetics,* 30**,** 398-415.

RUSCELLO, D. M. 1995, Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders,* 28**,** 279-362.

SECORD, W. A., BOYCE, S. E., DONOHUE, J. S., FOX, R. A. and SHINE, R. E. 2007, *Eliciting Sounds: Techniques and Strategies for Clinicians*, Clifton Park, NY, Thomson Delmar Learning.

SHAWKER, T. H. and SONIES, B. C. 1985, Ultrasound biofeedback for speech training. Instrumentation and preliminary results. *Investigative radiology,* 20**,** 90-93.

SHRIBERG, L. D., FOURAKIS, M., HALL, S. D., KARLSSON, H. B., LOHMEIER, H. L., MCSWEENY, J. L., POTTER, N. L., SCHEER-COHEN, A. R., STRAND, E. A., TILKENS, C. M. and WILSON, D. L. 2010, Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics and Phonetics,* 24**,** 795-824.

SJOLIE, G. M., LEECE, M. C. and PRESTON, J. L. 2016, Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders,* 64**,** 62-77.

SUGDEN, E., BAKER, E., MUNRO, N., WILLIAMS, A. L. and TRIVETTE, C. M. 2018, Service delivery and intervention intensity for phonology-based speech sound disorders. *International Journal of Language & Communication Disorders,* 53**,** 718-734.

TATE, R., ROSENKOETTER, U., WAKIM, D., SIGMUNDSDOTTIR, L., DOUBLEDAY, J., TOGHER, L., MCDONALD, S. and PERDICES, M. 2015, *The Risk of Bias in N-*

of-1 trials (RoBiNT) Scale: An expanded manual for the critical appraisal of single-case reports, Sydney, Australia, Author.

VERITAS HEALTH INNOVATION *Covidence Systematic Review Software*. Melbourne, Australia: Veritas Health Innovation. Available at http://www.covidence.org.

WARING, R. and 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**,** 25-40.

WARREN, S. F., FEY, M. E. and YODER, P. J. 2007, Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews,* 13**,** 70-77.

WREN, Y., MILLER, L. L., PETERS, T. J., EMOND, A. and 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**,** 647-673.

Table 1. Study design, quality rating, and level of evidence of the included studies.

Study reference	Study design[a]	Quality rating[b]	TIDieR rating[c]	Level of evidence[d]
Shawker & Sonies (1985)	Case study	7/30	5	III
Foss et al. (1990)	Case study	2/30	4	III
Bernhardt et al. (2003)	Case series	12/30	5	III
Bernhardt et al. (2005)	Case series	9/30	5	III
Adler-Bock et al. (2007)	Case study	9/30	6	III
Bacsfalvi et al. (2007)	Pre-post	8/30	4	IIb
Berhnardt et al. (2008)	Case series	9/30	4	III
Fawcett et al. (2008)	AB design	9/30	6	III
Modha et al. (2008)	ATD (ABCBCA)	6/30	4	IIb
Bacsfalvi (2010)	MBD (participants)	9/30	6	IIb
Bacsfalvi & Bernhardt (2011)	Long-term follow-up study	7/30	2	III
Lipetz & Bernhardt (2013)	Single subject two-phase	8/30	7	IIb
Preston et al. (2013)	MBD (behaviours)	16/30	6	IIb
McAllister Byun et al. (2014)	Study 1: MBD (participants)	20/30	9	IIb
	Study 2: MBD (participants)	17/30	8	IIb
Preston et al. (2014)	MBD (participants)	14/30	7	IIb
Cavin (2015)	Case study	2/30	4	III
Cleland et al. (2015)	Case series	14/30	7	III
Hitchcock & McAllister Byun (2015)	Case study	14/30	9	III
Lee et al. (2015)	Case study	7/30	6	III
Bressmann et al. (2016)	RCT	5/11	3	Ib
Heng et al. (2016)	MBD (participants)	13/30	5	IIb
Preston et al (2016a)	Case series	16/30	8	III
Preston et al. (2016b)	MBD (behaviours)	14/30	8	IIb
Roxburgh et al. (2016)	Case series	11/30	3	III
Sjolie et al. (2016)	Single subject randomisation block design	18/30	8	IIb
Preston & Leece (2017)	Case series	18/30	8	III
Preston et al. (2017a)	ABACA/ACABA with MBD (behaviours and participants)	17/30	7	IIb
Preston et al. (2017b)	ATD with MBD (participants)	20/30	8	IIb

