

Research Article

Auditory Perception and Ultrasound Biofeedback Treatment Outcomes for Children With Residual /ɹ/ Distortions: A Randomized Controlled Trial

Jonathan L. Preston,^a Elaine R. Hitchcock,^b and Megan C. Leece^a

Purpose: This study evaluated whether outcomes from treatment, which includes ultrasound visual feedback (UVF), would be more or less effective when combined with auditory perception training for children with residual /ɹ/ errors.

Method: Children ages 8–16 years with /ɹ/ distortions participated in speech therapy that included real-time UVF of the tongue. Thirty-eight participants were randomized to speech therapy conditions that included a primary focus on articulation using UVF or a condition that included auditory perceptual training plus UVF (incorporating category goodness judgments and self-monitoring). Generalization of /ɹ/ production accuracy to untrained words was assessed before and after 14 hr of therapy. Additionally, the role of

auditory perceptual acuity was explored using a synthetic /ɹ-/w/ continuum.

Results: There was no difference between the treatment groups in rate of improvement of /ɹ/ accuracy (increase of 34% for each group; $p = .95$, $\eta_p^2 = .00$). However, pretreatment auditory acuity was associated with treatment progress in both groups, with finer perceptual acuity corresponding to greater progress ($p = .015$, $\eta_p^2 = .182$).

Conclusion: Similar gains in speech sound accuracy can be made with treatment that includes UVF with or without auditory perceptual training. Fine-grained perceptual acuity may be a prognostic indicator with treatment.

Supplemental Material: <https://doi.org/10.23641/asha.11886219>

In American English, /ɹ/ distortion is among the most common residual speech errors (RSEs). Distortions of /ɹ/ can be a result of ineffective articulatory movements resulting in productions that listeners perceive to be unclear. Thus, treatment strategies frequently aim to address speech movements for /ɹ/. Specifically, one articulatory approach to speech therapy for RSE-/ɹ/ is ultrasound visual feedback (UVF), which provides opportunities for real-time feedback of tongue movements during speech; this visual information enables more explicit cueing and self-monitoring (Preston, McAllister Byun, et al., 2017). A treatment program that includes UVF has been found to

facilitate correct /ɹ/ productions in a number of studies (Adler-Bock et al., 2007; McAllister Byun et al., 2014; Modha et al., 2008; Preston et al., 2014, 2018). As with any intervention, treatment programs that include UVF are not effective for all children (Preston et al., 2018; Preston, Leece, & Maas, 2017), possibly because poor auditory perception of /ɹ/ may also be a factor underlying /ɹ/ misarticulations (Shuster, 1998). Adding auditory perceptual training to speech production-focused programs has been found to enhance treatment outcomes in preschool children with speech sound disorders (Rvachew et al., 2004). Enhancing awareness of both the articulatory requirements (with UVF) and auditory perceptual aspects of /ɹ/ may be beneficial for school-age children with RSEs to maximize treatment outcomes. This study therefore explores whether auditory perceptual training can enhance speech sound learning when paired with an articulatory-focused treatment for school-age children with /ɹ/ misarticulations and whether auditory perceptual skills relate to treatment progress.

There are several subtypes of speech sound disorder, and even among children with RSEs, there can be variation in developmental history, etiology, psycholinguistic

^aDepartment of Communication Sciences and Disorders, Syracuse University, NY

^bDepartment of Communication Sciences and Disorders, Montclair State University, NJ

Correspondence to Jonathan L. Preston: jopresto@syr.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Chao-Yang Lee

Received July 9, 2019

Revision received September 8, 2019

Accepted November 15, 2019

https://doi.org/10.1044/2019_JSLHR-19-00060

Disclosure: The authors have declared that no competing interests existed at the time of publication.

profiles, or distortion error types. For example, RSEs may be present following a period of otherwise typical speech development or following a period of delayed phonological acquisition (Flipsen, 2015; Shriberg, 1993, 1994; Shriberg & Kwiatkowski, 1994). These differences have been posited to relate to different etiological factors (Karlsson et al., 2002). Thus, there are likely numerous factors that contribute to RSEs; however, RSE subtyping is complicated by several factors, including the fact that collecting an accurate developmental history may not be possible. Further individual variation in children with RSEs affecting /ɪ/ may be evident in the nature of the allophones produced (e.g., substitutions of [w], various distortions of [ɪ], including multiple allophonic variations presenting within an individual child depending on phonetic context). Consequently, we will refer to these children as having “RSEs,” a general term which includes children whose speech histories, etiologies, or error allophones may vary to some degree but whose similarity is speech errors beyond the ages at which speech normalization would be expected. Such errors may continue even in children who have undergone extensive periods of treatment, suggesting the need for alternate forms of intervention, such as visual feedback.

