ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Rhotic Generalization Is More Rapid in Biofeedback Than Motor-Based Treatment for Residual Speech Sound Disorder: Secondary Outcomes of a Randomized Controlled Trial

Jonathan L. Preston,[a] Elaine R. Hitchcock,[b] Megan C. Leece,[a] Nina R. Benway,[a,c] Jennifer Hill,[d] and Tara McAllister[e]

[a] Department of Communication Sciences and Disorders, Syracuse University, NY [b] Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, NJ [c] Department of Electrical and Computer Engineering, University of Maryland, College Park [d] Department of Applied Statistics, Social Science, and Humanities, New York University, NY [e] Department of Communicative Sciences and Disorders, New York University, NY

ARTICLE INFO

Article History:
Received September 3, 2025
Revision received October 27, 2025
Accepted November 24, 2025

Editor-in-Chief: Jessica E. Huber
Editor: Breanna Krueger

https://doi.org/10.1044/2025_JSLHR-25-00700

ABSTRACT

Purpose: This study examined the effects of visual biofeedback approaches and nonbiofeedback motor-based treatment on generalization outcomes following speech therapy for children with residual speech sound disorders (RSSDs).
Method: A total of 108 children aged 9–15 years with RSSD affecting American English /ɹ/ were randomly assigned to receive 19 motor-based speech treatment sessions, with or without visual biofeedback (divided into ultrasound or visual-acoustic biofeedback). The treatment included practice designed to implement several motor learning principles, with task difficulty systematically adjusted based on the child's performance. Children's /ɹ/ accuracy on untreated words before and after treatment was rated as correct or incorrect by lay listeners who were blinded to participant characteristics, treatment conditions, and time points.
Results: The mixed-effects regression model revealed a statistically significant interaction between treatment type and time point. Specifically, both the biofeedback and nonbiofeedback motor-based treatment groups made progress over time, but the amount of generalization to untreated words was significantly greater in the biofeedback condition than in the motor-based treatment. In a subanalysis comparing biofeedback types, greater generalization was observed following ultrasound biofeedback than visual-acoustic biofeedback, although this effect was strongest at one treatment site.
Discussion: This randomized controlled trial found that adding biofeedback to motor-based treatment can increase the rate of accurate production of the American English /ɹ/ in untreated words.

Correspondence to Jonathan L. Preston: jopresto@syr.edu. ***Disclosure:*** *Tara McAllister is a member of Sonority Labs LLC, a for-profit entity that was created to explore the possibility of commercializing staRt, an iOS/web application for visual-acoustic biofeedback. The staRt app is currently distributed as a free download, but it may be sold or licensed for for-profit capacity in the future. The staRt app was not used in the present study, but the findings on the broader efficacy of biofeedback are relevant to its development and potential commercialization. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Residual speech sound disorder (RSSD) involves the distorted production of speech sounds that persist beyond approximately 8 years of age (e.g., Flipsen, 2015; Shriberg et al., 2010). These unresolved speech deviations can be associated with reduced speech intelligibility (Cronin et al., 2014) as well as negative social, emotional, and educational consequences (Hitchcock et al., 2015; Wren et al., 2021), warranting clinical management. However, RSSDs sometimes do not resolve despite interventions (Gray & Shelton, 1992; Lewis et al., 2019; McAllister Byun &

 • Copyright © 2026 American Speech-Language-Hearing Association

Hitchcock, 2012). Therefore, there is an important clinical need to identify interventions that can effectively accelerate improvements in speech sound learning. While treatments rooted in the principles of motor learning have been found to be efficacious for some children, visual biofeedback enhancements to motor-based treatments (MBTs) have the potential to facilitate the initial acquisition (McAllister et al., 2026) and longer-term learning (Preston et al., 2019) of speech movements.

In American English, the rhotic consonant /ɹ/ and rhotacized vowels /ɝ/ and /ɚ/ are among the most common sounds affected by RSSD (Shriberg, 2009). Rhotics tend to emerge relatively late in speech development (Crowe & McLeod, 2020) and have complex articulatory configurations. Specifically, rhotic production is characterized by an anterior lingual constriction near the palate, a posterior lingual constriction formed by tongue root retraction, bracing of the lateral margins of the tongue against the back molars, and slight lip rounding (Preston, Benway, et al., 2020). Distorted or derhotic productions tend to lack one or more articulatory elements. This articulatory configuration contributes to the acoustically distinct formant structure for rhotics, defined by a low third formant (F3) that is very close to a high second formant (F2), resulting in a small F3–F2 distance (Campbell et al., 2018; Espy-Wilson et al., 2000; Harper et al., 2020). Consequently, distorted or derhotic production commonly involves a higher F3 and larger F3–F2 distance than fully rhotic production.

The existing standard of care for RSSD can be broadly described as an MBT designed to integrate several principles of motor learning (Maas et al., 2008). While clinicians likely vary in how they implement MBT, the general approach includes a clinician model and verbal descriptions of articulator placement for sound, with repetitive practice and feedback (Preston, Benway, et al., 2020). Although some children respond well to MBT, others fail to show significant improvement, and there have long been calls for more effective intervention approaches (Ruscello, 1995). Supplementing speech practice with visual biofeedback has emerged as an influential factor in MBT (Hitchcock et al., 2019, 2023; McAllister Byun, 2017; McAllister Byun & Hitchcock, 2012). Among the different types of biofeedback, visual-acoustic biofeedback offers a visual representation of the acoustic features of speech (e.g., formant structure), and ultrasound biofeedback offers real-time feedback of the tongue shape during speech production. These technologies provide a new visual modality that may enhance clinician cueing and opportunities for client self-evaluation, especially for speech sounds that can be challenging to describe verbally. See Figure 1 from the companion paper (McAllister et al., 2026) for example images of ultrasound and visual-acoustic biofeedback displays.

Visual-acoustic biofeedback converts the acoustic speech signal into a visual representation, which can help more clearly define the acoustic goals of /ɹ/ for learners, which some children with RSSD affecting /ɹ/ may not readily perceive (Cabbage & Carrell, 2014; Sénéchal et al., 2004; Shuster, 1998). Specifically, for children whose /ɹ/ distortion involves a wide distance between F3 and F2, a visual display can be used to encourage them to modify their articulation to try to achieve a lower F3 (and subsequently reduce the F3–F2 distance). This representation of formants is often achieved using a linear predictive coding (LPC) display, which displays the resonant frequencies of the vocal tract as peaks in a wavelike shape. Several small-scale studies, including case series and single-case experimental designs, have found that treatment incorporating visual-acoustic biofeedback can facilitate improvements in /ɹ/ productions for individuals with RSSD (McAllister Byun, 2017; McAllister Byun & Campbell, 2016; Mcallister Byun et al., 2017; Ochs et al., 2023; Shuster et al., 1995). However, prior to the randomized controlled trial described here, no well-controlled randomized group study has directly compared treatment with and without visual-acoustic biofeedback, which is necessary to establish a robust evidence base (Robey, 2004).

Ultrasound biofeedback offers a potentially different path toward the remediation of RSSDs. Specifically, it provides a real-time display of the midsagittal or coronal lingual articulation patterns that are generally hidden in the mouth. As such, visual display of the tongue can be used to enhance cues to elevate the anterior tongue, lower the posterior tongue, retract the tongue root, or raise the lateral margins of the tongue. Ultrasound biofeedback has been used to remediate a number of lingual speech deviations, with a particularly large evidence base of case studies and single-case experimental designs supporting its use for addressing /ɹ/ (Adler-Bock et al., 2007; Bacsfalvi, 2010; Bernhardt et al., 2005; McAllister et al., 2022; McAllister Byun et al., 2014; Preston et al., 2014, 2018, 2019; Preston, Hitchcock, & Leece, 2020; Preston & Leece, 2017; Preston, Leece, & Maas, 2017; Raaz et al., 2021; Sjolie et al., 2016). A systematic review by Sugden et al. (2019) characterized ultrasound biofeedback as a promising intervention approach for RSSD and other clinical populations but also called for larger clinical trials.

Overall, while both biofeedback types have shown the potential to enhance clinical outcomes for RSSD, prior research is limited by small sample sizes and single-case designs that cannot fully determine the added benefit of biofeedback beyond MBT. The study reported here and in a companion paper (McAllister et al., 2026) was intended to address this limitation through a randomized controlled trial, which is generally considered the standard for evidence-based decision making (Dollaghan, 2007).

It is important to consider that the mechanisms by which ultrasound and visual-acoustic biofeedback enhance speech sound learning may differ. It is well established that speech production goals are informed by both auditory and somatosensory representations (Guenther & Vladusich, 2012; Tourville & Guenther, 2011). Visual-acoustic biofeedback provides learners with visual information about the acoustic characteristics of speech, which may primarily enhance the learner's underlying auditory representations for /ɹ/. However, ultrasound biofeedback of the tongue may primarily enhance learners' somatosensory knowledge of articulatory positioning and movement for /ɹ/ production. Thus, although both biofeedback approaches add a real-time visual component to enhance learning for /ɹ/, different learning pathways may be involved (Benway et al., 2021; McAllister et al., 2026). Because both biofeedback approaches have empirical and theoretical support, they are included in this study.