[a]A (e.g., in AB) = assessment / baseline / withdrawal phase; ATD = alternating treatment design, B (e.g., in AB) = treatment / intervention phase (treatment 1); C (e.g., in ABCBCA) = treatment / intervention phase (treatment 2); MBD = multiple baseline design (either across behaviours or participants); RCT = randomised controlled trial. [b]Group designs were rated with the PEDro-P tool (Perdices et al., 2009) for a total score out of 11; single-case experimental design studies and case studies were rated using the RoBiN-T scale (Tate et al., 2015) for a total score out of 30. [c]Template for Intervention Description and Replication checklist (Hoffmann et al., 2014).[d]Rated using ASHA's Level of Evidence System.

Table 2. Summary of participant details, intervention targets, and intervention outcomes of the included studies (reported chronologically)

Study	Participant details[a]	Intervention targets	Summary of outcomes[a]	Reported effect size[b]
Shawker & Sonies (1985)	*n* = 1 (9 years old)	/ɹ/	Positive	-
Foss et al. (1990)	*n* = 2 (ages not reported)	Unclear	Positive	-
Bernhardt et al. (2003)	*n* = 4 (aged 16 to 18)	/ɹ, s, ʃ, l/ and tense-lax vowel distinction	Positive	-
Bernhardt et al. (2005)	*n* = 4 (aged 16 – 18)	/ɹ, s, ʃ, l/ and tense-lax vowel distinction	Mixed	-
Adler-Bock et al. (2007)	*n* = 2 (aged 12 and 14)	/ɹ/	Positive	-
Bacsfalvi et al. (2007)	*n* = 3 (aged 18)	Vowels	Positive	-
Berhnardt et al. (2008)	*n* = 13 (aged 7 to 15;0)	/ɹ/, some work on sibilants for some children	Mixed	-
Fawcett et al. (2008)	*n* = 3 (aged 21 to 27)	/ɹ/	Positive	-
Modha et al. (2008)	*n* = 1 (13 year old)	/ɹ/	Positive	-
Bacsfalvi (2010)	*n* = 3 (aged 15 – 18)	/ɹ/	Positive	-
Bacsfalvi & Bernhardt (2011)	*n* = 7 (high school age)	/ɹ, s, ʃ, i, ɪ/	Mixed	-
Lipetz and Bernhardt (2013)	*n* = 1 (aged 15;9)	Awareness of articulatory space, sibilants	Positive	-
Preston et al. (2013)	*n* = 6 (9;10 to 15;10)	Sound sequences	Mixed	SMD range = −3.9 to 37.8 (per target)
McAllister Byun et al. (2014)	Study 1: *n* = 4 (6;1 to 10;3)	/ɹ/	Mixed	d_2 range = −3.2 to 2.3 (per target)
	Study 2: *n* = 4 (7;8 to 15;8)	/ɹ/	Positive	d_2 range = 1.0 to 16.7 (per target)
Preston et al. (2014)	*n* = 8 (aged 10 to 20)	/ɹ, s, ʧ/	Positive	d_2 range = 0 – 8.61 (per target)
Cavin (2015)	*n* = 1 (aged 22)	/ɹ/	Positive	-
Cleland et al. (2015)	*n* = 8 (aged 6-10.1)	/t, k, g, ʃ, ɹ/	Positive	100% non-overlapping data (PND)
Hitchcock & McAllister Byun (2015)	*n* = 1 (aged 11;2)	/ɹ/	Positive	d_2 range = −0.82 to 37.6 (per target)
Lee et al. (2015)	*n* = 1 (aged 13, male)	/ɹ/	Positive	-

Bressmann et al. (2016)	*n* = 6 (aged 7 to 10)	/ɹ/	Positive	-
Heng et al. (2016)	*n* = 2 (4;2 and 4;11)	/k, g/	Mixed	-
Preston et al (2016a)	*n* = 3 (10;8 – 14;3)	/ɹ, s, ʧ/	Mixed	d_2 range = 0.4 to 16.1 (per target)
Preston et al. (2016b)	*n* = 3 (aged 10 to 13)	/ɹ/	Mixed	-
Roxburgh et al. (2016)	*n* = 2 (9;2 and 6;2)	velars and /n/	Positive	-
Sjolie et al. (2016)	*n* = 4 (7;0 to 9;7)	/ɹ/	Mixed	*d* range = −0.756 to 0.783 (per participant)
Preston & Leece (2017)	*n* = 4 (13;11 – 22;8)	/ɹ/	Positive	d_2 range = 2.6 to 24.5 (per target)
Preston et al. (2017a)	*n* = 12 (10;1 – 16;7)	/ɹ/	Positive	Mean d_2 for ultrasound treatment = 4.90
Preston et al. (2017b)	*n* = 6 (8;2 - 16;8)	/ɹ, s/	Mixed	Mean d_2 for each condition= 14.52 and 8.31