UVF Interventions

In a recent systematic review, Sugden et al. (2019) reported that the effects of UVF have been investigated in over a dozen single-case experimental designs and several case studies, totaling over 100 individuals with speech sound disorders. Treatment that includes UVF has been found to yield improved production of a range of lingual speech sounds (e.g., /d/, /n/, /s/, /z/, /ʃ/, /tʃ/, /dʒ/, /k/, /g/, /l/, /ɪ/; Cleland et al., 2019, 2015; Preston et al., 2013, 2014). Candidate populations may include individuals with Down syndrome (Fawcett et al., 2008), hearing impairment (Bacsfalvi et al., 2007; Bernhardt et al., 2005), and childhood apraxia of speech (Preston et al., 2013; Preston, Leece, McNamara, & Maas, 2017). However, RSEs affecting American English /ɪ/ have been the most commonly studied sound with UVF perhaps due to the prevalence and the clinical challenge in remediating distortions in some children. In particular, the complex lingual requirements for /ɪ/ (including bunched, retroflex, and other complex tongue shapes) can be challenging for clinicians to describe to children and may also be difficult for children to understand given that such movements are concealed within the oral cavity. UVF may enable more explicit understanding of the different movement requirements of the various parts of the tongue for /ɪ/, such as elevation of the anterior tongue (tip/blade/anterior dorsum), lowering of the posterior tongue dorsum, posterior retraction of the tongue root, and elevation of the sides of the dorsum.

The data to date suggest that treatment programs, which include UVF, can yield increases in /ɪ/ accuracy in many children, although there remains an individual variation in treatment response. For example, Preston et al. (2018) reported significant variability in participant treatment

response following 16 treatment sessions, which included UVF and non-UVF speech practice on /ɪ/. Four of 12 children showed an increase above baseline performance in /ɪ/ accuracy on untrained words of less than 5%, four of 12 showed increases of 10%–31%, and the remaining four of 12 had an increase of 53%–91%. Several other studies have reported similar individual variation, with mean increases of about 30%–50% following 14–16 sessions of therapy (McAllister Byun et al., 2014; Preston et al., 2014, 2018; Preston & Leece, 2017; Preston, Leece, & Maas, 2017). Therefore, in an effort to increase treatment responsiveness, additional approaches might be explored in conjunction with the articulatory focus of UVF.

Auditory Perception of Speech in Children with Speech Sound Disorders

Under the assumption that speech sound production involves learning and refinement of multiple levels of representation—including not just articulatory, but also auditory perceptual features (Guenther & Vladusich, 2012; McAllister Byun et al., 2016; Shiller et al., 2010)—current assessment and treatment for speech sound disorders may be improved with an auditory perceptual component. Numerous studies have pointed out that a subset of children with speech sound disorders differ from children with typical speech in their auditory perception of the speech sounds that they produce in error (Cabbage et al., 2016; Hearnshaw et al., 2018; Hoffman et al., 1985; Nijland, 2009; Rvachew et al., 2003). Among children with RSE, perceptual differences may be evident at young ages (e.g., in preschool) due to several factors including periods of otitis media or genetic involvement (Shriberg et al., 2010). Additionally, among children with relatively typical phonological development in early childhood, RSEs may be a consequence of poor “phonological attunement” for specific late-developing speech sounds (Shriberg, 1994; Shriberg et al., 2005).

For children with speech sound disorders, including those with /ɪ/ errors, there may be difficulty perceiving errors in their own speech as well as in the speech of others. Specifically, Shuster (1998) reported children with /ɪ/ distortions demonstrated chance level accuracy when asked to judge /ɪ/ in their own incorrect productions. Performance was slightly better when asked to judge the accuracy of incorrect productions in other speaker’s productions but the difference was not statistically significant. The difficulty may be because children with /ɪ/ errors have relatively broad perceptual boundaries between /ɪ/ and its common substitute [w] (Hoffman et al., 1985; Ohde & Sharf, 1988). Moreover, even among children with typically developing /ɪ/ productions, there may be a relationship between subtle acoustic features that are both perceived and produced (McAllister Byun & Tiede, 2017). Therefore, within children with RSE-/ɪ/, individual differences in auditory perceptual acuity may be related to their learning to produce /ɪ/. Indeed, there may be a bidirectional relationship such that refining auditory perception may facilitate production

learning, and production learning may enhance auditory perception (Sénéchal et al., 2004).

Although auditory perception of speech may be weak in some children with RSE-/ɪ/ and clinical intuition has led many clinicians and researchers to advocate for various forms of “ear training” (Van Riper & Erickson, 1996), there are few empirically tested options for treating auditory perception for children with RSE-/ɪ/. Among younger (preschool-age) children with speech sound disorders, category goodness judgment training has been shown to be effective at enhancing speech production training (Jamieson & Rvachew, 1992; Rvachew, 1994; Rvachew et al., 1999, 2004). Category goodness judgment training requires that children judge the accuracy of others’ productions as “good” or “not good” exemplars of words (e.g., “Was that a good way for this person to say ‘rat?’”). Such training may help to refine children’s identification of speech sound errors in others, which may therefore facilitate self-awareness of sound errors. Although category goodness training has been found to be effective with preschool-age children, no studies have explicitly tested this approach with school-age children with RSE-/ɪ/. Furthermore, children with RSEs may differ in the degree to which speech perception skills are affected or the degree to which they respond to perceptual training, perhaps due to variation in etiology and/or developmental history.

An additional approach to facilitate auditory perceptual awareness in children with RSE-/ɪ/ is to engage in self-monitoring, which involves category goodness ratings of one’s own speech (e.g., “Do you think you just said ‘rat’ correctly?”). Several small-scale studies have shown that self-monitoring may enhance generalization of speech sound learning in children with distortions of late developing sounds such as /ɪ/ and /s/ (L. K. Koegel et al., 1986; R. L. Koege et al., 1988; Ruscello & Shelton, 1979), although no randomized controlled trials have tested this approach. However, given the reported individual differences in treatment response using perceptual training, it is unclear whether such training would be effective for all children (Gray & Shelton, 1992; Wolfe et al., 2003).

