1-1-2019

Speech Sound Disorder and Visual Biofeedback Intervention: A Preliminary Investigation of Treatment Intensity

Elaine Hitchcock
Montclair State University, hitchcocke@montclair.edu

Michelle T. Swartz
Montclair State University

Melissa Lopez
Montclair State University

Speech Sound Disorder and Visual Biofeedback Intervention: A Preliminary Investigation of Treatment Intensity

Elaine R. Hitchcock, Ph.D., CCC-SLP,[1]
Michelle T. Swartz, M.A., CCC-SLP,[1] and Melissa Lopez, M.A., CFY-SLP[1]

ABSTRACT

A growing body of research suggests that cases of speech sound errors that have not responded to previous intervention can sometimes be eliminated through speech therapy incorporating visual biofeedback. Aside from considerations related to the specific biofeedback type, acquisition and generalization of a motor plan may be linked to treatment intensity. Several researchers have raised the possibility that inadequate dosage levels may present a significant barrier to success. Thus, the current study aimed to assess the relationship between treatment intensity and treatment outcomes. Twenty-nine articles reporting the use of visual biofeedback intervention for speech sound disorder were identified and coded for treatment intensity using the cumulative intervention index and outcomes using mean level difference scores. Findings reveal small but significant relationships between measures of treatment intensity and efficacy, which should be interpreted with caution given the preliminary nature of this review. Further research in this area is necessary, as inconsistencies in reporting intensity and outcomes across studies underscore the need for more systematic terminology and reporting methods.

KEYWORDS: residual speech errors, childhood apraxia of speech, treatment intensity, dose, visual biofeedback

Learning Outcomes: As a result of this activity, the reader will be able to (1) identify three major types of visual biofeedback used as a method of treatment for speech sound errors and identify reasons for the variable nature of treatment outcomes; (2) define formal definitions for measuring treatment intensity including dose, dose frequency, and total intervention duration; (3) describe the challenges associated with measuring treatment intensity and efficacy for treatment of speech sound errors.

[1]Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, New Jersey.

Address for correspondence: Elaine R. Hitchcock, PhD, CCC-SLP, Department of Communication Sciences and Disorders, Montclair State University, 1515 Broad Street, 2nd Floor, Bloomfield, NJ 07003
(e-mail: hitchcocke@mail.montclair.edu).

Articulation and Phonological Disorders; Guest Editor, Kathryn Cabbage, Ph.D., CCC-SLP.

Semin Speech Lang 2019;40:124–137.
DOI: https://doi.org/10.1055/s-0039-1677763.
ISSN 0734-0478.

A growing body of research indicates that visual biofeedback techniques may be efficacious in treating speech sound disorder (SSD). However, across biofeedback studies, individual treatment response ranges from no measurable difference to complete generalization of target productions at the level of conversational speech.[1–7] Such outcomes raise questions about why treatment gains generalize for some, but not all, participants. Aside from individual differences (e.g., attention, motivation), relevant factors related to treatment planning may include the type of visual biofeedback, complexity of treatment targets, practice schedule, feedback type (e.g., knowledge of performance vs. knowledge of results), schedule of feedback, and treatment intensity (i.e., frequency, duration). Several researchers have begun to compare biofeedback versus non-biofeedback conditions[8,9] as well as the impact of practice and feedback,[10] yet presently, none of these findings have definitively revealed a favored treatment, method of practice, or feedback condition. *Treatment intensity*, however, has been gaining widespread attention as a significant factor in SSD research, particularly for disorders associated with motor speech skills.[11,12] Yet despite the growing focus on treatment intensity for traditional SSD intervention programs, little biofeedback research has directly addressed its importance for the successful acquisition and generalization of new speech skills.

VISUAL BIOFEEDBACK TECHNIQUES

Recent technological advances have the potential to revolutionize the clinical management of children with SSD. Many studies have focused on alternative treatments, specifically visual biofeedback intervention, for a subset of individuals whose errors persist despite intervention. This subset includes children with motor speech difficulties such as residual speech errors (RSE) and childhood apraxia of speech (CAS). Both populations include individuals whose errors often do not remediate with traditional speech therapy methods. Throughout the article, we will refer to these errors as "residual" errors, a general term in which we are including the category of "persistent" errors. Such errors may continue even in children who have undergone extensive periods of intervention.[13,14] When not remediated, the presence of speech sound errors may negatively impact a child's overall social, emotional, and academic well-being.[15,16]

Traditionally, learners are taught to alter inaccurate speech sounds through auditory feedback instruction. The introduction of biofeedback allows learners to use a visual modality to identify aspects of speech that are challenging for them to distinguish under typical circumstances.[17] Specifically, the use of *visual biofeedback* provides a real-time visual representation of the user's speech which can be compared against a model representing correct production of a target sound. The external representation of an accurate speech target facilitates correction of an error pattern instead of relying on internal self-perception. These effects have been documented across various biofeedback technologies, including acoustic biofeedback, in which the client views a computer-generated acoustic representation (e.g., LPC spectrum[5,18] or spectrogram[1]) of his/her speech; ultrasound biofeedback,[7,19] in which an ultrasound probe held beneath the chin generates an image of the client's tongue during speech; electropalatography,[20,21] which uses a pseudopalate to register and display areas of contact between the client's tongue and palate; and electromagnetic articulography, in which an animated 3D tongue avatar moves in real time with the client's own tongue.[22,23]

The benefit of incorporating visual biofeedback as a dynamic, visual teaching modality is rapidly gaining recognition in the field of speech–language pathology. Several small-scale studies have found that visual biofeedback treatment can be successful in eliminating speech sound errors. However, variations in treatment response may be, in part, secondary to differences in treatment intensity.

VISUAL BIOFEEDBACK AND TREATMENT INTENSITY

According to Warren et al,[24] treatment intensity may be an essential variable in optimizing the effectiveness of intervention in the field of communication disorders. This premise is not surprising given that intervention intensity is one of the well-cited principles of rehabilitation. The intensity of a training task, aimed at the improvement of a specific skill such as production of a speech sound, can affect neural connectivity[25] and serve to document intensity as a necessary element for achieving optimal progress in speech intervention. As noted previously, little biofeedback research has directly addressed the importance of treatment intensity for the successful acquisition and generalization of motor learning. Findings by Preston and colleagues[26] suggest that the impact of treatment intensity may be maximized during specific phases of treatment. For example, high intensity levels may be more valuable when the individual is learning a new motor plan compared with later, in the generalization phase. In a case series study composed of four participants presenting with residual rhotic errors that had not resolved with previous therapy, a high-frequency intervention program composed of twice daily sessions for a period of 1 week resulted in significant gains for all participants. Thus, clinicians might improve client outcomes by initially providing an intense frequency intervention schedule, ideally greater than the typical model of one to two sessions per week.