[a]For more detail, see Appendix 1. [b]Some studies reported more than one effect size; due to space, only one is included in the table; absence of information in this column indicates that effect sizes were not reported in the study; *d* = Cohen’s effect size; d_2 = a variation of Cohen’s *d*, used when there is no variance in the baseline phase; PND = percent non-overlapping data, an effect size used in single-case research; SMD = standard mean difference.

Appendix 1. Summary of studies included in systematic review of the use of ultrasound visual biofeedback in intervention for developmental speech sound disorders

This appendix provides a chronological summary of the 29 peer-reviewed studies that were included in this systematic review. Readers are encouraged to refer to the original publication for more detail about each study.

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Shawker and Sonies (1985)	Case study *n* = 1 (9 years old) “Single persistent articulatory defect” (p. 90) Previously received speech therapy	*Service delivery and intervention intensity:* 1 × individual 60-min session per fortnight for a total of 4 sessions, with an additional 30-min introductory session provided immediately before the first session. *Intervention target(s):* /ɹ/ *Procedures:* A pre-recorded training tape, which included ultrasound and audio recording of an SLT producing /ɹ/ in isolation, syllables and in words, was provided. The participant was instructed to repeat the targets and match the correct tongue position. The participant was not provided with additional prompting or reinforcement from the examiner.	Increased accuracy of target during and immediately following intervention, but with reduced accuracy at 3-month maintenance assessment.	7/30	5
Foss, Whitehead, Paterson, and Whitehead (1990)	Case study *n* = 2 (only 1 reported, ages not reported) Participant(s) had a hearing impairment	*Service delivery and intervention intensity:* 2 x 45-50-min sessions per week delivered during the academic quarter. One session in the SLT’s office, the other at the ultrasound laboratory. Subject 1 completed 10 weeks of therapy. *Intervention target(s):* unclear *Procedures:* Less than 30 minutes each session spent using the ultrasound.	U-VBF is potentially useful for improving articulation problems for some individuals with hearing impairment.	2/30	4

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Bernhardt, Gick, Bacsfalvi, and Ashdown (2003)	Case series (trained listener study) *n* = 4 (aged 16 to 18) Severe to profound sensorineural hearing impairment Three participants came from families that speak English as a second language All had received previous speech therapy	*Service delivery and intervention intensity:* Individual intervention, delivered by an SLT at a university clinic in 30-minute sessions 1 × week for a total of 14 sessions. *Intervention target(s):* /ɹ, s, ʃ, l/ and tense-lax vowel distinction *Procedures:* All participants received U-VBF and EPG intervention. The first 6 sessions included either U-VBF or EPG, then 3 sessions with other approach. The final 5 sessions incorporated both technologies. Intervention started with demonstrations by SLT, then practised in isolated segments following a hierarchy to production in phrases.	Treated targets improved significantly more than non-treated targets. Targets that were absent or marginal pre-treatment showed greater gains following intervention.	12/30	5
Bernhardt, Bacsfalvi, Gick, Radanov, and Williams (2005)	Case series (everyday listener study) *n* = 4 (aged 16 – 18) Severe to profound sensorineural hearing impairment All had received previous speech therapy Participants from Bernhardt et al. (2003)	*Service delivery and intervention intensity:* 14 × individual weekly sessions at university research laboratory delivered by SLT, with follow-up sessions at school (without visual biofeedback). *Intervention target(s):* /ɹ, s, ʃ, l/ and tense-lax vowel distinction *Procedures:* All participants received U-VBF and EPG intervention. The first 6 sessions included either U-VBF or EPG, then 3 sessions with other approach. The final 5 sessions incorporated both technologies. Intervention started with silent movements, then followed a hierarchy to production in phrases.	Everyday listeners observed changes in production for some but not all speakers or targets.	9/30	5