Research Questions

The primary research question was whether speech sound learning during a treatment program that includes UVF (which emphasizes articulatory aspects of speech sound production) could be enhanced by adding auditory perceptual training. It was hypothesized that treatment that included articulatory (UVF) and auditory perceptual training (P + UVF, including category goodness judgment and self-monitoring) would result in greater accuracy in /ɪ/ production than an approach that was only articulatory in nature.

The second research question addressed the question of whether pretreatment perceptual acuity predicted treatment outcomes for children with RSE-/ɪ/. It was hypothesized that children with poorer auditory perceptual acuity would make less progress in a production-oriented treatment such

as UVF than children with more refined auditory perceptual acuity; however, perceptual skills may interact with the type of treatment children receive (i.e., UVF only or P + UVF).

Method

Participants

Participants were recruited through announcements sent to local schools and clinics requesting referrals for children with /ɪ/ distortions. To be included in the study, children had to be between 8 and 16 years of age and demonstrate /ɪ/ errors/distortions. Individuals with hearing loss and structural abnormalities such as cleft palate or developmental disabilities (such as autism or Down syndrome) were not eligible to participate in the study. Participants were recruited in two locations: Montclair, NJ, and Syracuse, NY. Participants provided assent, and parents provided written consent. The study procedures were approved by the institutional review boards at Syracuse University and Montclair State University.

Additionally, children referred to the study were screened to determine eligibility for treatment. Participants were required to meet the following criteria: (a) pass a pure-tone hearing screening at 20 dB at 500, 1000, 2000, and 4000 Hz; (b) a percentile score of 7 or below on the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); (c) a *t* score of 37 or higher on the Matrix Reasoning subtest of the Wechsler Abbreviated Scales of Intelligence (Wechsler, 2011); (d) a scaled score of 6 or higher on both the Recalling Sentences and Formulated Sentences subtests of the Clinical Evaluation of Language Fundamentals–Fifth Edition (Wiig et al., 2013); (e) a standard score of 80 or higher on the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007); and (f) an /ɪ/ accuracy below 25% on a 125-item probe list.

In a second pretreatment visit, further descriptive testing was completed to characterize speech and language skills. These tasks included (a) the LinguSystems Articulation Test (Bowers & Huisingsh, 2011), (b) a stimulability probe assessing the ability to imitate /ɪ/ in 11 different syllables elicited three times each, and (c) the Elision and Phonemic Isolation subtests of the Comprehensive Test of Phonological Processing–Second Edition (Wagner et al., 2013).

Additionally, to assess auditory perception related to /ɪ/, an eight-step synthetic continuum from /ɪ/ to /w/ was administered. The stimuli used in this assessment have been developed and validated in a previous study to detect individual differences in children’s perceptual acuity and its relationship with production of /ɪ/ (McAllister Byun & Tiede, 2017). The stimuli, as described by these authors, were synthetically modified tokens from an 8-year-old female speaker’s productions of “rake” and “wake”; a 10-step continuum was derived from his “rake” production by morphing 13 linear predictive coding coefficients in equivalent intervals toward the coefficient values for “wake” (with fundamental

frequency and release noise of /k/ held constant). The end points were used to familiarize the child with the task; prior to the experimental trials, participants completed a practice block where the end point stimuli of “rake” and “wake” were presented four times each in random order in order to familiarize the participants with the stimuli presentation and mouse click response. Feedback on accuracy was provided during familiarization trials but not during experimental trials. The eight intermediate steps between the end points were used during the experimental trials to assess acuity. Each of these eight steps between “rake” and “wake” were presented 10 times each in random order for a total of 80 trials; the participant categorized each token by clicking on the words “rake” or “wake” with a mouse. The primary measure derived from this task was acuity (boundary width), which is the distance in continuum steps between the 25% and 75% probability points on the fitted logistic function for each participant. This acuity measure was used to describe the participant’s perceptual skills and was administered both pre- and posttreatment.

Randomization and Group Comparisons Before Treatment

This was a parallel group randomized controlled trial. After eligibility was determined, participants were randomized using a coin flip (conducted by the clinician who had contact with the family). This assignment allowed us to obtain approximately a 1:1 allocation to the two groups: UVF or P + UVF. Performance of the two groups on the pretreatment tasks and demographic information are reported in Table 1. Correlations among the pretreatment variables and correlations with posttreatment outcomes are presented in Supplemental Material S2.

A total of 52 participants were recruited and tested, 14 were ineligible, 38 were randomized to treatment, and

36 completed treatment and posttreatment visits (18 per group). Figure 1 outlines enrollment, allocation, follow-up, and analysis according to Consolidated Standards of Reporting Trials guidelines (Schulz et al., 2010).

Treatment

In both conditions, 14 treatment sessions were provided with a target schedule of two sessions per week. All treatment sessions were delivered by American Speech-Language-Hearing Association–certified speech-language pathologists (SLPs) who were trained in the protocol. Each child was assigned one SLP, although clinician substitutions occasionally occurred to accommodate participants’ schedules. There were a total of seven clinicians who provided treatment in each condition (P + UVF was provided by four SLPs in Syracuse and three in Montclair, whereas UVF was provided by five SLPs in Syracuse and two in Montclair). Participants did not receive additional treatment on /ɹ/ while enrolled in the study.