Perhaps because the field has yet to definitively adopt a standard for reporting treatment intensity variables, methodological descriptions vary widely across treatment studies' with most studies reporting some combination of the following treatment measures: (1) overall duration, (2) total number of sessions, (3) number of sessions per week, (4) length of treatment session, (5) session length in minutes, and (6) number of teaching episodes per session. Relatively few studies report and/or define these variables in the exact same fashion, making valid comparisons of study outcomes challenging.

Using the Cumulative Intervention Index to Evaluate Treatment Intensity

Warren et al[24] reviewed the existing literature relative to dosage in communication and language development, concluding that "treatment intensity research is of utmost importance in developing optimally efficacious interventions."[24] They noted that there was surprisingly little literature in communication disorders designed to specifically evaluate the impact of treatment intensity. They proposed formal definitions for measuring treatment intensity including *dose* (number of teaching episodes per session), *dose frequency* (number of treatment sessions per week), and *total intervention duration* (total time period of intervention) and developed the cumulative intervention index (CII). The CII is calculated by multiplying these dosage variables to generate a single numerical measure intended to aid in the investigation of treatment efficacy. Standardization of dosage parameters provides a framework well suited for comparing treatment outcomes on the basis of intensity. Furthermore, analysis of treatment intensity provides a structure for examining the range of treatment responses reported for participants using visual biofeedback. Some researchers, such as Preston and colleagues, have investigated the effects of high-intensity ultrasound biofeedback treatment for RSE and speech errors in childhood apraxia,[26,27] but these effects have not yet been compared with other biofeedback techniques using one unified outcome measure. Given evidence of both strong responders and nonresponders to biofeedback treatment,[1–4] exploring the role of treatment intensity may help identify dosage factors that align with positive participant treatment outcomes.[9]

Article Review

The purpose of the current study was to review the existing literature on biofeedback for treatment of speech sound errors using the CII as a measure of comparison for treatment intensity. The aim of this analysis was not to propose a singular cause–effect relationship between treatment intensity and efficacy, as it is clear that other variables such as treatment type, practice conditions, and feedback methods may impact treatment success. Instead, the aim of this paper was to *explore* the relationship between treatment intensity and treatment outcomes. Outcomes associated with dose parameters have the potential to influence future visual biofeedback intervention programs.

METHOD

Search Strategy

A computer-based search was undertaken to identify papers investigating the use of biofeedback for treatment of SSDs from January 1980 to July 2018. The search strategy followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) search guidelines.[28] All available databases related to health sciences were selected for inclusion in the present review. The flow diagram depicting the study selection process is presented in Fig. 1.

Search terms. Ten databases related to health sciences were comprehensively searched via Montclair State University's library portal for peer-reviewed journal articles. For a detailed list of databases and search terms, see Table 1.