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Adler-Bock, Bernhardt, Gick, and Bacsfalvi (2007)	Case study *n* = 2 (aged 12 and 14) SSD of unknown origin Monolingual Canadian English speakers Participants had previously received intervention, but not with visual biofeedback	*Service delivery and intervention intensity:* 14 × 60-mins session delivered by an SLT at the university laboratory. Participant 1 attended 2 – 3 × weekly for 6 weeks, participant 2 attended "on occasional weekends" (p. 131) over 20 weeks, with one visit to their hometown with school-based SLT. *Intervention target(s):* /ɹ/ *Procedures:* Session 1 involved use of traditional elicitation techniques to determine stimulability; subsequent sessions provided U-VBF. Protocol began with awareness and progressed through a production hierarchy. *Home practice:* Homework (non-U-VBF) provided for 10-mins per day.	Increased accuracy in single-words and some phrases following intervention. For these participants, more time with U-VBF resulted in better outcomes.	9/30	6
Bacsfalvi, Bernhardt, and Gick (2007)	Pre-post *n* = 3 (aged 18) Severe to profound sensorineural hearing impairment All had participated in the Bernhardt et al. (2003) study	*Service delivery and intervention intensity:* 2 × weekly for 6 weeks. One session delivered at a university research laboratory for 60 – 90 minutes, the other delivered at school for 45 minutes. Intervention delivered by an SLT, and all sessions had some individual and some group instruction. *Intervention target(s):* vowels *Procedures:* All participants received both EPG and U-VBF intervention in most sessions. All sessions began with awareness, then practice in a hierarchy from isolation to phrases. *Home practice:* Home practice completed with a family member or school assistant, without the use of visual feedback. Home practice was not completed every time by all participants.	Both U-VBF and EPG can facilitate production accuracy of vowels in adolescents with hearing impairment.	8/30	4

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Berhnardt, Bacsfalvi, Alder-Bock, Shimizu, Cheney, Giesbrecht, O'Connell, Sirianni, & Radanov (2008)	Case series (BCB research design) *n* = 13 (aged 7 to 15;0) Residual speech impairment Some minor deviations for 10 participants on oral-motor exam All had received previous speech therapy	*Service delivery and intervention intensity:* All children received a period of non-U-VBF treatment, delivered by local SLT for 7 to 9 sessions. Children then received either 3-4 × U-VBF consultations (total 2-3 hours) over two days (for "Northern communities") or 1 × U-VBF consultation (for "south-central" communities) delivered by project SLT. Finally, children received 7 – 8 sessions delivered by local SLT. *Intervention target(s):* mainly /ɹ/, some work on sibilants for some children *Procedures:* Major components of targets introduced as silent gestures, then vocalisation added. *Home practice:* Completed without U-VBF.	Eleven children showed improvement in production of target following the ultrasound consultation For these children, more U-VBF may have been related to better outcomes.	9/30	4
Fawcett, Bacsfalvi, and Bernhardt (2008)	AB design *n* = 3 (aged 21 to 27 years) All participants diagnosed with Down Syndrome All had received previous speech therapy	*Service delivery and intervention intensity:* 1 × session per week over 4 months (total 16 sessions), delivered at a research centre by an SLT. The first 5 sessions (45-minutes long) involved a mix of group and individual therapy, subsequent sessions were delivered individually for 20 minutes. Participants attended 13 or 14 of the planned 16 sessions. *Intervention target(s):* /ɹ/ *Procedures:* U-VBF used in 10 of the sessions. Included auditory discrimination, then followed a hierarchy from silent posturing to production in words. *Home practice:* Parents asked to complete practice of the movement, speech sounds, and words for 10-15 minutes per day. Participants reported completing practice ranging from several days per week to daily.	All participants improved production of the target and were able to produce it in single words.	9/30	6

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Modha, Bernhardt, Church, and Bacsfalvi (2008)	Alternating treatment design (ABCBCA) *n* = 1 (13 year old male) Only speech error was on /ɹ/ Participants had received previous speech therapy	*Service delivery and intervention intensity:* 1 × 30-45 minute individual session per week for a total of 9 sessions delivered by an SLT. *Intervention target(s):* /ɹ/ *Procedures:* The first session did not involve U-VBF. This was followed by 4 sessions with U-VBF and 2 sessions without U-VBF. Following this, 2 more non-U-VBF sessions were provided. Intervention followed a hierarchy from isolation to phrases; for the U-VBF, components of /ɹ/ were practised individually and in combination. *Home practice:* Completed with parent.	Ratings of speech samples indicated improvement in production of target, particularly after introduction of ultrasound. U-VBF facilitated acquisition of the target.	6/30	4
Bacsfalvi (2010)	Non-concurrent multiple baseline design across participants *n* = 3 (aged 15 – 18) Severe to profound sensorineural hearing impairment. Participants had previously received speech therapy. Two participants had previous experience with U-VBF	*Service delivery and intervention intensity:* 1 x 45-min session per week for a total of 7 or 8 sessions. Intervention delivered by an SLT. *Intervention target(s):* /ɹ/ *Procedures:* Intervention commenced with awareness of anatomy and ultrasound system. Components of /ɹ/ articulation taught sequentially, with each movement taught in isolation and then in combination. Once silent articulation was established, voicing was added and /ɹ/ was practised in a hierarchy from isolation to word level and clusters. *Home practice:* Provided, but details unclear.	All participants were able to learn the gestural components of the target, and one participant showed production changes at the word level.	9/30	6