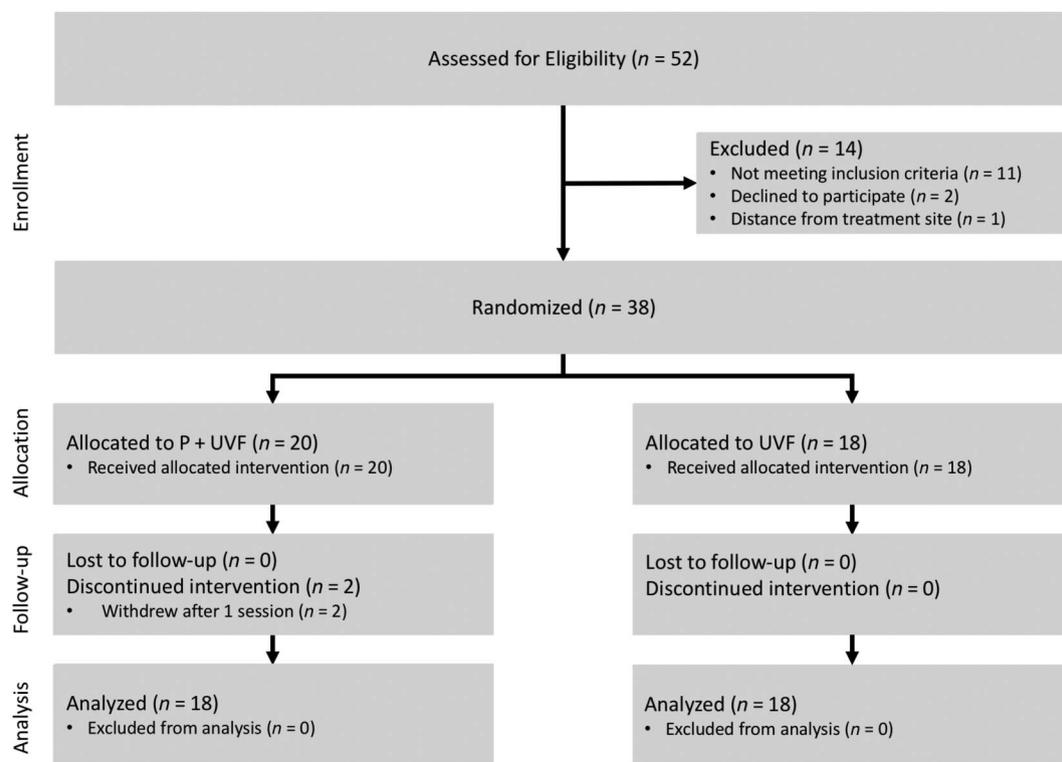
Sessions were structured to include prepractice (6–8 min) followed by practice (which continued for 162 trials or 45 min, whichever occurred first). For all participants, four treatment contexts were targeted in the first seven sessions, including one exemplar of each of the following: /ɹ/ + front vowel (i.e., one from /ɹi/, /ɹe/, /ɹaɪ/), /ɹ/ + back vowel (i.e., one from /ɹɑ/, /ɹo/, /ɹu/), front vowel + /ɹ/ (i.e., one from /iɹ/, /eɹ/), and back vowel + /ɹ/ (i.e., one from /ɑɹ/, /oɹ/). As treatment progressed past the syllable level, the selected syllables were embedded within words at the word, phrase, and sentence level. In the first seven sessions, UVF was provided on 88% of practice trials (with the remaining items practiced in a traditional manner with no visual feedback of the tongue). To facilitate generalization, Sessions 8–14 featured five additional targets, including /ʒ/, as well as a second exemplar of /ɹ/ + front vowel, /ɹ/ + back vowel,

Table 1. Pretreatment comparison of the two groups on demographic measures and on speech and language tasks.

| Variable | UVF group | P + UVF group |
|--|------------------|------------------|
| Gender | 9 male, 9 female | 14 male 6 female |
| Age (years;months) | 10;2 (28 months) | 10;3 (27 months) |
| GFTA-2 percentile | 2.4 (2.3) | 2.9 (2.5) |
| WASI-2 Matrix Reasoning <i>t</i> score | 55.1 (13.2) | 53.3 (9.1) |
| CELF-5 Formulated Sentences scaled score | 12.3 (3.2) | 11.5 (2.8) |
| CELF-5 Recalling Sentences scaled score | 12.0 (3.6) | 11.1 (3.4) |
| PPVT-4 standard score | 117 (14) | 116 (18) |
| LinguiSystems Articulation Test percentile | 0.56 (1.5) | 0.88 (2.29) |
| Pretreatment word probes percent /ɹ/ correct | 17.2 (12.5) | 15.5 (13.0) |
| Stimulability percent /ɹ/ correct on | 8.7 (18.1) | 5.7 (11.0) |
| CTOPP-2 Phonemic Isolation scaled score | 10.1 (1.6) | 9.3 (2.3) |
| CTOPP-2 Elision scaled score | 11.2 (2.1) | 10.5 (1.6) |
| /ɹ/-/w/ Acuity | 2.5 (1.8) | 2.4 (1.8) |
| Percent of participants who had received prior /ɹ/ therapy | 67% | 90% |