Finally, when comparing biofeedback and non-biofeedback MBT, the context in which speech production is measured has both experimental and clinical importance. Within the broad framework of motor learning, skill *acquisition* may be reflected as improved speech production during practice, which contrasts with generalized *learning* that can be seen outside of practice (Maas et al., 2008). For example, skill acquisition may be evidenced by correct productions of /ɹ/ on practiced words during speech therapy sessions when a client is provided with feedback and a supportive treatment structure. Biofeedback can be characterized as providing detailed feedback about a production (knowledge of performance or KP feedback); KP feedback is thought to enhance performance in the acquisition phase, but may limit generalization (Hodges & Franks, 2001; Maas et al., 2008). In line with theoretical predictions derived from principles of motor learning, multiple studies have demonstrated that biofeedback facilitates acquisition of /ɹ/ for children with RSSD during treatment sessions (McAllister et al., 2026; McAllister Byun & Hitchcock, 2012; Preston, Leece, & Maas, 2017). In particular, in a companion paper reporting the primary outcome of the randomized controlled trial presented here (McAllister et al., 2026), children's acoustically measured progress in an early stage of treatment for RSSD was found to be 2.4 times faster with biofeedback than with MBT alone.

However, the long-term goal of intervention is not skill acquisition but *motor learning*, which involves the generalization or transfer of new motor plans beyond the immediate training context, including unpracticed situations. To fully understand the benefits of biofeedback, it is important to compare MBT and biofeedback treatments in terms of speech motor learning outcomes in the absence of clinician support. Although there is theoretical tension between elements of treatment that facilitate motor acquisition versus motor learning (Maas et al., 2008), there is preliminary evidence that adding biofeedback to motor-based /ɹ/ treatment may also facilitate generalization (Preston et al., 2019). It is possible that biofeedback could boost the acquisition of speech sounds in a manner that subsequently leads to an increase in generalization to untreated words. This finding also supports the clinical utility of biofeedback approaches.

Purpose and Hypothesis

This study aimed to evaluate generalization learning in children with RSSD affecting /ɹ/ after a period of speech therapy with MBT or biofeedback. This study is part of a larger project, the multisite C-RESULTS RCT (Correcting Residual Errors With Spectral, ULtrasound, Traditional Speech therapy Randomized Controlled Trial). In this project, all participants received a course of structured treatment that was randomly assigned to feature MBT or biofeedback conditions. Our companion paper describes the primary outcome of the C-RESULTS RCT, a comparison of acoustic improvements within treatment sessions to characterize the acquisition of /ɹ/ in MBT and biofeedback conditions. Acquisition was measured as our primary outcome based on theoretical predictions of the principles of motor learning. In the present study examining the secondary outcome, we shifted our focus from theory to the clinically important question of generalization to untreated words. Our preregistered hypothesis for the secondary outcome was that when the total number of treatment sessions was held constant, the group of children who received biofeedback treatment would demonstrate greater generalization of rhotics than the group of children who received MBT without biofeedback (McAllister et al., 2020). Additionally, because participants receiving biofeedback were subrandomized to receive either ultrasound or visual-acoustic biofeedback, we explored whether there were differences in generalization of /ɹ/ for the two biofeedback types; however, there was not a clear theoretical reason to expect one type of biofeedback to be more effective than the other on average.

Method

Study procedures were approved by the Biomedical Research Association of New York (protocol #18–10-393). A parent/guardian provided written consent and the child participants provided written assent. Data were collected from three sites: New York University (NYU), Montclair State University (MSU), and Syracuse University (SU). The New York University site was initially

intended to provide data processing and statistical analysis; it was only added as a treatment site in the third year of the project to offset the impact of the COVID-19 pandemic on recruitment. Details of the study procedures are reported in the study registration (McAllister et al., 2020) and in a companion paper on the primary outcome (McAllister et al., 2026). The major elements of the participants, study design, data coding, and analysis are summarized below.

Participants

Participants were referred by local speech-language pathologists (SLPs) or recruited through public flyers. A phone screening with parents/guardians of the potential participant verified initial inclusionary criteria: ages of 9;0 to 15;11 (years;months); English as a dominant or equally dominant language; started learning English at no later than 3 years of age; were acquiring a rhotic variety of English, per parent report; and reported difficulty with /ɹ/ production.

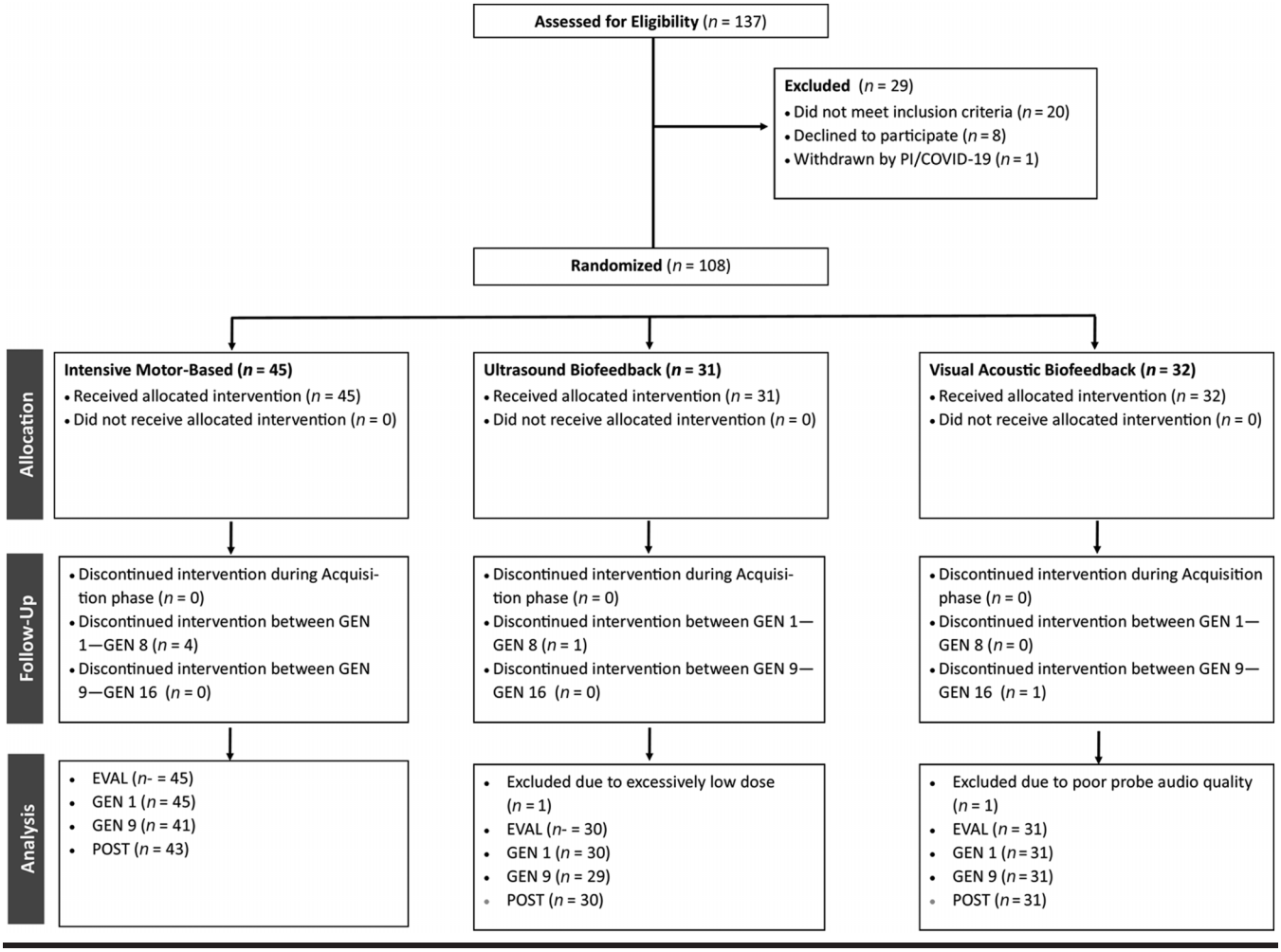
Following the initial screening, 137 participants completed an in-person evaluation session to verify their eligibility. Inclusionary criteria at this session included: passing a pure-tone hearing screening at 20 dB HL; passing a brief examination of oral structure and function; achieving a passing score on the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5) screening measure, or a score of at least 80 on the CELF-5 Core Language Index (Wiig et al., 2013); scoring no more than 30% correct on a 50-item probe of /ɹ/ in single words per clinician rating; and scoring at or below the eighth percentile on the Goldman-Fristoe Test of Articulation–Third Edition (GFTA-3; Goldman & Fristoe, 2015). Finally, to screen out childhood apraxia of speech (CAS), the Syllable Repetition Task (Shriberg et al., 2012) and the Apraxia Screening task of the LinguiSystems Articulation Test (Bowers & Huisingh, 2011) were administered. Participants whose scores exceeded a cutoff suggestive of CAS on both measures were excluded. Participants who passed one CAS screening task and not the other (n = 3) completed a third task as a tie-breaking measure: a maximum performance task assessing sustained phonemes and repeated single-syllable and trisyllable sequences (Thoonen et al., 1996, 1999). Participants were included in the study if they showed no signs of CAS on two of the measures. A total of 108 participants met eligibility criteria for the study. As reported in our companion paper, the MBT and biofeedback groups did not differ in mean age, GFTA-3 score, raw score on the CELF Screener, and percent of /ɹ/ sounds rated correct on the 50-word probe administered at the pretreatment time point (see Table 1 from McAllister et al., 2026).