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Bacsfalvi and Bernhardt (2011)	Long-term follow-up study *n* = 7 (high school age) Severe to profound sensorineural hearing loss All had participated in previous U-VBF intervention studies	*Service delivery and intervention intensity:* Mixed. Participants had participated in previous intervention studies conducted by this group of researchers (e.g., Bernhardt et al., 2003; Bacsfalvi et al., 2007). Most participants continued to have school-based intervention following participation in these studies. *Intervention target(s):* /ɹ, s, ʃ, i, ɪ/ *Procedures:* Refer to original intervention studies for more details. *Home practice:* Refer to original intervention studies for more details.	Five of seven participants maintained and/or generalised productions.	7/30	2
Lipetz and Bernhardt (2013)	Single subject two-phase *n* = 1 (aged 15;9, male) Participant diagnosed with ASD, although speech and language skills WNL except for residual frontal lisp Previously received speech therapy	*Service delivery and intervention intensity:* 11 x 60-minute intervention sessions held at the university and home delivered by an SLT student or SLT over two phrases, separated by a 10-week break. Phase 1 comprised 6 sessions over 8 weeks. Phase 2 comprised 5 sessions over 4 weeks. *Intervention target(s):* Phase 1: awareness of articulatory space; Phase 2: sibilants *Procedures:* Phase 1 focused on general awareness of the articulatory setting, with U-VBF incorporated in two sessions for less than 20 minutes each time. Phase 2 included traditional articulation therapy hierarchies and visual-acoustic biofeedback (spectrograms). *Home practice:* Participant completed home practice independently following each intervention session. Home practice in Phase 1 included observation of ultrasound images for a range of sounds, and vocal training exercises.	No changes to production accuracy following Phase 1. Direct training of speech (Phase 2) led to more improvements but may be related to training from Phase 1.	8/30	7

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Preston, Brick, and Landi (2013)	Multiple baseline across behaviours *n* = 6 (9;10 to 15;10) Persisting speech sound errors associated with CAS Participants had received previous speech therapy Some participants had a range of other "clinical concerns" (p. 630)	*Service delivery and intervention intensity:* 2 × 60-min individual sessions per week for a total of 18 sessions delivered over 10-16 weeks by an SLT or graduate SLT student. Of the approx. 20% of sessions reviewed, an average of 228 trials were elicited per session. *Intervention target(s):* sound sequences *Procedures:* Thirty minutes of each session involved U-VBF (split into 2 blocks of 15 min each), with 15-20 min of each session incorporating traditional table-top approaches.	All participants demonstrated improved accuracy at the word level. Some participants generalised to sequences that were phonetically similar to those targeted. SMD range = −3.9 to 37.8 (per target)	16/30	6

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
McAllister Byun, Hitchcock, and Swartz (2014)	Study 1: Multiple-baseline across participants n = 4 (6;1 to 10;3) Three participants had received previous speech therapy	*Service delivery and intervention intensity:* 2 × 30-45 minute individual sessions per week for 8 weeks (total 16 sessions) delivered by an SLT and student assistant. 60 trials elicited per session in the practice phase of intervention. *Intervention target(s):* /ɹ/ *Procedures:* Two instructional sessions (covering interpretation of ultrasound images and familiarisation with a bunched tongue shape for rhotics) followed by 14 biofeedback sessions. Biofeedback sessions included pre-practice (review of rhotic articulation) and 3-5 minutes of free-play with ultrasound, and then progressed to practice.	Three participants were able to produce perceptually more accurate rhotics when using the ultrasound, however generalisation to non-biofeedback contexts was minimal. d_2 range = −3.2 to 2.3 (per target)	20/30	9
	Study 2: Multiple-baseline across participants n = 4 (7;8 to 15;8) All participants had received speech therapy	*Service delivery and intervention intensity:* 2 × 30-45 minute individual sessions per week for 8.5 weeks (total 17 sessions) delivered by an SLT and student assistant. In the practice phase of intervention a dose of 60 was elicited. *Intervention target(s):* /ɹ/ *Procedures:* Identical to Study 1, above, except that a third instructional session was added. This session discussed the range of tongue shape possibilities for a perceptually correct /ɹ/ and to "try out" (p. 2124) different tongue shapes.	Large treatment gains for all participants, with generalisation observed. d_2 range = 1.0 to 16.7 (per target)	17/30	8