Note. Data represent all randomized participants prior to treatment. UVF= ultrasound visual feedback; P + UVF = perception training plus ultrasound visual feedback; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; WASI-2 = Wechsler Abbreviated Scales of Intelligence–Second Edition; CELF-5 = Clinical Evaluation of Language Fundamentals–Fifth Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; CTOPP-2 = Comprehensive Test of Phonological Processing–Second Edition.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram depicting the enrollment, allocation, follow-up, and analysis process for the randomized controlled trial. CONSORT diagram based on Schulz et al. (2010). UVF = ultrasound visual feedback group; P + UVF = perception training plus ultrasound visual feedback group.



front vowel + /ɪ/, and back vowel + /ɪ/. In these last seven sessions, biofeedback was provided on only 44% of trials (Preston et al., 2018). Visual feedback of the tongue was provided using either a Micrus or EchoBlaster 128 Ultrasound (Teleded Medical Systems), which the participant held beneath the chin (with occasional positioning cues or support from the clinician to ensure the images were sufficiently clear and properly oriented). Visual feedback included both sagittal view (imaging the tongue from tongue tip/blade to root) and coronal (which allowed for imaging the sides of the tongue) at the clinician's discretion.

Both the prepractice and practice phases differed for the two treatment conditions (see Condition Differences section below). However, in both conditions, practice trials were guided by the Challenge Point Program (CPP),¹ a research-developed open-source program. The CPP software presented stimuli and prompted the clinician when to provide feedback. Both the stimulus complexity and the feedback were adaptive based on the participant's performance. The CPP software began with syllable-level stimuli. After each trial, the clinician input her or his rating of the child's

accuracy (1 = *correct* or 0 = *incorrect*). After each block of six trials, CPP tallied the results and determined whether the subsequent block of six trials should increase in level (based on at least five of six correct), remain at the same level (if four of six correct), or decrease in level (if three or fewer correct). Higher levels included increases in stimulus complexity (from syllables to one-syllable words, one-syllable words with competing /l/ or /w/, two-syllable words, two-syllable words with competing /l/ or /w/, phrases, or sentences). To enhance speech motor learning, higher levels also prompted a reduction in the frequency of clinician-provided knowledge of results feedback decreasing from four to three to two trials per block of six trials (Maas et al., 2008).

Condition Differences

Although both treatment conditions included an articulatory focus with UVF training, they differed in the inclusion of auditory perception training during each session's prepractice and practice. During prepractice, children in the UVF condition engaged only in articulatory training that included real-time ultrasound images of the tongue for 6–8 min. Phonetic cues and shaping strategies were used in conjunction with the visual feedback. During prepractice, the clinician was free to choose whatever stimuli (syllables, words, phrases) were deemed appropriate

¹The challenge point concept is based on motor learning research by Guadagnoli and Lee (2004) and was more recently applied to speech motor learning by Rvachew and Brosseau-Lapr e (2012). It is a free and open-source program (available at <http://cpp.umd.edu/>).

for the child and cues were required to focus on the visual display of the tongue.

However, children in the P + UVF condition completed a prepractice that emphasized category goodness judgment but no articulatory training. Each session, a researcher-developed perceptual training software program composed of natural recordings of correct and misarticulated /ɪ/ was used to present 50 recordings of single words with /ɪ/ in various word positions. The participant indicated via mouse click whether the /ɪ/ was correct or incorrect and feedback was immediately provided about their judgment relative to the consensus judgment for that token. Tokens could be replayed up to five times before a judgment was made. Seven different perceptual training modules (each containing 50 words) were developed, and each module was presented for two consecutive sessions. During treatment, the seven perceptual training modules were presented in the order of word-initial /ɪ/ singleton: List A, postvocalic /ɪ/: List A, word-initial /ɪ/ singleton: List B, postvocalic /ɪ/: List B, mixed word position: List A, mixed word position: List B, mixed word position: List C. Thus, in P + UVF, prepractice focused exclusively on listening for correct and distorted /ɪ/ productions and no production training was provided. With respect to the stimuli used in training, the consensus rating of at least four listeners was used to determine whether a token was judged as “correct” or “incorrect.” Across the seven modules, participants were exposed to a total of 74 different speaker voices of children and adults, speaking 132 different lexical items, including pre-, inter-, and postvocalic contexts, 42.5% of which were judged to be correct and 57.5% were judged to be incorrect. Stimuli were collected from children and adult participants who gave permission to use their recordings from four different locations in the United States (Montclair, NJ; Syracuse, NY; New Haven, CT; and Cincinnati, OH). Some correct tokens came from children whose speech had improved following treatment and some were from individuals with no histories of speech disorders. Stimuli were presented to participants as wav files sampled at 22,050 Hz and normalized to 70dB SPL, but participants were allowed to adjust the volume to achieve a comfortable listening level.