Randomization

Participants who remained eligible after the evaluation completed a Dynamic Assessment session (Phase 0). This 2-hr session of nonbiofeedback treatment was intended to facilitate stimulability for /ɹ/ and evaluate initial response to MBT so that allocation to the experimental groups could be balanced with respect to this factor. The Dynamic Assessment session began with approximately 25 min of instruction on the articulation of /ɹ/, followed by 15–20 min of unstructured prepractice with individualized cueing and feedback, then approximately 45 min of structured practice (with a goal of producing 200 syllables). After the Dynamic Assessment session, participants were categorized as either high initial responders (> 5% /ɹ/ accuracy on syllables during structured practice trials in the Dynamic Assessment session) or low initial responders (0%–5% /ɹ/ accuracy). Randomization of participants into treatment conditions was therefore stratified by response group (high vs. low initial responders) as well as by research site, using concealed envelopes. For each response type and location, participants were block randomized in batches of 10, with an allocation ratio of 3 biofeedback to 2 MBT. Within the biofeedback group, the participants were randomized to receive ultrasound or visual-acoustic biofeedback. Thus, each batch of 10 envelopes contained three allocations for the ultrasound condition, three for the visual-acoustic treatment condition, and four for the MBT condition.

Following randomization, participants were assigned to complete three treatment sessions in Phase 1 (Acquisition) and 16 treatment sessions in Phase 2 (Generalization, defined in more detail below). For the purposes of this study, we focused on children's speech outcomes following the Generalization phase of treatment. Figure 1 shows the flow of participants through the study, following the Consolidated Standards of Reporting Trials guidelines (Schulz et al., 2010). Of the 108 participants who met all eligibility requirements, all 108 completed treatment Phase 0 (Dynamic Assessment) and Phase 1 (Acquisition). Six participants withdrew at some point during Phase 2 (Generalization): four from the MBT condition, one from the ultrasound biofeedback condition, and one from the visual-acoustic biofeedback condition. Participants who withdrew from treatment were invited to complete a posttreatment evaluation despite not completing the full treatment program. The posttreatment word probe used to evaluate generalization gains in the present study was collected from all participants, except for two individuals in the MBT condition. However, one participant in the visual-acoustic biofeedback condition was excluded from the analysis owing to multiple probes with corrupted audio recording quality, and one ultrasound biofeedback

Figure 1. Consolidated Standards of Reporting Trials diagram of participant flow. EVAL = evaluation; GEN = Generalization treatment; POST = posttreatment; PI = Principal Investigator.



participant was excluded because the treatment dosage was below the minimum threshold of 3,800 practice trials (minimum 600 Phase 1 and 3,200 Phase 2). This left 104 participants for whom pretreatment and posttreatment data were analyzed (43 from the MBT condition, 30 from the ultrasound biofeedback condition, and 31 from the visual-acoustic biofeedback condition).

Treatment Condition Similarities

Individual treatment sessions were delivered by certified SLPs trained in study procedures, as described by McAllister et al. (2020). Participants in all treatment conditions completed the Acquisition and a Generalization phase of treatment. The Acquisition phase (Phase 1) involved three sessions of up to 90 min each of /ɹ/ practice that were intended to be completed in 8 days or less (see McAllister et al., 2026). The treatment then transitioned to a Generalization phase (Phase 2) consisting of 16 semiweekly sessions lasting 45–60 min. The MBT and biofeedback conditions featured similar session structures and schedules. During structured practice in both Phase 1 and Phase 2, stimuli were presented, and clinician responses were recorded using the Challenge Point Program (McAllister et al., 2021), a custom open-source software for presenting speech stimuli with adaptive difficulty.

Acquisition Phase

As detailed by McAllister et al. (2026), Phase 1 (Acquisition) included three 90-min sessions within 8 days. Phase 1 targeted five specific syllables: one syllabic /ɝ/, one postvocalic /ɹ/ with a front vowel, one postvocalic /ɹ/ with a back vowel, one onset /ɹ/ with a front vowel, and one onset /ɹ/ with a back vowel. At the start of the first Acquisition session, the participants were provided with instructions aligning with their randomly assigned intervention condition,

which included an orientation to the biofeedback modality for children randomized to biofeedback. During prepractice, clinicians verbally modeled each syllable and provided detailed cues and feedback reflecting the participant's assigned condition. Prepractice ended once the participant produced each target syllable correctly three times or after 43 min, whichever came first.

Following prepractice, participants advanced to structured practice, which was designed to elicit a minimum of 150 and a maximum of 200 trials of syllables selected in the prepractice phase. These trials were grouped into blocks of 10, with each block repeating the same syllable. During structured practice, the Challenge Point Program prompted clinicians to provide qualitative verbal feedback (KP feedback) after every trial. KP feedback focused on articulatory aspects of speech (e.g., "Try to move the back of your tongue further back") and, in biofeedback conditions, referenced the visual display (e.g., "Align the third peak with the target"). Feedback concluded with the clinician modeling the target syllable to cue the next trial. In the three sessions of the Acquisition phase, there was no adaptation of the target complexity or the type and frequency of clinician feedback.

Generalization Phase

Phase 2 (Generalization) included 16 biweekly sessions lasting up to 1 hr each: 10 min of prepractice, followed by up to 250 trials of structured practice on syllables, words, phrases, or sentences containing /ɹ/. During the structured practice in Phase 2, task difficulty was automatically adjusted by the Challenge Point Program based on clinician-entered ratings. This adaptive mechanism aims to ensure that speech practice remains at an appropriately challenging level of difficulty in optimizing learning (Guadagnoli & Lee, 2004). Practice in this phase also occurred in blocks of 10 trials. The Challenge Point Program increased difficulty after blocks with eight or more correct productions, decreased difficulty after blocks with five or fewer correct productions, and maintained difficulty at the same level after blocks with six or seven correct productions. The next session began at the difficulty level where the participant left off.

The Challenge Point Program included 15 levels of increasing complexity, around which the session difficulty could be adapted. Difficulty adjustments modulated several parameters including linguistic complexity, feedback frequency (both verbal feedback and biofeedback, if applicable), and mode of elicitation. Modulations to linguistic complexity involved varying syllable counts per word, the inclusion of competing phonemes such as /l/ and /w/, and the use of carrier phrases or sentence contexts. At advanced levels, participants created their own sentences using the target words. Modulations to feedback involved gradually decreasing the frequency of KP and knowledge of results (KR) feedback, which increased the frequency of trials with no feedback. At advanced levels, participants self-evaluated their own production. In biofeedback conditions, the frequency of visual feedback also decreased as participants advanced through levels from 80% to 0% of trials. Modulations of elicitation mode involved imitative versus read production of speech stimuli and the presence or absence of prosodic manipulations (e.g., with cues to use interrogative or exclamatory intonation patterns). Initially applied in a blocked format, intonation patterns were randomized.

In addition to within-session adjustments during Phase 2 (Generalization), the Challenge Point Program also included modulations to stimulus presentation that occurred between sessions. Specifically, based on the participant's cumulative performance in a session, the practice schedule for the next session was adjusted as follows: items were initially presented in a fully blocked order (10 trials of a single stimulus, with successive blocks of the same phonetic context grouped together), then in random-blocked order (10 trials of a single stimulus, with phonetic contexts randomized across blocks), and finally in fully random practice (where different stimuli were mixed within each block).

Treatment Condition Differences

The differences between the treatment conditions are described below. Videos illustrating the treatments are available at https://osf.io/9uztn/.

MBT

No biofeedback was provided in the MBT condition. Clinicians provided cueing through auditory models, verbal descriptions of the intended articulator placement (supplemented by visual aids), and guidance based on their perceptual assessment of the participant's current production. The MBT condition was designed to reflect the best practices in treatment without access to real-time visual feedback. That is, clinicians provided verbal cues for articulator placement based on detailed knowledge of the articulatory phonetics of /ɹ/ (Preston, Benway, et al., 2020), and practice difficulty was adaptively adjusted in accordance with principles of motor learning using the Challenge Point Program, as described above. Images and diagrams of the vocal tract (available at https://osf.io/afb8k) were used to introduce different functional regions of the tongue (e.g., tip, blade, root) and various tongue shapes associated with perceptually accurate /ɹ/. The images included computer animations and drawings, magnetic resonance images, electropalatography images, and ultrasound images of the tongue during /ɹ/ production.

These images were available in the MBT condition as well as in biofeedback conditions because they demonstrated important articulatory information for cueing, and they did not constitute biofeedback because they were static images. Clinicians explained that different tongue shapes and positions can lead to a correct /ɹ/, and participants were encouraged to try different articulatory movements to find the best fit for their vocal tracts. To maintain consistency across clinicians, a standardized list of suggested cues was made available (https://osf.io/fb5dk). All participants were exposed to MBT because in the biofeedback condition, they had practice opportunities that did not feature biofeedback.

Ultrasound Biofeedback

In the ultrasound biofeedback condition, clinicians incorporated all cueing elements used in MBT, while also providing real-time ultrasound visualization of tongue shape and movement. Participants viewed the ultrasound display using either an EchoBlaster 128 CEXT-1Z, rev.C Beamformer with a PV6.4/10/128-Z-3 microconvex transducer, a MicrUS EXT-1H, rev.D Beamformer with a MC8-4R20S-3 microconvex transducer (both Telemed Medical Systems), or an Acuson X300 with a C8–5 transducer (Siemens Corporation). Ultrasound images were displayed using Echo Wave or AverMedia software, with optional annotation either by tracing on a sheet protector placed over the screen or using digital annotation tools such as EpicPen (Tank Studios).

To ensure stable probe positioning while allowing head movement, the ultrasound probe was secured using a chest plate (Hitchcock & Reda, 2020). During the first session, the participants were introduced to the basics of ultrasound imaging and learned to recognize tongue movements on the display. Key features of /ɹ/ articulation were visually examined, followed by comprehension checks to confirm that participants could visually differentiate between accurate and inaccurate /ɹ/ productions as seen through ultrasound.