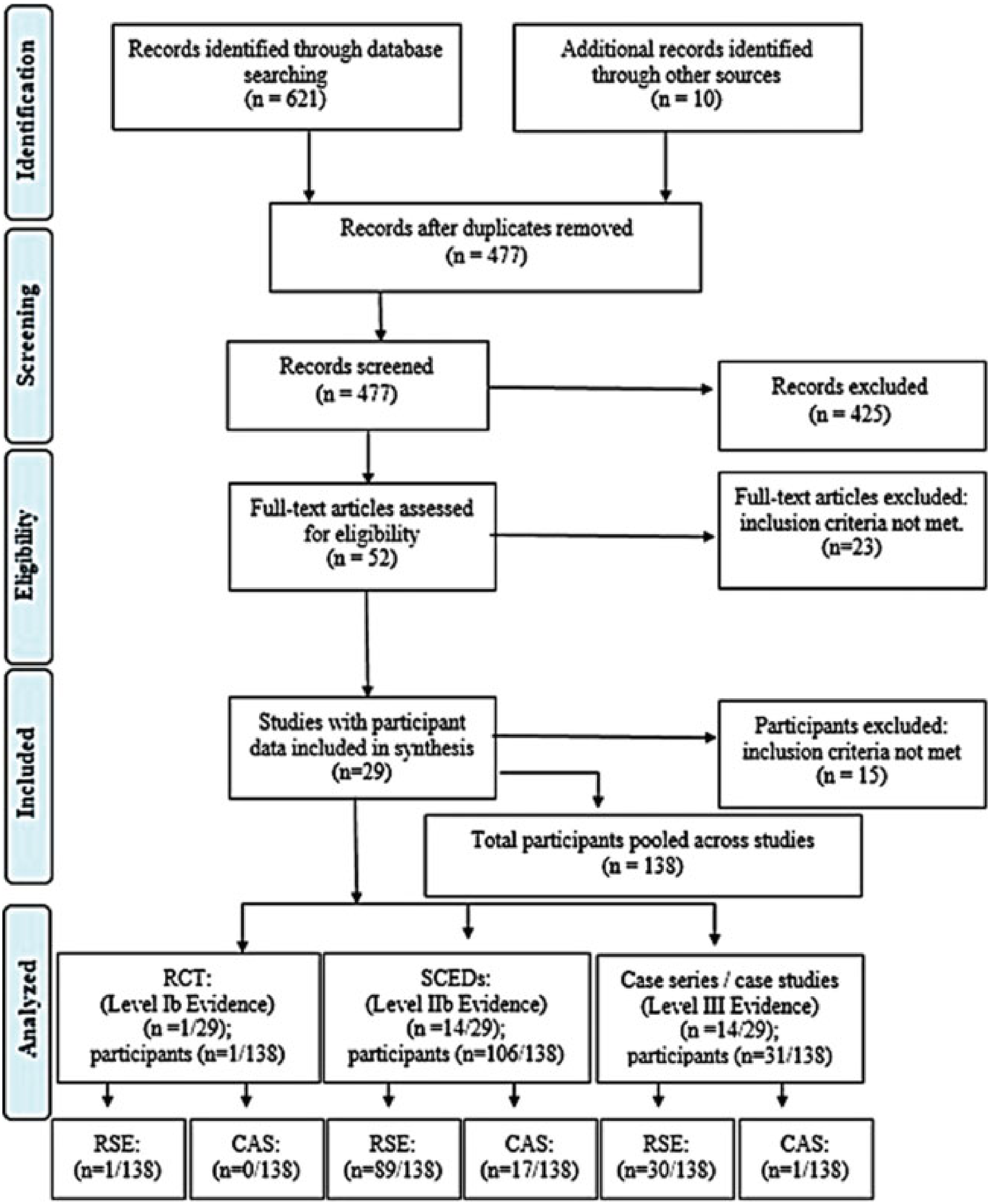

Figure 1 Flow diagram of study selection (adapted from Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA]).[28] RCT, randomized controlled trial; SCED, single-case experimental design; RSE, residual speech error; CAS, childhood apraxia of speech.

Table 1 Databases and Terms used in Systematic Review

Databases searched	Search terms
Social Sciences Citation Index	"biofeedback," "ultrasound," electropalatography,"
Science Citation Index	or "visual acoustic" and
Academic Search Complete	"speech sound disorder," "speech sound disorders,"
Education Research Complete	"phonological intervention," "residual speech errors,"
CINAHL Complete	"articulation impairments,"
PsycINFO	"speech impairments,"
ERIC	"speech therapy,"
ScienceDirect	or "speech"
SciELO	
British Library EthOS	

Screening. Articles were exported to EndNote X7,[29] where duplicates were removed, resulting in 477 studies. An initial screening of the 477 studies revealed a diverse set of etiological factors associated with a diagnosis of SSD. Thus, several diagnostic exclusions were put in place to limit the diversity of the participants being reviewed and to reduce the likelihood of multiple conditions influencing reported outcome measures. The participants in the articles were required to have an SSD. The SSD could be classified as a residual or persistent speech error of unknown etiology or due to a motor speech disorder such as CAS without any comorbid excluded diagnoses (e.g., autism). Article titles and abstracts were screened to exclude studies considered to be outside of the scope of the present review. Thus, all papers related to participants diagnosed with the following disabilities were excluded: structural anomalies (e.g., cleft palate), genetic syndromes (e.g., Down's syndrome), cerebral palsy, hearing impairment, and autism. Screening removed 425 articles, leaving 52 to be assessed for eligibility. A follow-up online search containing the names of the authors and intervention types was conducted to ensure all relevant articles were found.

Eligibility. Copies of articles were obtained and assessed against the final inclusion criteria before being reviewed. These criteria were as follows: (1) peer-reviewed articles published between January 1980 and July 2018; (2) written in English (to allow analysis by monolingual English-speaking authors); (3) reporting quantitative participant data outcomes focused on articulation/phonology within treatment, during probe measures, or in spontaneous speech; and (4) using visual biofeedback during intervention. All visual biofeedback methods were included, as were all levels of evidence, except for systematic reviews, as they do not contain individual participant data.[30] During the eligibility phase, researchers were not blinded to article title or authors. Additionally, the first and second authors of the current article, both of who have published articles under consideration, did not review their own publications. Excluded articles were not analyzed further.

Final review. A final review of the remaining qualified articles was conducted to assess individual cases within each study. Eligibility decisions regarding inclusion of participant data were made using a predetermined set of criteria: (1) participant presented with no physiological and physical limitations, (2) participant was under 22 years of age with a speech sound error, (3) the study demonstrated performance of within-treatment practice trials, and (4) the study provided individual participant data for review. This selection yielded 29 articles from which 138 participants were identified for analysis. Each study was assigned a level of evidence according to published ASHA guidelines.[30]

Lastly, the following data were collected and summarized for analysis: participant information (i.e., age, gender, speech and language diagnosis), treatment information (i.e., biofeedback type, speech sound targets), study design (i.e., case study, single-case experimental design, level of evidence), session length in minutes, dose, dose frequency, treatment duration, and effect size (see Table 2).

Analysis

There were 29 visual biofeedback studies that met the inclusion criteria for the present review. All of the preestablished categories were independently coded by two members of the research team (i.e., a trained graduate student

Table 2 Biofeedback Type, Gender, Age, Reported Diagnoses/Speech Targets Per Study

Biofeedback type	Author(s)	Publication date	Gender			Age (mean; months)	Dx	Speech targets
			M	F	NR			
EPG	Dagenais et al[46]	1994	–	2	–	103	RSE	Sibilants
	Gibbon et al[47]	1993	–	1	–	108	RSE	Stops
	Gibbon et al[48]	1990	1	–	–	132	RSE	Sibilants
	Hitchcock et al[21]	2017	2	3	–	93	RSE	Rhotics
	Schmidt[49]	2007	5	2	–	125	RSE	Multiple
Ultrasound	Adler-Bock et al[3]	2007	1	–	–	144	RSE	Rhotics
	Bressmann et al[50]	2016	1	–	–	108	RSE	Rhotics
	Byun et al[7]	2014	4	4	–	112	RSE	Rhotics
	Cleland et al[51]	2015	5	2	–	89	RSE	Multiple
	Heng et al[52]	2016	1	1	–	54	RSE	Velars
	Hitchcock and Byun[53]	2015	–	1	–	134	RSE	Rhotics
	Lee et al[54]	2015	1	–	–	156	RSE	Rhotics
	Modha et al[55]	2008	1	–	–	156	RSE	Rhotics
	Preston et al[6]	2013	3	–	–	156	CAS	Multiple
	Preston et al[8]	2014	6	2	–	149	RSE	Multiple
	Preston et al[41]	2015	3	–	–	132	CAS	Rhotics
	Preston et al[27]	2016	1	–	–	158	CAS	Rhotics
	Preston and Leece[26]	2017	2	2	–	119	RSE	Rhotics
	Preston et al[56]	2017	8	4	–	144	RSE	Rhotics
	Preston et al[57]	2017	7	4	–	119	CAS	Rhotics
	Preston et al[58]	2018	12	–	–	136	RSE	Rhotics
	Shawker and Sonies[59]	1985	–	1	–	108	RSE	Rhotics
	Sjolie et al[60]	2016	–	–	4	102	RSE	Rhotics
Visual	McAllister Byun[61]	2017	5	2	–	147	RSE	Rhotics
acoustic	McAllister Byun et al[18]	2016	6	3	–	116	RSE	Rhotics
	Byun and Hitchcock[5]	2012	10	1	–	108	RSE	Rhotics
	McAllister Byun and Campbell[9]	2016	7	4	–	135	RSE	Rhotics
	Shuster et al[1]	1995	1	1	–	144	RSE	Rhotics
(StaRt App)	McAllister Byun et al[62]	2017	–	1	–	156	RSE	Rhotics

Abbreviations: CAS, childhood apraxia of speech; multiple, multiple treatment targets per study; RSE, residual speech error.

or an author). Lastly, the authors met to review the coding of the 29 studies and resolve any discrepancies to finalize the dataset.