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Preston, McCabe, Rivera-Campos, Whittle, Landry, & Maas (2014)	Multiple baseline across participants *n* = 8 (aged 10 to 20) Residual speech sound errors Seven participants had previously received speech therapy	*Service delivery and intervention intensity:* 2 × 60-min sessions per week for a total of 14 sessions delivered by an SLT. An average of 210 practice trials provided per session. *Intervention target(s):* /ɹ, s, ʧ/ *Procedures:* Each session included 2 × 13-min periods with U-VBF alternated with 2 × 13-min periods without U-VBF. Each session included pre-practice and structured chaining practice. Time with the ultrasound was gradually reduced as participants met step-up criterion in the structured practice. Sessions 1-7 used either a prosodic or a non-prosodic cueing condition and sessions 8-14 used the alternate condition. *Home practice:* No home practice provided.	Most participants demonstrated increased accuracy at the word level, with generalisation and retention (at a 2-month follow-up) observed. d_2 range = 0 – 8.61 (per target)	14/30	7
Cavin (2015)	Case study *n* = 1 (22 year old male) No previous speech therapy	*Service delivery and intervention intensity:* 9 × 60-minute sessions. *Intervention target(s):* /ɹ/ *Procedures:* Followed a hierarchy from awareness to production of sentences.	Improvement in production accuracy in untreated words following intervention.	2/30	4

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Cleland, Scobbie, and Wrench (2015)	Case series *n* = 8 (aged 6-10.1) Persistent primary SSD All had received previous speech therapy	*Service delivery and intervention intensity:* 1 × 60-minute session per week for 12 weeks, delivered by an SLT in a sound-treated room. *Intervention target(s):* velar fronting, post-alveolar fronting, backing, idiosyncratic production of /ɹ/ *Procedures:* Each session included between 10 and 40 minutes of U-VBF, with the remainder of the session incorporating non-U-VBF therapy. Therapy individualised, but followed a general hierarchy and ultrasound videos of other children producing targets. *Home practice:* Provided with home practice to be completed 5 × weekly. Parents reported completing it 1 – 2 × weekly.	All participants demonstrated significant progress on targets following intervention. 100% non-overlapping data (PND)	14/30	7
Hitchcock and McAllister Byun (2015)	Case study *n* = 1 (“Lilianne”, 11;2) Received previous speech therapy, and participated in McAllister Byun et al., 2014 study	*Service delivery and intervention intensity:* Individual sessions, each 30-45 minutes in duration delivered by an SLT 1 × weekly. In total, 11 sessions delivered over 11 weeks. Sessions elicited 60 trials in the practice phase of intervention. *Intervention target(s):* /ɹ/ *Procedures:* Intervention structured according to a challenge point hierarchy. Each session commenced with 5-mins of free play with the ultrasound (pre-practice).	Generalisation and maintenance of target observed. d_2 range = −0.82 to 37.6 (per target)	14/30	9

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Lee, Wrench, and Sancibrian (2015)	Case study *n* = 1 (aged 13, male) Developmental articulation disorder Range of language learning diagnoses Received previous speech therapy	*Service delivery and intervention intensity:* 1 × 30-minute individual session delivered per week for 12 weeks by an SLT intern in a university clinic (total 12 sessions). *Intervention target(s):* /ɹ/ *Procedures:* U-VBF used for the entirety of the first 5 sessions, then reduced for subsequent sessions. Intervention followed a hierarchy from silent posturing to production in phrases. Ultrasound transducer was handheld, and typically a mid-sagittal image was used.	Increase in production accuracy of target following intervention.	7/30	6
Bressmann, Harper, Zhylich, and Kulkarni (2016)	RCT *n* = 6 (aged 7 to 10, mean 8.8) “unresolved articulation error concerning the sound /ɹ/” (p. 347)	*Service delivery and intervention intensity:* Individual therapy 1 × 60-minute session per week for 10 weeks (total 10 sessions). *Intervention target(s):* /ɹ/ *Procedures:* 4 participants received articulation therapy supplemented with 10-minutes of U-VBF per session. 2 participants received the articulation therapy only. Note that this study was not designed as a comparison of U-VBF and non-U-VBF intervention.	Both interventions resulted in both quantitative and qualitative gains following intervention.	5/11	3