Clinicians provide articulator placement cues and guidance on tongue adjustment to refine accuracy (Preston, McAllister Byun, et al., 2017). Unlike other conditions, where articulatory positioning was inferred from auditory and acoustic signals, ultrasound provided direct visual feedback of the tongue. Clinicians were free to choose the coronal or sagittal orientation of the transducer for tongue imaging in response to the participants' productions. To balance biofeedback support with the ultimate goal of independence, the ultrasound display was initially visible for the first eight out of 10 trials in each block and concealed for the final two trials to encourage internalized learning. As the children progressed to higher levels in the Challenge Point Program, ultrasound biofeedback systematically faded. More details on ultrasound cueing can be found in https://osf.io/afb8k/ and McAllister et al. (2020).

Visual-Acoustic Biofeedback

Participants assigned to the visual-acoustic biofeedback condition viewed a real-time display of the acoustic signal of speech alongside standard MBT cueing. The acoustic display was generated using the Sona-Match module of the PENTAX Medical Computerized Speech Lab (CSL) software integrated with CSL hardware model 4500 B. This software produces a live LPC spectrum displaying the spectral envelope during speech, with peaks in the spectrum inferred to represent formant frequencies.

In the first biofeedback session, participants were introduced to the LPC spectrum and encouraged to observe how the formant peaks shifted when sustaining various sounds. They were then trained to recognize the defining acoustic characteristics of an accurate /ɹ/ production: a lowered F3 and a small distance between F2 and F3. Each participant was assigned a template from a library of child and adolescent LPC traces representing perceptually accurate /ɹ/ production. Further details on template selection are available in McAllister et al. (2020) and https://osf.io/afb8k/, with the full template library accessible at https://osf.io/kj4z2/.

Using articulator placement cues similar to those in MBT, participants were encouraged to experiment with different vocal tract configurations to better match their assigned templates. As treatment progressed, new visual templates were generated to reflect the participant's improved acoustic production. To balance biofeedback support with independent learning, the acoustic display was visible for the first eight trials in each block and concealed for the last two trials, following the same approach used to fade visual support in the ultrasound biofeedback condition.

Treatment Fidelity

Treatment fidelity assessed the extent to which clinicians adhered to the prespecified treatment protocol. Fidelity ratings were obtained by a research clinician from another clinical site. Clinicians were unaware of which sessions were selected for fidelity review. Fidelity from Phase 0 (Dynamic Assessment) and Phase 1 (Acquisition) was reported by McAllister et al. (2026). In Phase 2 (Generalization), two randomly selected sessions were reviewed for fidelity for each participant. For each fidelity session, the rater reviewed the video and compared the clinician's actions to a record generated by the Challenge Point Program for each trial: (a) whether the clinician was expected to provide a verbal model before the trial, (b) whether the clinician was expected to provide KP feedback, and (c)

whether biofeedback should have been provided or withheld for participants randomized to a biofeedback treatment condition. If KP feedback was indicated, the rater also evaluated whether the clinician included the following components of KP feedback: reference to the child's articulation, reference to the visual display (if applicable), and a verbal model for the next trial.

Generalization Probes (Outcome Variable)

Participants' learning over the course of treatment was evaluated using standard probes consisting of 50 untreated words with /ɹ/ in various phonetic contexts. Each item contained only one /ɹ/. Participants read the words off the screen or were provided with cues to elicit the word if they had difficulty independently decoding an item. These probe words were elicited during the initial evaluation before the Dynamic Assessment session, at the start of the first Generalization session (after the completion of Phase 1 and before the start of Phase 2), at the start of the ninth Generalization session (halfway through Phase 2), and within approximately 1 week at the end of Phase 2 (and no later than 2 weeks after the end of Phase 2). While participants' performance on probes at all time points will be visually presented for qualitative characterization, our preregistered hypothesis (McAllister et al., 2020) focuses on comparing the first and final probes.

Generalization probe recordings were collected at all sites using a head-mounted microphone (AKG C520 Professional Head-Worn Condenser microphone) directed to a Marantz solid-state digital recorder via a Behringer UMC 404HD audio interface. Audio was recorded at 44.1 kHz and 16-bit resolution, resulting in lossless 16-bit Pulse-Code Modulation audio inside .wav containers.

Each generalization probe recording was split into 50 separate word-level productions using Praat software (Boersma & Weenink, 2022). These recordings were pooled across speakers and time points to be presented in a randomized order for rating. Perceptual ratings of these words were obtained from naïve raters recruited through Amazon's MTurk online crowdsourcing platform. The use of crowdsourced raters to provide perceptual ratings of /ɹ/ has been validated in previous studies (McAllister Byun et al., 2015, 2016). Although ratings provided by listeners recruited online are not identical to those provided by expert clinicians, the use of everyday raters arguably enhances the real-world relevance of this outcome measure (Ziegler et al., 2021). Raters were required to originate from U.S.-based IP addresses, report speaking American English as their native language, and report no significant history of hearing loss or speech language difficulty. Finally, all raters were required to pass a qualification task consisting of 100 words containing /ɹ/; to determine eligibility, their responses were compared against ratings from expert listeners as described in McAllister Byun et al. (2015). A total of 54 raters (37 women, 12 men, five nonbinary) contributed to the task, with each rater completing an average of 37.67 blocks of 200 trials (SD = 64.23, range: 1–244). The mean age was 45.34 years (SD = 8.7, range: 29–64 years).

When providing their ratings, crowdsourced raters were provided with the orthographic representation of the target word, but received no information about participant characteristics, time point of elicitation, or treatment condition. They were instructed to rate the accuracy of the /ɹ/ sound in each word in a binary fashion (correct/incorrect) and disregard other sounds in the word. Stimuli were presented in batches of 200 trials. Each block of experimental trials contained 20 "catch trials" considered unambiguously correct or incorrect; listeners had to classify at least 80% of catch trials correctly for their ratings for a given batch to be retained. Each batch also contained recordings from children with typical speech of the same age as our experimental participants (Ayala et al., 2023) so listeners heard a balance between rhotic and derhotic productions. Ratings were collected until at least nine unique listeners rated each token (McAllister Byun et al., 2015). The proportion of "correct" ratings out of the total number of ratings for a token served as the measure of /ɹ/ production accuracy for the generalization probes.

Statistical Analysis

Data analysis was carried out in the R software environment (R Core Team, 2023) using the tidyverse family of packages (Wickham et al., 2019) for data wrangling and visualization and the lme4 package (Bates et al., 2015) for mixed-effects modeling. A linear mixed-effects regression model was used to evaluate the statistical significance of differences in the perceptually rated accuracy of /ɹ/ productions across groups over time. The outcome variable was the proportion of "correct" ratings from crowdsourced listeners. Fixed effects included site, treatment condition (biofeedback or MBT), and time point (pre- and posttreatment), as well as the interaction between treatment condition and time point. Proportion of /ɹ/ rated correct at the pretreatment time point and responder type (high or low initial responder from the Dynamic Assessment session) were included in the model to control for the fact that the potential for growth may differ across participants with differing levels of ability at the baseline time point. The model also included random effects of participants and words, reflecting the fact that multiple observations were nested in these categories. Because we wanted to know if the rate of change over time differed between treatment conditions, the condition–time point interaction was the effect of primary interest. We hypothesized a significant

interaction between the treatment condition and time point, with participants who received biofeedback treatment showing greater generalization of the probes.

We also conducted a subanalysis using a similar mixed-effects model, but this time, we explored whether the two biofeedback conditions differed from one another. As in the first analysis, the outcome variable was the proportion of "correct" ratings from crowdsourced listeners assigned to the items from the generalization probe. Fixed effects included biofeedback type (ultrasound or visual-acoustic), time point (pre- and posttreatment), site, responder group (high or low initial responder from the Dynamic Assessment session), as well as the interaction between treatment condition and timepoint. Random effects of participants and words were included in the model. An alpha of .05 was used to test statistical significance.

Results

Treatment Fidelity

Table 1 reports average fidelity scores for clinicians for Generalization (Phase 2, see McAllister et al., 2026, for fidelity data for Dynamic Assessment and Phase 1). Table 1 reveals high levels of fidelity in providing biofeedback, auditory models, KR and KP feedback, and eliciting self-monitoring during the sessions selected for fidelity review. This indicates that, across treatment conditions, clinicians could readily follow the Challenge Point software prompts as prescribed and that the treatment conditions were implemented as designed.

Achieved Treatment Dosage and Schedule

Figure 2 shows the dosage across structured practice trials for the Biofeedback and MBT groups. The mean number of trials completed over the course of all treatment sessions by a given individual was 4,528.3 total trials in the biofeedback condition and 4,560.0 total trials in the MBT condition. This difference was not statistically significant in a t test for independent samples, $t(97.99) = 1.43$, $p = .16$.

Furthermore, data from the Challenge Point Program were examined across all sessions to characterize the amount of biofeedback that was provided. These data indicated that the proportion of trials in which children in each group received biofeedback in Phase 2 was 0% for the MBT condition, 52.5% for the ultrasound biofeedback condition, and 45.4% for the visual-acoustic biofeedback condition. Differences in biofeedback frequency are due to the algorithm in the Challenge Point Program systematically reducing the amount of biofeedback available as children progress in treatment, as rated by their clinicians.

With regard to fidelity to the intended schedule, the MBT group completed their 16 Generalization treatment sessions in an average of 8.8 weeks ($SD = 1.31$, range: 7.0–12.6) and the Biofeedback group completed their sessions in an average of 8.8 weeks ($SD = 1.32$, range: 6.7–14.7). The intended upper limit for completing treatment sessions was 10 weeks, and five participants exceeded this limit (three MBT, two Biofeedback). All participants completed their posttreatment probe within the specified 2 weeks after the final Generalization treatment session.