Treatment intensity measures. As stated previously, a primary goal of the present research was to explore the nature of the relationship between treatment intensity and the efficacy of visual biofeedback intervention for individuals with SSD. To investigate these variables, the CII was selected as a composite measure of treatment intensity. The CII, as described earlier, is the product of *dose* × *dose frequency* × *total intervention duration*.[24] Eighteen of the 29 studies adequately reported the variables of dose, dose frequency, and total intervention duration, making it possible to calculate a CII score for 114 participants out of a total of 138 participants.[a] A summary of the calculated CII values per study is reported in Table 3.

Treatment effect measures. Initially, an *improvement rate difference* (IRD), defined as the improvement rate of the baseline phase subtracted from the improvement rate of the

[a] *Five of the reviewed studies reported an average dose value and one reported an average CII value. Due to the preliminary nature and limited accessible data for the current review, we did not exclude these studies.*

Table 3 Biofeedback Type, Research Design, LOE, Effect Size, MLD, CII Per Study

Biofeedback type	Author(s)	Publication date	Research design	LOE	Reported Effect size	Reported MLD	Calculated CII
EPG	Dagenais et al[46]	1994	Cseries	III	N	N	2,300[a]
	Gibbon et al[47]	1993	CS	III	N	N	–
	Gibbon et al[48]	1990	CS	III	N	N	–
	Hitchcock et al[21]	2017	SCED	IIb	N	N	840
	Schmidt[49]	2007	Cseries	III	N	N	–
Ultrasound	Adler-Bock et al[3]	2007	CS	III	N	N	–
	Bressmann et al[50]	2016	RCT	Ib	N	N	–
	Byun et al[7]	2014	SCED	IIb	Y	Y	840
	Cleland et al[51]	2015	Cseries	III	N	Y	–
	Heng et al[52]	2016	SCED	IIb	N	N	–
	Hitchcock and Byun[53]	2015	CS	III	Y	Y	1,020
	Lee et al[54]	2015	CS	III	N	N	–
	Modha et al[55]	2008	CS	III	N	N	–
	Preston et al[6]	2013	SCED	IIb	Y	Y	4,104
	Preston et al[8]	2014	SCED	IIb	Y	Y	2,940
	Preston et al[41]	2015	G w/in MB	IIb	N	Y	1,988
	Preston et al[27]	2016	Cseries	III	Y	Y	–
	Preston and Leece[26]	2017	Cseries	III	Y	Y	5,117[a]
	Preston et al[56]	2017	SCED	IIb	Y	Y	4,115
	Preston et al[57]	2017	SCED	IIb	Y	Y	1,144[a]
	Preston et al[58]	2018	SCED	IIb	Y	Y	2,592
	Shawker and Sonies[59]	1985	CS	III	N	N	–
	Sjolie et al[60]	2016	SCED	IIb	Y	N	3,010
Visual	McAllister Byun[61]	2017	SCED	IIb	Y	Y	1,200
acoustic	McAllister Byun et al[18]	2016	SCED	IIb	Y	N	1,920
	Byun and Hitchcock[5]	2012	SCED	IIb	N	N	2,400
	McAllister Byun and Campbell[9]	2016	SCED	IIb	Y	Y	1,200
	Shuster et al[1]	1995	Cseries	III	N	N	1,600[a]
(StaRt App)	McAllister Byun et al[62]	2017	CS	III	N	Y	1,200

Abbreviations: CII, cumulative intervention index; CS, case studies; Cseries, case series; LOE, level of evidence; MLD, mean level difference; SCED, single-case experimental designs.
[a]Average CII score.

post-baseline phase, was selected to assess treatment outcomes across studies because an IRD is sensitive to changes in treatment masked by other standardized effect sizes. Additionally, it is easy to calculate, relatively straightforward to interpret, and has been previously used to analyze treatment effects in single-case research pertaining to SSDs.[31] Unfortunately, the IRD was ultimately abandoned, as the majority of studies did not report enough evidence to calculate this type of effect size. A standardized *effect size* by participant was also ruled out because it can be difficult to interpret as an independent measure of treatment outcomes and cannot be calculated when there is zero variance in the measured data periods.[b21,32] Moreover, effect size has not been consistently reported in the biofeedback literature. Due to these concerns, use of the *mean level difference* was judged to be the best available independent indicator of a change in participants' speech behavior. The mean level difference is defined as the raw difference

[b] *Zero variance can occur in the baseline period of a treatment program where target sound production accuracy is zero or a maintenance period where mastery of a target sound is complete.*

between the mean percentage of items rated correct in maintenance and baseline intervals. For the purposes of this paper, when multiple targets were reported, the target with the largest accuracy gain was selected for inclusion to identify instances where the treatment intensity was paired with a change in behavior (i.e., acquisition or improvement in target speech sound). We recognize that there are clear limitations of using the proposed measure of treatment efficacy, yet significant gaps in the reported data resulted in few options for estimating biofeedback treatment gains across studies. We revisit this issue in the discussion.

RESULTS

The 29 studies that met inclusion criteria consisted of 1 level Ib study, 14 level IIb (single-case experimental designs or SCEDs), and 14 level III (case series/case studies). There was a shift toward higher quality single-case studies over time as well as an increase in the number of studies using visual biofeedback, with all six studies prior to 2010 representing only level III evidence and a preponderance of level IIb evidence between 2011 and 2018 (22 studies). The one level Ib study was from 2016.