During structured practice with the CPP software, the two conditions differed only in whether children were required to self-rate the accuracy of their productions. In the UVF condition, participants practiced /ɪ/ as described above (see Treatment) and feedback on the accuracy of productions was provided only by the treating clinician. However, children in the P + UVF condition were prompted by the software to self-rate the accuracy of their productions on 50% of trials. Thus, during trials in which the clinician provided feedback, the child first rated the accuracy of his or her productions, followed by clinician agreement/disagreement of the child’s self-rating. This was intended to provide an opportunity for the clinician to help hone the child’s perception of his or her own /ɪ/ productions. During the self-monitoring trials, children were expected to focus on /ɪ/ accuracy based on auditory percept, rather than visual or

somatosensory information (e.g., if a production “looked” like a correct tongue shape on the ultrasound display but still sounded distorted, the token was to be judged as incorrect).

Treatment Fidelity

Prior to treatment, clinicians were trained in the protocol, in the use of ultrasound and CPP software, and in cueing strategies through meetings with the first and second authors, and clinicians completed mock treatment sessions prior to treatment. Furthermore, an investigator or clinician who was experienced with the procedures sat in during at least the first three treatment sessions for each clinician to monitor adherence to the protocol and to provide immediate feedback, and check-ins every other week were used to address questions or concerns related to the protocol.

Following treatment, video recordings of the computer screen, which included the ultrasound images and CPP software, were reviewed by research assistants for fidelity checking for a total of 76 treatment sessions. The CPP software prompted the treating clinician to provide either knowledge of results feedback (i.e., indicating the production was correct or incorrect) or knowledge of performance feedback (i.e., information about articulatory actions that resulted in correct or incorrect production; Maas et al., 2008). Based on all trials reviewed in these 76 sessions, knowledge of results feedback was provided as expected on 95% of trials (i.e., provided when prompted by the software, and not provided when not prompted by the software). Similarly, clinicians provided knowledge of performance feedback as expected for 96% of trials. For children in the P + UVF condition, the CPP software also cued the clinician to elicit a self-rating from the participant, which was properly elicited on 88% of trials. Additionally, ultrasound feedback was prompted by the software on select trials as described above, and all trials (100%) that were reviewed for fidelity were delivered with (or without) visual feedback as expected.

With respect to fidelity to the prescribed schedule, treatment was intended to be completed twice per week, with 14 sessions completed in 7 weeks. The mean number of weeks from Session 1 to Session 14 was 7.71 weeks ($SD = 1.23$, range: 6.14–11.43 weeks).

Probe Data Coding and Analysis

The primary outcome of the study was performance on word-level probes assessing untrained /ɪ/ words. Children were required to read a list (or name pictures, if they were poor readers) consisting of 125 words, which included 25 /ɪ/ onset singletons, 50 /ɪ/ onset clusters, 25 postvocalic /ɪ/, and 25 /ɛ/. Probe words were recorded as .wav at 44100-Hz sampling rate with no visual feedback.

Probes were collected during assessment sessions before treatment, midway through treatment (before treatment Session 8), within 1 week after the final treatment session, and at a 2-month follow-up. Recordings were labeled

in Praat (Boersma & Weeninck, 2018) using TextGrids to identify all 125 words. Individual words were then saved as separate .wav files, normalized to 70 dB, and randomized. Probe words from all pre-, mid-, and posttreatment sessions were randomly assigned to modules of 100 words. Four listeners (undergraduate or graduate students trained in articulatory phonetics and speech sound disorders) rated each item as a correct or incorrect /ɪ/ production. For each word that was rated, listeners were blind to treatment time (pre-, mid-, or posttreatment) and treatment condition. Praat software was used for randomized stimulus presentation and response recording. Raters completed all blocks in a self-paced fashion and were allowed to replay each token a maximum of four times before selecting an accurate or inaccurate rating. The percent of correct /ɪ/ productions for each child, averaged across four listeners, was then used as the primary dependent variable. That is, for each day on which a probe was administered, 125 words were recorded and later rated independently by four listeners, resulting in a percent correct /ɪ/ that was the average of the four listeners. Treating clinicians did not conduct probe ratings and were therefore blind to performance on the progress monitoring probe.

Reliability

Although the average across four listeners was used as the primary outcome of /ɪ/ accuracy, we also explored reliability among listeners. Interrater reliability was analyzed using Fleiss' kappa, which was found to be kappa = .44 ($p < .00$, 95% confidence interval [CI] [.435, .445]). This is in the moderate range of agreement. The four raters had unanimous agreement on 53.4% of tokens, and three of four raters were in agreement on an additional 32.7% of tokens; on the remaining 13.9% of tokens, the four raters were evenly split in their judgment of /ɪ/ accuracy.

Data Analysis

Analyses were conducted in SPSS 24.0.0 (IBM Corp., 2016) with the alpha level set at .05. The outcome variable was posttreatment /ɪ/ accuracy (percent correct based on four listeners' ratings on the probes). To address the first hypothesis regarding whether perceptual training could enhance treatment outcomes when paired with a motor-based treatment such as UVF, a two-group analysis of covariance (ANCOVA) was conducted to predict /ɪ/ accuracy from treatment group (UVF vs. P + UVF), while controlling for pretreatment /ɪ/ accuracy (O'Connell et al., 2017).²

²Thirty-eight participants were randomized, and 36 completed treatment. Two participants withdrew after the first treatment session. An intention-to-treat analysis was conducted using the group mean for missing data points for the two participants who withdrew prior to completion. The conclusions of all analyses were unchanged. As only pretreatment /ɪ/ accuracy data and perceptual acuity data were available for these two participants, we elect to report analyses from only those 36 participants who were exposed to at least half of the treatment sessions.