Listener Demographics and Reliability

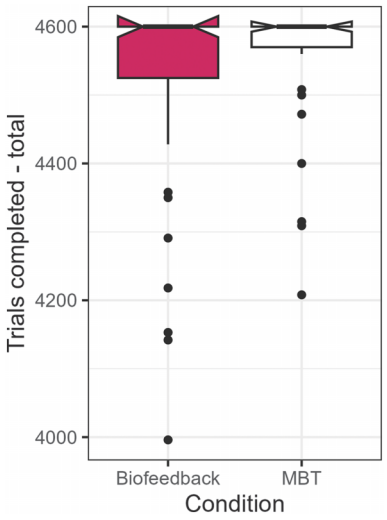
In addition to the trials presented for the experimental ratings, listeners heard filler files drawn from a previous study (Ayala et al., 2023). Files were drawn randomly from the same master list across rating batches, creating repeated presentations that could be used to compute intrarater reliability. In certain cases, a single listener rated the same token more than twice. For tractability of the analysis, we computed intrarater reliability statistics on exactly two ratings of a given file from each listener, randomly selected across their total set of ratings. This yielded 153,505 observations across 19 raters (not every

Table 1. Mean fidelity achieved in generalization sessions for seven parameters of interest.

Fidelity category	Motor-based treatment (n = 82 sessions)	Biofeedback (n = 126 sessions)
Biofeedback as indicated	N/A	99.6 (SD = 1.8, range: 84–100)
Model as indicated	99.3 (SD = 3.0, range: 82–100)	99.6 (SD = 2.3, range: 83–100)
Knowledge of results as indicated	99.7 (SD = 1.0, range: 94–100)	99.6 (SD = 1.7, range: 83–100)
Knowledge of performance (reference to articulation) as indicated	99.4 (SD = 2.0, range: 88–100)	99.0 (SD = 3.7, range: 72–100)
Knowledge of performance (reference to biofeedback display) as indicated	N/A	99.2 (SD = 3.0, range: 78–100)
Knowledge of performance (model) as indicated	99.4 (SD = 2.5, range: 84–100)	99.2 (SD = 3.6, range: 72–100)
Self-eval as indicated	99.9 (SD = 0.2, range: 98–100)	100 (SD = 0.0, range: 100–100)

Note. N/A = not applicable.

Figure 2. Achieved dosage (practice trials) in structured practice. MBT = motor-based treatment.



rater had repeated measurements because some raters completed only one block of trials). The number of repeated trials completed by a single rater depended on the number of blocks completed, with raters completing an average of 1,114.9 repeated trials (SD = 995.15, range: 3–2099). The mean agreement across repeated trials was 95.97% (SD = 4.46, range: 83.33–100).

As a measure of interrater agreement, the proportion of raters who agreed with the modal classification for each token (i.e., correct or incorrect) was calculated. On average, agreement across raters was 86.04%. A one-proportion z test indicated that agreement across raters was significantly greater than chance (50%), X-squared = 101394.86, $p < .0001$.

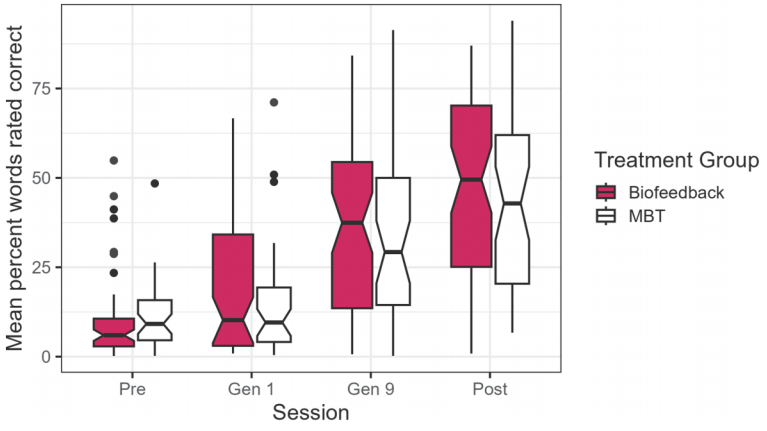
Descriptive Statistics

The boxplots in Figure 3 represent the distribution of participants' average percent of /ɹ/ words rated correct at each time point. In each boxplot, the horizontal black line represents the median, the box represents the interquartile range (25th–75th percentile), and outliers are represented as points. The notch in each boxplot represents a confidence interval around the median value based on the median +/− 1.58 times the interquartile range divided by the square root of the number of observations. Figure 3 shows that the average perceptually rated accuracy increased across the four time points represented by participants in both the biofeedback and MBT conditions. In the biofeedback group, percent /ɹ/ words rated correctly started at a median of 5.97 in the pretreatment evaluation session, increasing to 49.2 at the posttreatment time point. For the MBT group, percent /ɹ/ words correct started at a slightly higher median of 9.16 and ended at a slightly lower median of 42.86. Impressionistically, the degree of change was higher in the biofeedback group than in the MBT group early in the course of treatment (e.g., Pre vs. Gen 1 and Gen 1 vs. Gen 9), but in the final part of the study, both groups made a similar degree of progress.

Inferential Statistics

Linear mixed-effects models were used to statistically analyze changes in perceptually rated accuracy over the course of treatment. Following the preregistered analysis plan, only accuracy at the pre- and posttreatment time points were included in the analysis. The model controlled for both proportion of correct /ɹ/ at the baseline time point and baseline accuracy group (based on clinician-rated accuracy during the Dynamic Assessment session),

Figure 3. Plot comparing mean percent correct in biofeedback versus motor-based treatment conditions at four timepoints: Pretreatment, at the start of Generalization treatment (Gen 1, after Acquisition treatment), halfway through Generalization treatment (Gen 9), and Posttreatment. MBT = motor-based treatment.



since the latter was used to generate the randomization strata but the former was likely to explain more of the variance in the outcome variable.[1]

The average percent of correct /ɹ/ ratings at the Post time point were statistically significantly higher than ratings at Pre for both the biofeedback condition (β = 36.73, *SE* = 0.56, $p < .0001$) and the MBT condition (β = 33.61, *SE* = 0.68, $p < .0001$). There was no statistically significant difference in the average scores across treatment sites (MSU vs. NYU: β = 4.11, *SE* = 4.23, $p = .33$; SU vs. NYU: β = 3.69, *SE* = 4.22, $p = .38$; MSU vs. SU: β = −.42, *SE* = 2.56, $p = .87$). Baseline accuracy (the percentage of items scored perceptually correct at the pretreatment time point) was a significant predictor of accuracy at the posttreatment time point: unsurprisingly, higher accuracy at the start of treatment was associated with higher accuracy after treatment (β = 0.84, *SE* = 0.13, $p < .0001$), even after adjusting for being in the a high or low baseline accuracy group (high initial responders vs. low initial responders, which was not a statistically significant predictor of accuracy at posttreatment (β = 5.02, *SE* = 2.68, $p = .065$) apart from this linear trend).

The MBT group did not differ significantly in accuracy from the biofeedback group at either the Pre time point (β = 0.40, *SE* = 2.5, $p = .87$) or Post time point (β = −2.72, *SE* = 2.5, $p = .28$). However, as a critical test of our hypothesis, we observed a statistically significant interaction between the treatment condition and time point (β = −3.12, *SE* = 0.88, $p < .001$), consistent with a smaller rate of change from pre- to post-intervention for the MBT condition than the biofeedback treatment condition. The full model results are presented in Supplemental Table 1 at https://osf.io/sb68u.

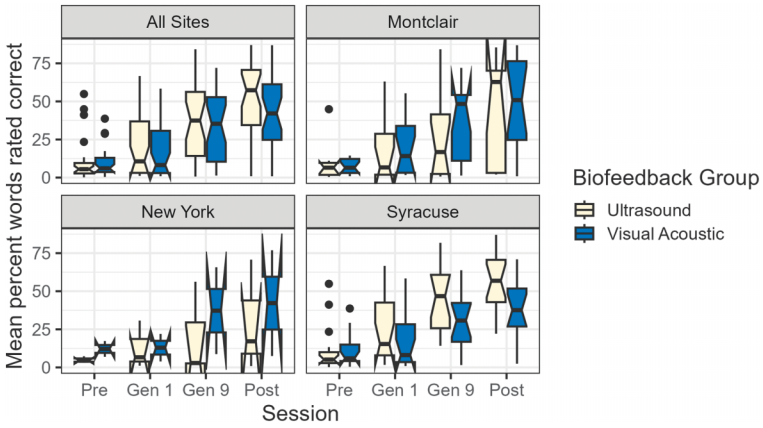
Biofeedback Type Comparison

In a subanalysis, we examined whether there were any differences between the effects of ultrasound biofeedback and visual-acoustic biofeedback conditions with regard to generalization. As can be seen in the upper left panel of Figure 4, the two biofeedback types were similar at the pretreatment time point. A mixed-effects model was run for only the two biofeedback conditions. The outcome variable was the mean percentage of correct words, and there were fixed effects for biofeedback type (ultrasound and visual-acoustic), time (Pre and Post), site (MSU, NYU, SU), and pretreatment responder groups (high and low). Random effects of participants and words were included in the model. As in the model comparing biofeedback and MBT, there were no significant differences in average scores across treatment sites (MSU vs. NYU: β = 6.55, *SE* = 6.29, $p = .30$; SU vs. NYU: β = 5.18, *SE* = 6.18, $p = .41$; SU vs. MSU: β = −1.37, *SE* = 3.55, $p = .70$) or across baseline accuracy groups (β = 2.32, *SE* = 3.75, $p = .54$). Within these groups, baseline accuracy score was a statistically significant predictor of the outcome (β = 0.9, *SE* = 0.16, $p < .0001$). Scores at the posttreatment time point were statistically significantly higher than scores at the pretreatment time point for both ultrasound biofeedback (β = 39.63, *SE* = 0.79, $p < .0001$) and visual-acoustic biofeedback (β = 33.92, *SE* = 0.78, $p < .0001$). There was no statistically significant effect of biofeedback type at either the Pre time point (β = −0.38, *SE* = 3.39, $p = .91$) or the Post time point (β = −6.09, *SE* = 3.39, $p = .078$). However, there was a significant group-by-time interaction (β = −5.71, *SE* = 1.1, $p < .0001$), suggesting that greater pre- to posttreatment improvement in /ɹ/ production accuracy was seen with ultrasound biofeedback compared to visual-acoustic biofeedback.