Participant profiles included 120 RSE (mean age: 10; 7) and 18 CAS (mean age: 10; 10) participants. A higher proportion of male ($n = 93$) versus female representation ($n = 41$) was observed in the gender distribution data. This finding is in keeping with previous literature indicating a greater prevalence of RSE among males than in females.[33–35] A summary of the data by study with relevant participant characteristics can be found in Table 2.

After coding each of the categories defined previously, we derived a measure of treatment intensity, or CII score, for 114 of 138 participants. Participants were excluded from this subgroup due to missing data in one or more of the composite CII categories. Mean level difference scores were reported for 86 of the participants with a calculated CII score. Unreported baseline or maintenance scores resulted in the smaller sample size for the combined CII and mean level difference subset. Table 3 offers a summary of the review data for the CII and mean level difference scores by research study.

Correlations

Pearson's product–moment correlation coefficients were computed to assess the relationship between the measures of treatment intensity and treatment effects. Correlation coefficients were calculated for CII and mean level difference. Correlation coefficients were also calculated for the component variables of the CII score (dose, dose frequency, treatment duration) and mean level difference.

The strongest correlations observed between the mean level difference scores and predictor variables were CII, dose, and duration. Significant positive correlations between mean level difference and CII ($r = 0.227$, $n = 86$, $p = 0.035$) as well as mean level difference and dose ($r = 0.223$, $n = 86$, $p = 0.039$) indicated weak but statistically significant associations between the effects of treatment and treatment intensity, specifically related to the individual factor of dose (teaching trials per session). Mean level difference score and duration were significantly (negatively) correlated but were small in magnitude ($r = -0.226$, $n = 86$, $p = 0.037$). The correlation between mean level difference and frequency was not significant ($r = 0.164$, $n = 86$, $p = 0.131$). The preliminary nature of the present data precludes further analysis.

DISCUSSION

Overview

The aim of this study was to explore the nature of the relationship between treatment intensity and efficacy of visual biofeedback treatment by conducting a review of the existing biofeedback literature. Several decades of research have documented the use of biofeedback for evaluation and treatment of speech error patterns resulting from a variety of diagnoses including articulation and phonological disorders, cleft palate, apraxia, dysarthria, hearing impairment, Down's syndrome, and cerebral palsy, among others.[18,36–41] The heterogeneous nature of these diagnoses was considered too broad to assess as one body of research; therefore, the present review was limited to research studies involving visual biofeedback for treatment of speech sound errors of unknown or motor-based etiology. Twenty-nine studies, covering three major types of biofeedback, qualified for in-depth review.

Summary of Findings

The findings from this preliminary review of biofeedback intervention revealed that the CII and one of the component parts, dose, showed small but significant positive relationships with treatment efficacy as measured by mean level difference scores. This finding suggests that changes in speech behaviors are influenced by treatment intensity or dosage, with higher dose values providing additional opportunities for practice and, thus, improvement. This is consistent with past research indicating that the motor component of speech production responds well to multiple repetitions of a task to improve performance.[42]

Additionally, the correlation results indicated a negative relationship between mean level difference and duration suggesting that the parameters of dose, dose frequency, and duration may be manipulated quite broadly to facilitate changes in speech behaviors. More clearly, it is possible that a short duration of treatment with a clustered session frequency distribution (i.e., 3 weeks $\times$ 4 sessions per week) might be preferable to a longer duration of treatment with sessions widely distributed (i.e., 12 weeks $\times$ 1 session per week). Taken in context with the reviewed literature, this finding raises the possibility that a short study duration may have an equal or greater impact on treatment progress assuming other parameters (i.e., dose, dose frequency) are increased accordingly to compensate for shorter duration and maintain overall treatment intensity. In the study conducted by Preston and Leece,[26] all four participants demonstrated measurable and statistically significant gains when treated twice per day for a period of 1 week at a high dose (M dose = 366; SD = 165). Adjusting the variables of duration and frequency alters the schedule of treatment, offering different practice schedules, such as mass versus distributed practice of a target speech task. According to Maas et al,[10] the optimal practice schedule may depend on the nature of the motor speech disorder. Manipulating the variables of intensity to facilitate best outcomes recognizes that an effective intensity level for one type of treatment may not be equally as effective for another type of treatment.[24] Thus, the need for further investigation of intensity measures is justified by the presence of both responders and nonresponders reported across the analyzed studies of different treatment intensities and the knowledge that optimal intensity may vary with disorder type. Additionally, the current review demonstrates that the parameters of treatment intensity and outcomes are not systematically reported in such a way that makes it easy to draw firm conclusions about relationships between them.

Limitations

Levels of evidence. The studies examined were primarily divided between level III evidence and level IIb evidence, although a promising trend toward more rigorous research designs was observed. This upward shift seems to coincide with the development of ASHA guidelines[33] recommending the use of evidence-based practice (EBP) in clinical treatment. Even though the published literature does not yet include large-scale, peer-reviewed studies using visual biofeedback for treatment of SSDs, several of the reviewed studies employed sound methodological practices (e.g., single-case experimental designs). It should be noted, however, that this review included a disproportionate number of studies from a small pool of authors, which may reduce the overall strength of our analysis. Although visual biofeedback is currently a developing research area, we expect that future studies will demonstrate increasing quality, more rigorous study designs, and a larger set of researchers.

Lack of consistency in reporting across studies. The present review revealed a surprising lack of unity when reporting parameters of treatment intensity and outcomes. Approximately one-third of the reviewed studies were missing at least one of the defined CII components (dose, dose frequency, and duration), complicating the data coding. In addition, several authors reported session length in lieu of the number of teaching episodes per session. Adopting a standard set of parameters for reporting intensity, such as the CII, would move the state of the literature forward in a clinically meaningful way.

The parameter most frequently unreported was dose, identified in the current study as being a factor of interest when studying treatment intensity. The current findings provide preliminary evidence that this variable is potentially meaningful for overall efficacy, but more work needs to be done to evaluate the strength of this relationship. Research to explore the proposed explanation for the negative durational relationship and the lack of a correlation between frequency and efficacy would also be beneficial.