To address the second hypothesis regarding whether pretreatment auditory perceptual skills predicted treatment progress, we first computed a partial correlation between pretreatment /ɪ/-/w/ perceptual acuity³ and posttreatment /ɪ/ accuracy (while controlling for pretreatment /ɪ/ accuracy). Additionally, the ANCOVA was repeated to determine whether posttreatment /ɪ/ accuracy could be predicted from treatment group while controlling for both pretreatment /ɪ/ accuracy as well as pretreatment perceptual acuity. Individual performance is reported in Supplemental Material S1.

Results

Treatment Similarities

Table 1 shows the group comparisons on demographic variables and on pretreatment assessments. Treatment fidelity was similar between groups with respect to the proportion of trials in which appropriate feedback was provided by the clinician (UVF group mean = 96.1%, $SD = 6.9\%$, P + UVF group mean = 94.6%, $SD = 8.2\%$, $t[34] = 0.58$, $p = .568$). There was also no difference between groups in the number of weeks to completion of the 14 treatment sessions (UVF group mean = 7.8, $SD = 1.4$ weeks, P + UVF group mean = 7.6, $SD = 1.1$ weeks, $t[34] = 0.31$, $p = .762$). However, the total number of practice trials differed, with the UVF group attempting more total practice trials during the 14 treatment sessions ($M = 2,181$, $SD = 123$ trials) than the P + UVF group ($M = 1,986$, $SD = 324$ trials, $t[34] = 2.38$, $p = .023$). This difference may be attributed to the time spent engaging in self-monitoring for the P + UVF, which the UVF group did not have to complete.

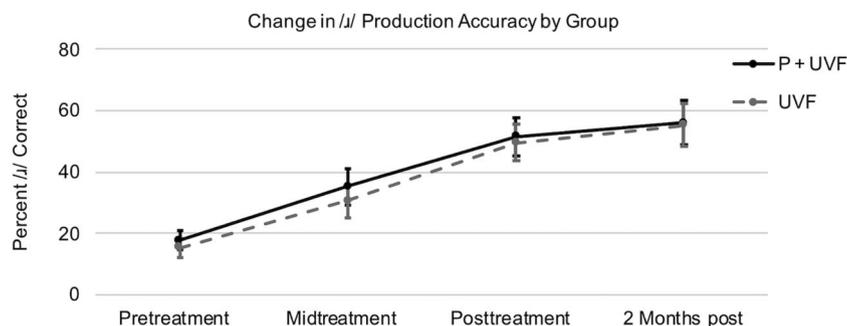
Question 1: Effects of Perceptual Training

Figure 2 shows pre-, mid-, and posttreatment accuracy, as well as 2-month follow-up accuracy; these data reflect the mean of four raters' judgments on the probes for untreated /ɪ/ words in the two treatment groups. As shown in Figure 2, the mean improvements for the two groups are quite similar. The UVF group ($n = 18$) had a mean increase of 34.1% in /ɪ/ accuracy above pretreatment levels ($SD = 19.6\%$, 95% CI [22.4%, 45.9%]), whereas the P + UVF group ($n = 18$) had a mean increase of 33.7% ($SD = 28.9\%$, 95% CI [21.8%, 45.4%]). Following treatment, there was a trend toward continued improvement at the 2-month follow-up visit for both groups.

The results of the two-group ANCOVA ($n = 36$) testing the effects of treatment group on posttreatment /ɪ/ production accuracy (with pretreatment /ɪ/ production

³One participant whose acuity value from the /ɪ/-/w/ perception task exceeded 8 (the maximum acuity detectable with this task) was replaced with 8. Additionally, one participant's data on the perceptual acuity task was missing due to file corruption; therefore, there were 35 participants with available data. Acuity values were not normally distributed. To meet the normality assumptions required for parametric testing, a square root transformation was used for all statistical analyses.

Figure 2. Changes in /l/ production accuracy over time for the ultrasound visual feedback (UVF) and perception training plus ultrasound visual feedback (P + UVF) groups ($n = 18$ per group). Error bars represent one standard error of the mean.



accuracy as a covariate) revealed no significant difference between treatment conditions, $F(1, 33) = 0.004$, $p = .95$, $\eta_p^2 = .00$. Pretreatment /l/ accuracy was not significantly related to posttreatment /l/ accuracy, $F(1, 33) = 3.595$, $p = .074$, $\eta_p^2 = .09$.