As this effect was unexpected, we conducted additional data visualization to better understand the factors underlying this finding. Figure 4 shows the progress of each biofeedback condition over time at each of the three treatment sites. Figure 4 reveals somewhat different patterns of progress in the two biofeedback conditions across the sites. In particular, at the SU site, participants in the ultrasound condition outperformed participants in the visual-acoustic biofeedback condition at all points after the initial evaluation, whereas at MSU, participants in the visual-acoustic biofeedback condition appeared to make more rapid progress than participants in the ultrasound biofeedback condition at all points until the final evaluation, when the ultrasound group showed a higher median accuracy (albeit with substantial variability). The NYU site, which was activated as part of a COVID-19 contingency plan and saw many fewer participants than the other sites, tended to lag behind in accuracy in both conditions, most notably in ultrasound. Based on these observations, we explored a statistical model that further included a three-way interaction between biofeedback type, time, and site. In this model, the three-way interaction between time point, biofeedback type, and site comparing Syracuse to NYU was statistically significant (β =

[1]The models reported here deviate from the preregistered plan in one respect: The preregistered plan indicates a random effect of listener (i.e., the individual raters on Amazon Mechanical Turk). However, this was an error—if we were analyzing individual listeners' ratings of each token, our model would be a logistic regression, and the preregistered plan indicates a linear regression. We corrected this inconsistency by collapsing across raters, yielding one measure (proportion of "correct" votes) for each token, which we fit with a linear mixed-effects model. As specified in the preregistered plan, models were fit with the maximal random-effects structure that was supported by model comparison and yielded no convergence or singular fit errors. This resulted in the inclusion of random intercepts for subject and word. A random slope of session on subject was examined but not retained because it yielded a singular fit error.

Figure 4. Subanalysis comparing biofeedback types, with panels by site.



−24.24, *SE* = 4.09, $p < .0001$); the three-way interaction between time point, biofeedback type, and site comparing Syracuse to Montclair was statistically significant (β = 25.36, *SE* = 2.31, $p < .0001$); the three-way interaction between time point, biofeedback and type and site comparing Montclair to NYU was statistically significant (note the two-way interaction between biofeedback condition and time was not significant for the NYU site, β = 5.59, *SE* = 3.8, $p = .14$). While this does not negate the finding in our primary statistical model, it does suggest that other factors are at play that warrant consideration when comparing biofeedback types. We return to this point in detail in the Discussion section. Complete results of the regression model can be found in Supplemental Table 1 at https://osf.io/sb68u.

Discussion

This multisite randomized controlled trial compared changes in /ɹ/ production by American English–speaking children with RSSD, who were randomly assigned to receive biofeedback treatment or a comparison condition of MBT without biofeedback. A companion study testing a theoretically motivated prediction about the effects of biofeedback in the early stages of treatment found that progress was significantly more rapid in the biofeedback condition than in motor-based treatment. This study reports on the project's secondary outcome, investigating generalization of improved /ɹ/ in untreated words as a measure of motor learning after the end of all treatment. The results from blinded listeners' perceptual ratings indicated that, after treatment, productions of /ɹ/ from children randomly assigned to receive biofeedback treatment were more likely to be rated as correct on untreated words compared to productions of /ɹ/ from children who received treatment without biofeedback (MBT). The results confirmed our hypothesis that generalization associated with biofeedback would significantly exceed that observed in MBT without biofeedback.

Whereas prior studies utilizing single-subject experimental designs have offered promising or inconclusive results regarding the effects of biofeedback on generalization, this is the largest study on biofeedback treatment to date, and it employed a rigorous randomized group design. With randomized controlled trials serving as the gold standard research design for evidence-based decision making (Dollaghan, 2007), the study provides support for the adoption of treatment incorporating biofeedback in clinical practice when learning to produce American English /ɹ/. The results are an important clinical follow-up to our companion study, which established that the addition of biofeedback accelerated acoustic changes in /ɹ/ during the first three treatment sessions for the same participants (McAllister et al., 2026). The results also align with a recent biofeedback study for children with apraxia of speech, in which ultrasound biofeedback improved generalization in /ɹ/ and other phonemes over nonbiofeedback treatment when the treatments began within a week of intensive practice (Preston et al., 2024).

Generalization learning at the end of all treatments was treated as a secondary rather than a primary outcome in the present study because the principles of motor learning predict an advantage for biofeedback in the acquisition of a new motor plan, but not necessarily in its retention and generalization. In fact, research on nonspeech motor learning suggests that providing detailed visual KP feedback may inhibit generalization (Hodges & Franks, 2001). Our study found the opposite: Specifically, treatment incorporating biofeedback was more effective at facilitating generalization than treatment without biofeedback. These differences may be explained by several factors, including the fact that our study included children

with RSSD learning to override an existing off-target motor plan, whereas Hodges and Franks studied visual knowledge of performance feedback with unimpaired adults learning a motor skill for which they did not have an existing motor schema. In addition, we intentionally decreased the frequency of biofeedback as participants' accuracy increased in a specific effort to avoid excessive dependency on visual feedback and to encourage generalization. As structured in the present study, it appears that biofeedback is an effective means of training a new motor plan both during the acquisition phase (McAllister et al., 2026) and when generalizing the motor plan.

It is noteworthy that while the addition of biofeedback resulted in higher average accuracy on untreated words after 20 treatment sessions, substantial improvement in /ɹ/ production was also seen in the MBT condition. Thus, while the addition of biofeedback can facilitate progress in treatment for RSSD affecting /ɹ/, the systematic implementation of the MBT approach also has evidence of efficacy. This finding may seem surprising in light of the fact that many of our participants had been receiving traditional treatment for /ɹ/ for months or years prior to their enrollment in our study. We posit that a high dosage of individualized intervention (cumulative total of over 4,500 trials) played a role in the robust response to treatment observed across treatment conditions. It is also possible that the participants benefited from other aspects of our MBT methodology that have been refined over the course of our previous research studies, including phonetically informed cues for articulator placement (Preston, Benway, et al., 2020) and a schedule of clinician feedback grounded in the principles of motor learning (Maas et al., 2008).

Additionally, while we did not have an initial hypothesis regarding differences between biofeedback types, we did observe greater generalization in /ɹ/ production on average for children who received ultrasound biofeedback compared to those who received visual-acoustic biofeedback. There are several possible explanations for this finding. As discussed in our companion paper (McAllister et al., 2026), visual-acoustic and ultrasound biofeedback may target different underlying processes related to speech motor learning. It is possible that ultrasound provides a more concrete description for children to modify their articulation and that directing children's focus to motor aspects of production (i.e., tongue shapes and movements) could more directly address a modifiable aspect of their speech sound representations. Alternatively, it is possible that there may be distinct profiles of children with RSSD with varying differences in motor and auditory representations in speech, and that our sample included children whose profiles were a better fit for treatment with specific feedback targeting articulatory movements. This possibility will be explored in a companion paper analyzing children's treatment gains in relation to the profile of sensory strengths and weaknesses they exhibited on a battery of measures of auditory and oral somatosensory function administered prior to treatment. It may also be the case that ultrasound biofeedback enables clinicians to observe information about a child's particular misarticulation, which they can continue to leverage even after the ultrasound biofeedback has faded or withdrawn. This possibility is of particular relevance for the present study since we deliberately reduced the rate at which biofeedback was available as participants made progress in treatment. Related to this question of biofeedback fading, we noted that the ultrasound biofeedback group received slightly more practice with biofeedback available (52.5% of trials) than children in the visual-acoustic condition (45.4% of trials). This suggests that participants in the VAB condition made faster progress early in the course of treatment (leading to a faster schedule for fading of biofeedback), whereas the progress of participants in the ultrasound condition was slower and steady, but led to more successful overall outcomes. Given that prior research suggests that more biofeedback may lead to greater generalization (Preston et al., 2018), it is possible that a slower schedule of biofeedback fading could be more effective in promoting generalization. This possibility is worth exploring using research designs that systematically manipulate the timing of changes in biofeedback availability. Finally, another plausible explanation, which is suggested by the significant interaction with the treatment site, is that clinician expertise or experience differs by biofeedback type. Further exploration of factors that affect clinician delivery of interventions, as well as children's individualized responses to these interventions, is warranted.

Caveats and Limitations

Although this is the largest clinical trial on biofeedback approaches to SSD treatment, there are several important caveats. For example, the study focused on American English /ɹ/ because it is commonly described as the most frequent and challenging speech deviation in RSSD in mainstream American English. However, other sounds are also commonly misarticulated in English (e.g., /s/, /z/, /l/), and other sound targets are appropriate for individuals with different language or dialect backgrounds. We also limited treatment to 20 total sessions (one Dynamic Assessment, three Acquisition, and 16 Generalization), and it is therefore unclear what the effects might be with more extended treatment or different session parameters. Furthermore, our protocol involved a systematic reduction of the amount of biofeedback available to children assigned to that condition as their speech improved, so it is possible that greater or less access to biofeedback might impact generalization outcomes

(Preston et al., 2018). Finally, we evaluated participants' progress at a time point immediately after the end of treatment. Future research should additionally probe participants' performance at later time points (e.g., several weeks or months after the end of treatment) to better understand the long-term retention of gains made through treatment with and without biofeedback.