Measures of treatment efficacy were also found to be inconsistently reported across studies. Reporting on treatment outcomes ranged from inclusion of (1) a posttreatment measure at the spontaneous word level, (2) baseline and posttreatment measures of spontaneous speech, (3) a baseline and posttreatment untreated probe score, (4) mean level difference scores for untreated probes, (5) mean-level difference and effect sizes of untreated probes, and (6) mean level differences for treated and untreated probes. It should be acknowledged that the previously identified shift toward studies with higher levels of evidence should resolve some of these inconsistencies assuming certain minimum standards of reporting outcomes are established/met in future research.

Also noteworthy is the confounding variable of multiple treatment targets in a treatment session. The current study reported only the gains of the most improved treatment target per participant. This strategy was based on the rationale that if gains were observed, it would be most meaningful to identify the treatment intensity of targets acquired rather than the treatment intensity of targets not acquired. However, in some instances, reported measures of intensity were not specific to a target, making it impossible to determine dose per target. The importance of reporting target-specific intensity measures cannot be understated if the ultimate goal is to use such findings to guide clinical practice.

Future Recommendations

The current data included multiple biofeedback types, such that ultrasound, visual-acoustic (LPC and spectrographic), and EPG measures were treated as one biofeedback method. The diversity of these intervention types is significant considering that ultrasound and EPG offer the ability to view articulator movements concealed within the oral/pharyngeal cavity, while visual acoustic displays show a graphic representation of an acoustic speech signal. Presently, it is not possible to determine if the intensity/efficacy relationship identified here is consistent across all methods or more strongly related to a particular type of biofeedback and/or speech sound error type. For example, the least optimal biofeedback type for rhotic errors may be EPG, which shows information about lateral bracing of the tongue but yields little information about other more salient features of rhotic articulation.[21] Given the preliminary nature of this review and the disproportionate number of ultrasound participants compared with visual-acoustic and EPG, identification of an interaction suggesting a preferred biofeedback type was considered premature.

The calculation of treatment outcomes presented several challenges in the current investigation. Significant gaps and inconsistencies in the reported data resulted in few options for estimating the efficacy of biofeedback treatment across studies. As a result, we chose to use a mean level difference score for reasons previously cited. However, it is possible that the mean level difference accuracy score does not adequately reflect the degree of change in participant behavior. Assessments conducted only in the pre- and posttreatment periods and, more importantly, only on untreated targets could mask the early acquisition phase of a speech sound. Several researchers have indicated that the addition of the visual modality in biofeedback seems to be most effective for the learner during this early acquisition period,[2,20] which suggests that tracking accuracy improvement rates for within-treatment trials is particularly relevant to biofeedback research. In future investigations of treatment intensity, a more reliable measure of a change in participant speech behavior, such as an IRD, needs to be clearly established and validated as a measure of treatment efficacy.

Finally, the ultimate goal of an intervention program is generalization of the target speech behavior to naturalistic settings. This type of improvement is observed when a treated target is mastered and spontaneously produced in novel utterances and conversational speech. In the present research, generalization to spontaneous speech was evaluated by 31% of the studies, and all of these studies were found to be level III evidence, primarily case studies and case series. Although a relatively high percentage of these participants were reported to achieve generalization to spontaneous speech, the strength of the conclusions that can be drawn from the results is intrinsically limited by the use of a single-subject case study or case series study design. The remaining 69% of the studies reported posttherapy probe measures in structured contexts (14 level IIb, 4 level III, and 1 level Ib studies). Regardless of the level of evidence, using probe measures as the only method of assessing generalization does not differentiate between gains observed in elicited versus spontaneous contexts, a fact which merits consideration in future research.

Clinical Relevance

A primary goal of both the researcher and clinician is to identify and implement best practice procedures. To identify best practice for a population such as SSD, it is necessary to assess proposed treatment variables in an organized and systematic fashion. However, it is a reality that many speech–language pathologists encounter overwhelming caseloads and have little time to review and compare evidence for emerging treatment programs. Given limited time and resources, identifying, implementing, and tracking measures of treatment (e.g., dosage, practice conditions) supported by EBP can be challenging for any dedicated clinician; yet, it is a vital part of the clinical decision-making process.

A composite score such as the CII offers clinicians the ability to effectively quantify an overall treatment intensity as well as the individual components of dose, dose frequency, and treatment duration. Using the CII, clinicians can directly investigate its effect on treatment outcomes, a concept that has previously been suggested as a necessary direction for speech and language research.[24] In the present study, we reviewed biofeedback treatment both as a preliminary review of the literature summarizing the existing evidence for biofeedback intervention (see previous review) and as an example of how clinicians can use CII in the clinical decision-making process.

Typically, clinicians select an appropriate intervention based on clinical education about the efficacy of the treatment for the communication disorder of the individual while simultaneously considering other variables that may facilitate or impede treatment outcomes.[43] Of these variables, service delivery, and more specifically treatment intensity, is often defined as the frequency and duration of treatment. The additional consideration of dose, a less commonly reported factor of treatment intensity, may help clinicians explain variations in client outcomes for a given treatment program administered with different intensities. Allen,[12] investigating treatment intensity and a multiple oppositions approach for SSD, reported that increasing frequency of intervention yielded greater improvements for participants when the overall CII remained the same. As noted previously, Preston and colleagues[26] significantly altered variables of treatment intensity compared with other biofeedback studies and reported positive treatment outcomes. These findings align with the premise that increasing frequency of intervention associated with high overall intensity will lead to improvement of a desired behavior.[44] However, Baker and McLeod[45] reported that many SSD studies document measures of dose frequency and treatment duration; yet, very few reported dose, a finding also noted in the current study, and one that makes it difficult for clinicians to replicate treatment programs from studies with positive outcomes. Our preliminary findings suggest that planning, implementing, tracking, and adjusting dose, as well as frequency and duration, may help improve treatment efficacy.

Our findings also highlight the need to establish a unified measure of treatment efficacy. We recognize that several past researchers have used effect size as a measure of treatment outcomes. However, an effect size cannot be calculated when there is no change in a measured period of time (i.e., maintenance period where mastery of a target sound is complete), making it difficult to use as the primary measure of treatment outcomes, nor is it easily calculated by clinicians. We used the MLD score to measure treatment outcomes, a measure which is easily calculated but does not calculate within-treatment accuracy gains. Other measures, such as an IRD, seem promising for future use but need further research to determine overall clinical utility. Until more definitive treatment intensity results are available in the literature, we suggest that clinicians track pre-, within-, and posttreatment accuracy for treated and untreated targets, and correlate these outcomes with each component of treatment intensity.