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Heng, McCabe, Clarke, and Preston (2016)	Multiple-baseline across participants *n* = 2 (4;2 and 4;11) Both participants had received previous speech therapy	*Service delivery and intervention intensity:* Intervention delivered individually by an SLT student. Sessions were 50 minutes in duration, for a total of 6 sessions delivered over 3 weeks. In pre-practice, a mean of 65.8 (participant 1) or 52.8 (participant 3) trials were provided per session. In practice, P1 received an average of 61.7 trials per session. Participants received a cumulative dose of 317 or 765. *Intervention target(s):* /k, ɡ/ *Procedures:* U-VBF used in pre-practice phase of intervention only. P3 did not progress to the practice phase of intervention. Participants received between 10 and 50 minutes of U-VBF per session. *Home practice:* No home practice provided.	P1 improved production of targets at syllable level during treatment and achieved correct production at the word level at a follow-up assessment. P3 showed no acquisition or change in production.	13/30	5
Preston, Leece, and Maas (2016)	Case series *n* = 3 (10;8 – 14;3) CAS and residual speech sound errors All participants had received previous speech therapy	*Service delivery and intervention intensity:* 2 × 60-min sessions per day, 5 days per week, for 2 weeks, for a total of 16 hr delivered by an SLT. *Intervention target(s):* /ɹ/ for two participants, /s, ʧ/ for one participant *Procedures:* Each session included 2 × 12-min periods with U-VBF alternated with 2 × 12-min periods without U-VBF. Each session included pre-practice and structured chaining practice. Each session began with 6-8 min (50 trials) of auditory perceptual training.	All participants demonstrated acquisition of the target. Generalisation and retention were mixed. d_2 range = 0.4 to 16.1 (per target)	16/30	8

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Preston, Maas, Whittle, Leece, and McCabe (2016)	Multiple-baseline across behaviours *n* = 3 (aged 10 to 13) All participants had CAS, and one was diagnosed with flaccid dysarthria All had weak language abilities Participants had previously received speech therapy	*Service delivery and intervention intensity:* 14 × 60-min sessions delivered by SLT. For sessions in which participants progressed to the practice phase of intervention, an average of 142 trials were provided. *Intervention target(s):* /ɹ/ *Procedures:* Each session included 2 × 13-min periods with U-VBF alternated with 2 × 13-min periods without U-VBF. Each session included pre-practice and structured chaining practice (if the participant met the criteria to progress). Sessions 1-7 used either a prosodic or a non-prosodic cueing condition and sessions 8-14 used the alternate condition.	Two participants demonstrated some acquisition within treatment sessions. No participants demonstrated generalisation to untreated words.	14/30	8
Roxburgh, Cleland, and Scobbie (2016)	Case series (expert listener study) *n* = 2 (9;2 and 6;2) Repaired submucous cleft palate Participants had previously received speech therapy	*Service delivery and intervention intensity:* 2 blocks of intervention delivered by an SLT, each with 8 × 60-min sessions delivered over 8 weeks, with a 6 week break in between blocks (total 16 sessions). *Intervention target(s):* velars and /n/ *Procedures:* Block 1 provided intervention using Visual Articulatory Models. Block 2 used U-VBF.	Children made improvements following both interventions.	11/30	3