We also acknowledge that the dosage and intensity of the intervention delivered in the present study are challenging to achieve in real-world clinical practice, especially in school settings, where large caseloads often necessitate group service delivery for children with RSSD. In a five-phase clinical research (Robey, 2004), this study represents an efficacy trial that measures the response to treatment under idealized conditions. The next step in the sequence is an effectiveness trial, which measures the benefit of a treatment method when it is "provided in a typical fashion by typical practitioners to typical patients in typical clinical settings" (p. 402). Thus, future studies should examine the effects of treatment with and without biofeedback when delivered under real-world constraints such as lower intensity or group service delivery. Future studies should also explore strategies to increase treatment dosage while addressing these practical limitations. For example, artificial intelligence–supported home practice tools can supplement clinician-guided sessions, boosting the total dosage without adding to clinicians' workloads (Benway et al., 2024; Benway & Preston, 2024). Finally, the adoption of biofeedback in clinical practice might be limited by the availability of specialized equipment and clinicians with the requisite skills. Although an increasing number of clinics are investing in ultrasound devices, the cost of the equipment remains a significant barrier. Additionally, patient access to ultrasound biofeedback is limited because of the requirement that the clinician and patient are in the same room; it cannot be delivered via telepractice. However, access to visual-acoustic biofeedback may be accomplished with only a microphone and software, and visual-acoustic biofeedback can readily be conducted via telehealth. Lack of familiarity with best practices for the delivery of biofeedback can present another barrier to its implementation. While there are published tutorials aimed at educating clinicians on biofeedback approaches (Hitchcock et al., 2023; Preston, McAllister Byun, et al., 2017), successful implementation requires ongoing training. Increasing the representation of biofeedback in graduate training and continuing education is an important goal in increasing the adoption of this evidence-based method.

Future Directions

Having documented the impact of biofeedback on /ɹ/ acquisition in our companion paper and its effect on generalization in the current study, future research may include studies designed to answer additional questions. For example, would biofeedback yield similar advantages when treating different sounds, speakers of other languages, or individuals with different clinical profiles (e.g., younger ages or other concomitant diagnoses)? As noted above, it will also be important to conduct studies on the effectiveness of biofeedback when implemented in real-world settings, which may differ from the present study in terms of factors such as the dosage and intensity of treatment. Furthermore, because biofeedback interventions may require some financial outlay for technology and/or training, studies examining the cost-effectiveness of treatment with biofeedback are also worthwhile. Finally, while the present study included only visual-acoustic and ultrasound biofeedback, these are not the only types of biofeedback for speech; other methods such as electropalatography or electromagnetic articulography are also worthy of investigation.

The current study investigated the group-averaged results of MBT and biofeedback conditions. However, within each condition, a range of treatment responses was observed. Further research is needed to identify individual predictors of responses to different treatment conditions. Prior research suggests that individuals who are not stimulable for /ɹ/ may be good candidates for biofeedback, whereas those who are readily stimulable may achieve success with MBT (McAllister et al., 2022). Other individual factors, such as auditory or somatosensory profiles related to speech processing, may also be informative in determining which treatment is most suitable for a child (Benway et al., 2021).

Finally, while we emphasized the potential value of biofeedback in promoting the acquisition and generalization of speech targets, it is important to reiterate that MBT interventions remain a viable approach when designed and implemented following best-practice guidelines (Preston, Benway, et al., 2020). Modifications to MBT that do not require biofeedback tools, such as varying the distribution of treatment sessions (Herbst et al., 2025; Preston et al., 2024), may also affect treatment outcomes.

Conclusions

This randomized controlled trial provides evidence that biofeedback enhances generalization in the production of /ɹ/ in children with RSSD, surpassing the improvements observed with MBT alone. While both treatment conditions facilitated gains, biofeedback led to more accurate productions of untreated words after 20 sessions, reinforcing its clinical utility. Differences between biofeedback modalities suggest that ultrasound may offer unique advantages, potentially due to its emphasis on articulatory

movements. Future research should explore individual predictors of treatment response and the broader applicability of biofeedback across different speech sounds, populations, and clinical settings.

Data Availability Statement

Project-related materials, including code and de-identified data to reproduce the analyses reported here, can be found at https://osf.io/6qs4d/.

Acknowledgments

This research was funded by National Institutes of Health Grant R01DC017476, Biofeedback-enhanced treatment for speech sound disorder: Randomized controlled trial & delineation of sensorimotor subtypes (2019–2023, PI: Tara McAllister). The research team is grateful for the contributions of the following individuals (ordered alphabetically): Samantha Ayala, Molly Beiting, Nicole Caballero, Twylah Campbell, Erin Doty, Amanda Eads, Jayce Garner, Sarah Granquist, Lynne Henry, Heather Kabakoff, Danielle Kealy, Roberta Lazarus, Megan Matson, Laura Ochs, José A. Ortiz, Cory Pinto, Amy Schwartz, Michelle Swartz, and Stephanie Urlass. The authors thank Siemens Corporation for the loan of an Acuson X300 ultrasound device to New York University.

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology, 16*(2), 128–139. https://doi.org/10.1044/1058-0360(2007/017)

Ayala, S. A., Eads, A., Kabakoff, H., Swartz, M. T., Shiller, D. M., Hill, J., Hitchcock, E. R., Preston, J. L., & McAllister, T. (2023). Auditory and somatosensory development for speech in later childhood. *Journal of Speech, Language, and Hearing Research, 66*(4), 1252–1273. https://doi.org/10.1044/2022_JSLHR-22-00496

Bacsfalvi, P. (2010). Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech-Language Pathology and Audiology, 34*(3), 206–217. https://cjslpa.ca/files/2010_CJSLPA_Vol_34/No_03_153-225/Bacsfalvi_CJSLPA_2010.pdf [PDF]

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting-linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benway, N. R., Hitchcock, E. R., McAllister, T., Feeny, G. T., Hill, J., & Preston, J. L. (2021). Comparing biofeedback types for children with residual /ɹ/ errors in American English: A single-case randomization design. *American Journal of Speech-Language Pathology, 30*(4), 1819–1845. https://doi.org/10.1044/2021_AJSLP-20-00216

Benway, N. R., & Preston, J. L. (2024). Artificial intelligence-assisted speech therapy for /ɹ/: A single-case experimental study. *American Journal of Speech-Language Pathology, 33*(5), 2461–2486. https://doi.org/10.1044/2024_AJSLP-23-00448

Benway, N. R., Preston, J. L., Salekin, A., Hitchcock, E., & McAllister, T. (2024). Evaluating acoustic representations and normalization for rhoticity classification in children with speech sound disorders. *JASA Express Letters, 4*(2), Article 025201. https://doi.org/10.1121/10.0024632

Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics, 19*(6–7), 605–617. https://doi.org/10.1080/02699200500114028

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (Version 6.2.10) [Computer program]. Retrieved March 20, 2022, from https://praat.org

Bowers, L., & Huisingh, R. (2011). *Linguisystems Articulation Test*. LinguiSystems. https://www.mindresources.com/education/059955

Cabbage, K., & Carrell, T. (2014). The relationship between speech perception and production: Evidence from children with speech production errors. *The Journal of the Acoustical Society of America, 135*(4), Article 2420. https://doi.org/10.1121/1.4878036

Campbell, H., Harel, D., Hitchcock, E., & McAllister Byun, T. (2018). Selecting an acoustic correlate for automated measurement of American English rhotic production in children. *International Journal of Speech-Language Pathology, 20*(6), 635–643. https://doi.org/10.1080/17549507.2017.1359334

Cronin, S. A., Blanchet, P. G., Klonsky, B. G., & Piazza, N. (2014). The effect of a modeled /r/ articulatory disorder on listener perceptions of speech skills and personality traits. *Contemporary Issues in Communication Science and Disorders, 41*(Fall), 169–178. https://doi.org/10.1044/cicsd_41_f_169

Crowe, K., & McLeod, S. (2020). Children's English consonant acquisition in the United States: A review. *American Journal of Speech-Language Pathology, 29*(4), 2155–2169. https://doi.org/10.1044/2020_AJSLP-19-00168

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Brookes.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America, 108*(1), 343–356. https://doi.org/10.1121/1.429469

Flipsen, P., Jr. (2015). Emergence and prevalence of persistent and residual speech errors. *Seminars in Speech and Language, 36*(4), 217–223. https://doi.org/10.1055/s-0035-1562905

Goldman, R., & Fristoe, M. (2015). *Goldman Fristoe Test of Articulation–Third Edition*. Pearson.

Gray, S. I., & Shelton, R. L. (1992). Self-monitoring effects on articulation carryover in school-age children. *Language, Speech, and Hearing Services in Schools, 23*(4), 334–342. https://doi.org/10.1044/0161-1461.2304.334

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Harper, S., Goldstein, L., & Narayanan, S. (2020). Variability in individual constriction contributions to third formant values in American English /ɹ/. *The Journal of the Acoustical Society of America, 147*(6), 3905–3916. https://doi.org/10.1121/10.0001413

Herbst, B. M., Beiting, M., Schultheiss, M., Benway, N. R., & Preston, J. L. (2025). Speech in ten-minute sessions: A pilot

randomized controlled trial of the chaining SPLITS service delivery model. *Language, Speech, and Hearing Services in Schools, 56*(1), 102–117. https://doi.org/10.1044/2024_LSHSS-24-00043

Hitchcock, E. R., Harel, D., & 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*(4), 283–294. https://doi.org/10.1055/s-0035-1562911

Hitchcock, E. R., Ochs, L. C., Swartz, M. T., Leece, M. C., Preston, J. L., & McAllister, T. (2023). Tutorial: Using visual–acoustic biofeedback for speech sound training. *American Journal of Speech-Language Pathology, 32*(1), 18–36. https://doi.org/10.1044/2022_AJSLP-22-00142

Hitchcock, E. R., & Reda, J. (2020). *A new ultrasound probe stabilizer for speech therapy* [Poster presentation]. Ultrafest IX, Indiana University, United States.