In sum, we recognize that the current review only represents a specific subset of the SSD literature. It is also clear that all individuals with SSD do not have a motor-based component and, thus, may not be appropriate candidates for visual biofeedback treatment. At the same time, we hope that the method employed to assess treatment intensity and outcomes provides clinicians with a developing framework to systematically review other studies in a controlled fashion. The comparability of interventions based on treatment intensity correlated with overall treatment outcomes may help direct future treatments to a "best fit" model for clients, regardless of underlying etiology.

CONCLUSIONS

It is clear that we are at the forefront of a rapidly developing area of research in SSD, one that extends well beyond the subset of data evaluated in the current work. As interest in visual biofeedback grows and lowering equipment costs support increased access to practicing clinicians, clear guidelines for treatment are imperative. The present review underscores the need for more systematic language and reporting methods when exploring the relationship between biofeedback intervention for SSDs and factors of treatment intensity.

DISCLOSURES

Elaine R. Hitchcock is a recipient of a separately budgeted research grant from Montclair State University. She has no relevant nonfinancial relationships.
Michelle T. Swartz has no relevant financial or nonfinancial relationships to disclose.
Melissa B. Lopez has no relevant financial or nonfinancial relationships to disclose.

CONFLICT OF INTEREST

None.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the following individuals: Jennifer Caceres, Carly Truglio, and Grace Lazarus, graduate student research assistants. This research was supported by a separately budgeted research grant from Montclair State University.

REFERENCES

1. Shuster L, Ruscello D, Toth A. The use of visual feedback to elicit correct /r/. Am J Speech Lang Pathol 1995;4(02):37–44
2. Gibbon F, Paterson L. A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. Child Lang Teach Ther 2006;22(03):275–292
3. Adler-Bock M, Bernhardt BM, Gick B, Bacsfalvi P. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. Am J Speech Lang Pathol 2007;16(02):128–139
4. Bacsfalvi P, Bernhardt B, Gick B. Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. Adv Speech Lang Pathol 2007;9(01):36–45
5. Byun TM, Hitchcock ER. Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. Am J Speech Lang Pathol 2012;21(03):207–221
6. Preston JL, Brick N, Landi N. Ultrasound biofeedback treatment for persisting childhood apraxia of speech. Am J Speech Lang Pathol 2013;22(04): 627–643
7. Byun TM, Hitchcock ER, Swartz MT. Retroflex versus bunched in treatment for rhotic misarticulation: evidence from ultrasound biofeedback intervention. J Speech Lang Hear Res 2014;57(06):2116–2130
8. Preston JL, McCabe P, Rivera-Campos A, Whittle JL, Landry E, Maas E. Ultrasound visual feedback treatment and practice variability for residual speech sound errors. J Speech Lang Hear Res 2014;57(06):2102–2115
9. McAllister Byun T, Campbell H. Differential effects of visual-acoustic biofeedback intervention for residual speech errors. Front Hum Neurosci 2016;10(567):567
10. Maas E, Robin DA, Austermann Hula SN, et al. Principles of motor learning in treatment of motor speech disorders. Am J Speech Lang Pathol 2008; 17(03):277–298
11. Austermann Hula SN, Robin DA, Maas E, Ballard KJ, Schmidt RA. Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. J Speech Lang Hear Res 2008;51(05):1088–1113

12. Allen MM. Intervention efficacy and intensity for children with speech sound disorder. J Speech Lang Hear Res 2013;56(03):865–877
13. Ruscello DM. Visual feedback in treatment of residual phonological disorders. J Commun Disord 1995;28(04):279–302
14. Ballard KJ, Robin DA, McCabe P, McDonald J. A treatment for dysprosody in childhood apraxia of speech. J Speech Lang Hear Res 2010;53(05): 1227–1245
15. McCormack J, McLeod S, McAllister L, Harrison L. A systematic review of the association between childhood speech impairment and participation across the lifespan. Int J Speech-Language Pathol 2009;11(02):155–170
16. Hitchcock ER, Harel D, Byun TM. Social, emotional, and academic impact of residual speech errors in school-aged children: a survey study. Semin Speech Lang 2015;36(04):283–294
17. Volin R. A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task. Am J Speech Lang Pathol 1998;7(01):81–90
18. McAllister Byun T, Swartz MT, Halpin PF, Szeredi D, Maas E. Direction of attentional focus in biofeedback treatment for /r/ misarticulation. Int J Lang Commun Disord 2016;51(04):384–401
19. Bernhardt MB, Bacsfalvi P, Adler-Bock M, et al. Ultrasound as visual feedback in speech habilitation: exploring consultative use in rural British Columbia, Canada. Clin Linguist Phon 2008;22 (02):149–162
20. Fletcher SG, Dagenais PA, Critz-Crosby P. Teaching consonants to profoundly hearing-impaired speakers using palatometry. J Speech Hear Res 1991;34(04):929–942
21. Hitchcock ER, Byun TM, Swartz M, Lazarus R. Efficacy of electropalatography for treating misarticulation of /r/. Am J Speech Lang Pathol 2017;26 (04):1141–1158
22. Mental R, Carey H, Schreiber N, Barnes A, Lee G, Vick J. Changes in lateralized /s/ production after treatment with opti-speech visual biofeedback. J Acoust Soc Am 2016;140(04):3442
23. Vick J, Mental R, Carey H, Schreiber N, Barnes A, Lee G. Plasticity of the internal model for speech sounds: articulatory changes during intensive visual biofeedback treatment. J Acoust Soc Am 2016;140 (04):3226
24. Warren SF, Fey ME, Yoder PJ. Differential treatment intensity research: a missing link to creating optimally effective communication interventions. Ment Retard Dev Disabil Res Rev 2007;13(01): 70–77
25. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008;51(01):S225–S239
26. Preston JL, Leece MC. Intensive treatment for persisting rhotic distortions: a case series. Am J Speech Lang Pathol 2017;26(04):1066–1079
27. Preston JL, Leece MC, Maas E. Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. Front Hum Neurosci 2016;10(440):440
28. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62(10):1006–1012
29. Reuters T. Endnote X7. Philadelphia, PA: Thomson Reuters; 2013
30. American Speech-Language-Hearing Association. Evidence-Based Practice in Communication Disorders: In: Communication Disorders: An Introduction. Rockville: American Speech-Language-Hearing Association; 2004
31. Murray E, McCabe P, Ballard KJ. A systematic review of treatment outcomes for children with childhood apraxia of speech. Am J Speech Lang Pathol 2014;23(03):486–504
32. Parker R, Vannest K, Brown L. The improvement rate difference for single-case research. Except Child 2009;75(02):135–150
33. National Institute of Health on Deafness and Other Communication Disorders. Apraxia of Speech. Bethesda: U.S. Department of Health and Human Services-National Institutes of Health; 2016. Available at: https://www.nidcd.nih.gov/sites/default/files/Documents/health/voice/Apraxia-508.pdf. Accessed January 2, 2019
34. National Institute on Deafness and Other Communication Disorders (NIDCD). National Institutes of Health (NIH). Available at: https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-institute-deafness-other-communication-disorders-nidcd. Published 2018. Accessed December 10, 2018
35. Shriberg L, Potter N, Strand E. Childhood apraxia of speech in children and adolescents with galactosemia. Presentation presented at the 2009 American Speech-Language-Hearing Association National Convention, New Orleans, LA
36. Bernhardt B, Gick B, Bacsfalvi P, Ashdown J. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. Clin Linguist Phon 2003;17(03):199–216
37. Carter P, Edwards S. EPG therapy for children with long-standing speech disorders: predictions and outcomes. Clin Linguist Phon 2004;18(6-8):359–372
38. Cleland J, Timmins C, Wood SE, Hardcastle WJ, Wishart JG. Electropalatographic therapy for children and young people with Down's syndrome. Clin Linguist Phon 2009;23(12):926–939
39. Dent H, Gibbon F, Hardcastle W. Inhibiting an abnormal lingual pattern in a cleft palate child using