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Sjolie, Leece, and Preston (2016)	Single subject randomisation block design n = 4 (7;0 to 9;7) All had previously received speech therapy	*Service delivery and intervention intensity:* 2 × 60-min sessions per week for a total of 14 sessions over 7 to 8 weeks delivered by an SLT student. Of the approx. 15% of sessions that we reviewed, an average of 215 trials were provided across both pre-practice and practice. *Intervention target(s):* /ɹ/ *Procedures:* Each week, one session used U-VBF and the other did not. In the U-VBF sessions, 3 × blocks of 13 min were included, with the first and third using the ultrasound. All sessions commenced with elicitation (pre-practice) and progressed to structured practice.	U-VBF may facilitate acquisition for some children. U-VBF neither facilitated nor inhibited retention or generalisation compared to intervention without U-VBF. d range = −0.756 to 0.783 (per participant)	18/30	8
Preston and Leece (2017)	Case series n = 4 (13;11 – 22;8) Residual speech sound errors, with one diagnosed with Participants had received previous speech therapy	*Service delivery and intervention intensity:* 1 × 2-hr session per morning and 1 × 2-hr session per afternoon for 5 days, delivered by SLTs over one week. In total, participants received 14 hr of intervention (7 sessions), with an average of 366 trials in the practice phase per session (cumulative dose 4475 to 6219 trials per participant). *Intervention target(s):* /ɹ/ *Procedures:* Each session included 2 × 12-min periods with U-VBF alternated with 2 × 12-min periods without U-VBF. Each session included pre-practice and structured chaining practice. The first 7 sessions included 6-8 min of auditory perceptual judgement training, the final 7 sessions included randomised production practice.	All participants demonstrated improved speech accuracy and generalisation following intervention. d_2 range = 2.6 to 24.5 (per target)	18/30	8

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Preston, Leece, and Maas (2017)	ABACA/ACABA single-case design with multiple baselines across behaviours and participants *n* = 12 (10;1 – 16;7) Residual speech sound errors Eight participants had previously received speech therapy	*Service delivery and intervention intensity:* 2 × 60-min sessions per week delivered by an SLT, for an average of 4115 total trials for each participant in structured practice. *Intervention target(s):* /ɹ/ *Procedures:* All participants received two treatment conditions for 7 sessions each, the order of which was counterbalanced across participants. Both conditions included a principles of motor learning (PML)-based intervention, with or without U-VBF (*PML + U-VBF* or *PML + no-U-VBF*). Each session included 4 × 13-min periods. In the *PML + U-VBF* condition, U-VBF was provided in the first and third periods only. For both conditions, all sessions included pre-practice and structured chaining practice.	Both conditions results in increased accuracy; however, some children responded differently to the different approaches. Mean d_2 for ultrasound treatment = 4.90	17/30	7

Reference	Study design[a] and participant details	Details of intervention[b]	Study outcomes and effect size[c]	Quality rating[d]	TIDieR[e] rating
Preston, Leece, McNamara, and Maas (2017)	Alternating treatments with multiple baseline across participants *n* = 6 (8;2 - 16;8) All participants had CAS, and two participants were rated as possibly having dysarthria	*Service delivery and intervention intensity: 2* × 60-min sessions per week delivered by an SLT for a total of 14 sessions over 7 weeks. An average of 69 practice trials provided per session in the prosodic-variation condition and 98 practice trials per session in the no-prosodic variation condition (average total practice trials: 970 and 1365, respectively). *Intervention target(s):* /ɹ, s/ *Procedures:* Each session was divided into two 30-min sections which used different treatment approach: a prosodic variation treatment condition, or a no-prosodic variation condition. Each 30-min section comprised approx. 8 min of auditory training (50 trials) and production training (including pre-practice and practice), with 10 min with U-VBF and 10-min without U-VBF.	All participants showed greater generalisation on targets treated using the prosodic condition. Mean d_2 for each condition= 14.52 and 8.31	20/30	8

[a]Key for study design: A (e.g., in AB) = assessment / baseline / withdrawal phase; B (e.g., in AB) = treatment / intervention phase (treatment 1); C (e.g., in ABCBCA) = treatment / intervention phase (treatment 2); RCT = randomised controlled trial.
[b]Note that not all studies provided information about all elements of service delivery, intervention intensity, or intervention procedures. Absence of this information within the table reflects absence of this information in the published paper.
[c]Some studies reported more than one effect size; due to space, only one is included in the appendix; absence of information in this column indicates that effect sizes were not reported in the study; *d* = Cohen's effect size; d_2 = a variation of Cohen's *d*, used when there is no variance in the baseline phase; PND = percent non-overlapping data, an effect size used in single-case research; SMD = standard mean difference.
[d]Group designs were rated with the PEDro-P tool (Perdices, Savage, Tate, McDonald, & Togher, 2009); single-case experimental design studies and case studies were rated using the RoBiN-T scale (Tate et al., 2015). The PEDro-P tool rates studies using a binary-scored 11-item scale. The RoBiN-T scale rates studies using a 3-point scale for 15 items, providing a total score out of 30.
[e]Template for Intervention Description and Replication checklist (Hoffmann et al., 2014).