Hitchcock, E. R., Swartz, M. T., & Lopez, M. (2019). Speech sound disorder and visual biofeedback intervention: A preliminary investigation of treatment intensity. *Seminars in Speech and Language, 40*(2), 124–137. https://doi.org/10.1055/s-0039-1677763

Hodges, N. J., & Franks, I. M. (2001). Learning a coordination skill: Interactive effects of instruction and feedback. *Research Quarterly for Exercise and Sport, 72*(2), 132–142. https://doi.org/10.1080/02701367.2001.10608943

Lewis, B. A., Freebairn, L., Tag, J., Igo, R. P., Jr., Ciesla, A., Iyengar, S. K., Stein, C. M., & Taylor, H. G. (2019). Differential long-term outcomes for individuals with histories of preschool speech sound disorders. *American Journal of Speech-Language Pathology, 28*(4), 1582–1596. https://doi.org/10.1044/2019_AJSLP-18-0247

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

McAllister, T., Eads, A., Kabakoff, H., Scott, M., Boyce, S., Whalen, D. H., & Preston, J. L. (2022). Baseline stimulability predicts patterns of response to traditional and ultrasound biofeedback treatment for residual speech sound disorder. *Journal of Speech, Language, and Hearing Research, 65*(8), 2860–2880. https://doi.org/doi:10.1044/2022_JSLHR-22-00161

McAllister, T., Hitchcock, E. R., & Ortiz, J. A. (2021). Computer-assisted challenge point intervention for residual speech errors. *Perspectives of the ASHA Special Interest Groups, 6*(1), 214–229. https://doi.org/10.1044/2020_PERSP-20-00191

McAllister, T., Preston, J. L., Benway, N. R., Hill, J., Lara, M. P., Leece, M. C., Liang, W., & Hitchcock, E. R. (2026). Rhotic acquisition is more rapid in biofeedback than motor-based treatment for residual speech sound disorder: Primary outcome of a randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 69*(4), 1342–1361. https://doi.org/10.1044/2025_JSLHR-24-00909

McAllister, T., Preston, J. L., Hitchcock, E. R., & Hill, J. (2020). Protocol for correcting residual errors with Spectral, ULtrasound, Traditional Speech therapy Randomized Controlled Trial (C-RESULTS RCT). *BMC Pediatrics, 20*(1), Article 66. https://doi.org/10.1186/s12887-020-1941-5

McAllister Byun, T. (2017). Efficacy of visual–acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language, and Hearing Research, 60*(5), 1175–1193. https://doi.org/10.1044/2016_JSLHR-S-16-0038

McAllister Byun, T., & Campbell, H. (2016). Differential effects of visual-acoustic biofeedback intervention for residual speech errors. *Frontiers in Human Neuroscience, 10,* Article 567. https://doi.org/10.3389/fnhum.2016.00567

Mcallister Byun, T., Campbell, H., Carey, H., Liang, W., Park, T. H., & Svirsky, M. (2017). Enhancing intervention for residual rhotic errors via app-delivered biofeedback: A case study. *Journal of Speech, Language, and Hearing Research, 60*(6S), 1810–1817. https://doi.org/10.1044/2017_JSLHR-S-16-0248

McAllister Byun, T., Halpin, P. F., & Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders, 53,* 70–83. https://doi.org/10.1016/j.jcomdis.2014.11.003

McAllister Byun, T., Harel, D., Halpin, P. F., & Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders, 64,* 91–102. https://doi.org/10.1016/j.jcomdis.2016.07.001

McAllister Byun, T., & Hitchcock, E. R. (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology, 21*(3), 207–221. https://doi.org/10.1044/1058-0360(2012/11-0083)

McAllister Byun, T., Hitchcock, E. R., & 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. https://doi.org/10.1044/2014_JSLHR-S-14-0034

Ochs, L. C., Leece, M. C., Preston, J. L., McAllister, T., & Hitchcock, E. R. (2023). Traditional and visual–acoustic biofeedback treatment via telepractice for residual speech sound disorders affecting /ɹ/: Pilot study. *Perspectives of the ASHA Special Interest Groups, 8*(6), 1533–1553. https://doi.org/10.1044/2023_PERSP-23-00120

Preston, J. L., Benway, N. R., Leece, M. C., Hitchcock, E. R., & McAllister, T. (2020). Tutorial: Motor-based treatment strategies for /r/ distortions. *Language, Speech, and Hearing Services in Schools, 51*(4), 966–980. https://doi.org/10.1044/2020_lshss-20-00012

Preston, J. L., Caballero, N. F., Leece, M. C., Wang, D., Herbst, B. M., & Benway, N. R. (2024). A randomized controlled trial of treatment distribution and biofeedback effects on speech production in school-age children with apraxia of speech. *Journal of Speech, Language, and Hearing Research, 67*(9S), 3414–3436. https://doi.org/10.1044/2023_JSLHR-22-00622

Preston, J. L., Hitchcock, E. R., & Leece, M. C. (2020). Auditory perception and ultrasound biofeedback treatment outcomes for children with residual /ɹ/ distortions: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 63*(2), 444–455. https://doi.org/10.1044/2019_JSLHR-19-00060

Preston, J. L., & Leece, M. C. (2017). Intensive treatment for persisting rhotic distortions: A case series. *American Journal of Speech-Language Pathology, 26*(4), 1066–1079. https://doi.org/10.1044/2017_AJSLP-16-0232

Preston, J. L., Leece, M. C., & Maas, E. (2017). Motor-based treatment with and without ultrasound feedback for residual speech-sound errors. *International Journal of Language and Communication Disorders, 52*(1), 80–94. https://doi.org/10.1111/1460-6984.12259

Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., & 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*(8), 1875–1892. https://doi.org/10.1044/2018_JSLHR-S-17-0441

Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., & Whalen, D. H. (2019). Remediating residual rhotic errors with traditional and ultrasound-enhanced treatment: A single-case experimental study. *American Journal of Speech-Language Pathology, 28*(3), 1167–1183. https://doi.org/10.1044/2019_AJSLP-18-0261

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., & Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *Journal of Visualized Experiments, 2017*(119), Article e55123. https://doi.org/10.3791/55123

Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research, 57*(6), 2102–2115. https://doi.org/10.1044/2014_JSLHR-S-14-0031

Raaz, C., Leece, M. C., McAllister, T., & Preston, J. L. (2021). Treatment generalization from trained /ɹ/ to untrained /l/: A case study of persisting distortion errors. *Clinical Linguistics & Phonetics, 35*(12), 1210–1219. https://doi.org/10.1080/02699206.2021.1879273

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.5). R Foundation for Statistical Computing. https://www.r-project.org/

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37*(5), 401–411. https://doi.org/10.1016/j.jcomdis.2004.04.003

Ruscello, D. M. (1995). Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders, 28*(4), 279–302. https://doi.org/10.1016/0021-9924(95)00058-X

Schulz, K. F., Altman, D. G., Moher, D., & The CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(1), Article 18. https://doi.org/10.1186/1741-7015-8-18

Sénéchal, M., Ouellette, G., & Young, L. (2004). Testing the concurrent and predictive relations among articulation accuracy, speech perception, and phoneme awareness. *Journal of Experimental Child Psychology, 89*(3), 242–269. https://doi.org/10.1016/j.jecp.2004.07.005

Shriberg, L. D. (2009). Childhood speech sound disorders: From postbehaviorism to the postgenomic era. In R. Paul & P. Flipsen (Eds.), *Speech sound disorders in children* (pp. 1–33). Plural.

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., & Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795–824. https://doi.org/10.3109/02699206.2010.503006

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 26*(5), 445–482. https://doi.org/10.3109/02699206.2012.655841

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research, 41*(4), 941–950. https://doi.org/10.1044/jslhr.4104.941

Shuster, L. I., Ruscello, D. M., & Toth, A. R. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology, 4*(2), 37–44. https://doi.org/10.1044/1058-0360.0402.37

Sjolie, G. M., Leece, M. C., & Preston, J. L. (2016). Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders, 64,* 62–77. https://doi.org/10.1016/j.jcomdis.2016.10.003

Sugden, E., Lloyd, S., Lam, J., & Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International Journal of Language & Communication Disorders, 54*(5), 705–728. https://doi.org/10.1111/1460-6984.12478

Thoonen, G., Maassen, B., Gabreëls, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13*(1), 1–23. https://doi.org/10.1080/026992099299211

Thoonen, G., Maassen, B., Wit, J., Gabreels, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10*(4), 311–336. https://doi.org/10.3109/02699209608985178

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. , François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Pearson.

Wren, Y., Pagnamenta, E., Peters, T. J., Emond, A., Northstone, K., Miller, L. L., & Roulstone, S. (2021). Educational outcomes associated with persistent speech disorder. *International Journal of Language & Communication Disorders, 56*(2), 299–312. https://doi.org/10.1111/1460-6984.12599

Ziegler, W., Lehner, K., & KommPaS Study Group. (2021). Crowdsourcing as a tool in the clinical assessment of intelligibility in dysarthria: How to deal with excessive variation. *Journal of Communication Disorders, 93,* Article 106135. https://doi.org/10.1016/j.jcomdis.2021.106135