electropalatography. In: Leahy M, Kallen J, ed. Interdisciplinary Perspectives in Speech and Language Pathology. Dublin: Trinity College; 1992: 211–221

40. Hardcastle WJ, Gibbon FE, Jones W. Visual display of tongue-palate contact: electropalatography in the assessment and remediation of speech disorders. Br J Disord Commun 1991;26(01):41–74
41. Preston JL, Maas E, Whittle J, Leece MC, McCabe P. Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. Clin Linguist Phon 2016;30(3-5):363–381
42. Schmidt R, Lee T. Motor Control and Learning: A Behavioral Emphasis. 4th ed. Champaign, IL: Human Kinetics; 2005
43. Schooling T, Venediktov R, Leech H. Evidence-Based Systematic Review: Effects of Service Delivery on The Speech and Language Skills of Children from Birth to 5 Years of Age. Rockville: American Speech-Language-Hearing Association; 2010. Available at: https://www.asha.org/uploadedfiles/ebsr-service-delivery.pdf. Accessed January 2, 2019
44. Duhon G, Mesmer E, Atkins M, Greguson L, Olinger E. Quantifying intervention intensity: a systematic approach to evaluating student response to increasing intervention frequency. J Behav Educ 2009;18(02):101–118
45. Baker E, McLeod S. Evidence-based practice for children with speech sound disorders: part 1 narrative review. Lang Speech Hear Serv Sch 2011;42(02):102–139
46. Dagenais P, Critz-Crosby P, Adams J. Defining and remediating persistent lateral lisps in children using electropalatography. Am J Speech Lang Pathol 1994;3(03):67–76
47. Gibbon F, Dent H, Hardcastle W. Diagnosis and therapy of abnormal alveolar stops in a speech-disordered child using electropalatography. Clin Linguist Phon 1993;7(04):247–267
48. Gibbon F, Hardcastle W, Moore A. Modifying abnormal tongue patterns in an older child using electropalatography. Child Lang Teach Ther 1990; 6(03):227–244
49. Schmidt A. Evaluating a new clinical palatometry system. Adv Speech Lang Pathol 2007; 9:73–81
50. Bressmann T, Harper S, Zhylich I, Kulkarni GV. Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors. Clin Linguist Phon 2016;30(3-5):345–362
51. Cleland J, Scobbie JM, Wrench AA. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. Clin Linguist Phon 2015; 29(8-10):575–597
52. Heng Q, McCabe P, Clarke J, Preston JL. Using ultrasound visual feedback to remediate velar fronting in preschool children: a pilot study. Clin Linguist Phon 2016;30(3-5):382–397
53. Hitchcock ER, Byun TM. Enhancing generalisation in biofeedback intervention using the challenge point framework: a case study. Clin Linguist Phon 2015;29(01):59–75
54. Lee S, Wrench A, Sancibrian S. How to get started with ultrasound technology for treatment of speech sound disorders. Perspect Speech Sci Orofac Disord 2015;25(02):66–80
55. Modha G, Bernhardt BM, Church R, Bacsfalvi P. Case study using ultrasound to treat /r/. Int J Lang Commun Disord 2008;43(03):323–329
56. Preston JL, Leece MC, Maas E. Motor-based treatment with and without ultrasound feedback for residual speech-sound errors. Int J Lang Commun Disord 2017;52(01):80–94
57. Preston JL, Leece MC, McNamara K, Maas E. Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: a single case experimental study. Am J Speech Lang Pathol 2017;26(03):840–852
58. Preston JL, McAllister T, Phillips E, et al. Treatment for residual rhotic errors with high- and low-frequency ultrasound visual feedback: a single-case experimental design. J Speech Lang Hear Res 2018;61(08):1875–1892
59. Shawker TH, Sonies BC. Ultrasound biofeedback for speech training. Instrumentation and preliminary results. Invest Radiol 1985;20(01):90–93
60. Sjolie GM, Leece MC, Preston JL. Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. J Commun Disord 2016;64:62–77
61. McAllister Byun T. Efficacy of visual–acoustic biofeedback intervention for residual rhotic errors: a single-subject randomization study. J Speech Lang Hear Res 2017;60(05):1175–1193
62. Byun TM, Campbell H, Carey H, Liang W, Park TH, Svirsky M. Enhancing intervention for residual rhotic errors via app-delivered biofeedback: a case study. J Speech Lang Hear Res 2017;60(6S):1810–1817