Question 2: Role of Pretreatment Auditory Perception on Treatment Outcomes

With respect to the secondary question of whether pretreatment auditory perceptual acuity (derived from participant responses to the *rake-wake listening task*) related to treatment progress, a partial correlation indicated a significant negative relationship between pretreatment perceptual acuity and change in /l/ production ($r = -.424$, $p = .012$, $df = 32$). The negative association suggests that lower pretreatment auditory perceptual acuity values (i.e., sharper perceptual boundaries) were associated with greater improvement in /l/. Furthermore, a two-group ANCOVA was again used to predict posttreatment /l/ accuracy from the main effect of treatment group ($n = 35$), with pretreatment perceptual acuity and pretreatment /l/ accuracy as covariates, as well as the interaction of Treatment Group \times Perceptual Acuity. As with the first model, the main effect of treatment group did not significantly predict posttreatment /l/ accuracy, $F(1, 30) = 0.12$, $p = .73$, $\eta_p^2 = .004$. Treatment group did not significantly interact with perceptual acuity, $F(1, 30) = 0.13$, $p = .72$, $\eta_p^2 = .004$, suggesting that, although auditory perceptual skills were directly targeted in one condition, pretreatment perceptual acuity was not a better predictor of change in either group. Importantly, as was observed with the partial correlation, pretreatment perceptual acuity was a significant covariate in the ANCOVA model, indicating that perceptual acuity was significantly related to change in /l/ production accuracy, $F(1, 30) = 6.67$, $p = .015$, $\eta_p^2 = .182$. Figure 3 shows the association between children's pretreatment perceptual acuity on the /l/-/w/ continuum and their posttreatment /l/ accuracy (while controlling for pretreatment /l/ accuracy).

Finally, because pretreatment perceptual acuity explained some variance in treatment progress, in an ancillary

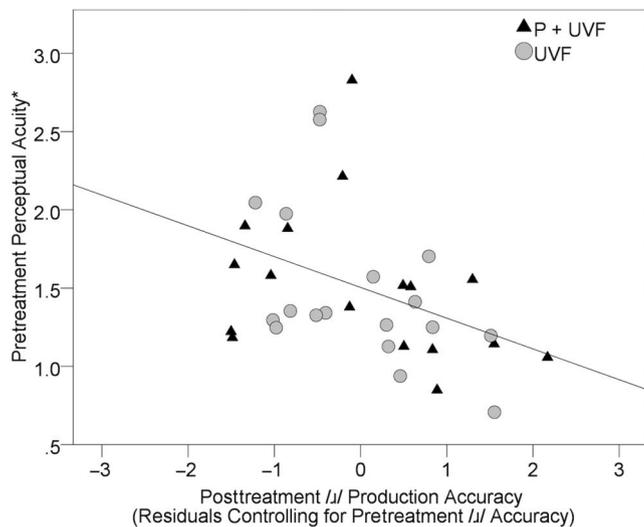
analysis, we explored whether there were any changes in auditory perceptual acuity following treatment. There was a significant decrease (i.e., improvement) in perceptual acuity in the UVF group (mean change from pre- to posttreatment = -1.13 , 95% CI [$-1.55, -0.72$]) and in the P + UVF group (mean change from pre- to posttreatment = $-.60$, 95% CI [$-1.01, -0.19$]).⁴ A two-group ANCOVA was conducted to analyze posttreatment perceptual acuity with pretreatment acuity as a covariate. Results indicated no significant difference between the groups in posttreatment perceptual acuity, $F(1, 31) = 0.65$, $p = .43$, $\eta_p^2 = .020$. Thus, perceptual acuity sharpened to a similar degree following both treatment conditions (see Figure 4). Pretreatment acuity was significantly related to posttreatment acuity, $F(1, 31) = 7.49$, $p = .010$, $\eta_p^2 = .195$, with higher pretreatment acuity values (i.e., poorer acuity) corresponding to greater change in acuity following treatment.

Discussion

When added to a visual feedback treatment program, perceptual training (category goodness judgment and self-monitoring) did not enhance speech production outcomes after 14 hr of therapy. Both the UVF and P + UVF treatments were equally effective at improving /l/ production accuracy. Thus, treatment that includes UVF can be effective in treating /l/ distortions in many children, and perceptual training, as delivered here, may not be necessary to improve outcomes. For both conditions, which included visual feedback and traditional (nonbiofeedback) practice on /l/, there was a mean increase of about 34% on /l/ in untreated words over 14 sessions; specifically, the 95% CI did not encompass 0, suggesting statistically significant improvements in /l/ accuracy following both treatments. The magnitude of improvement in both groups is generally consistent with previous studies exploring the effects of UVF on RSEs affecting /l/ (McAllister Byun et al., 2014; Preston

⁴One participant's pretreatment auditory perceptual data file was corrupted as was one posttreatment file, leaving 17 participants in each group.

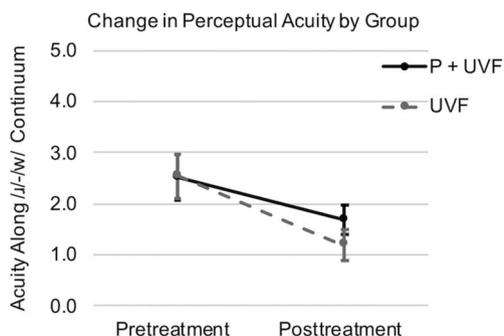
Figure 3. Relationship between pretreatment perceptual acuity and posttreatment /ɪ/ production accuracy (while controlling for pretreatment /ɪ/ accuracy). Acuity (boundary width) is measured by the distance in continuum steps between the 25% and 75% probability points of rating a token as “rake” using a fitted logistic function for each participant. Larger acuity values represent broader perceptual boundaries. *Acuity values are square root transformed. UVF = ultrasound visual feedback group; P + UVF = perception training plus ultrasound visual feedback group.



et al., 2014, 2018). Furthermore, at the 2-month follow-up visit, both groups showed evidence that they had retained similar levels of accuracy (see Figure 2).

Consistent with numerous previous studies, there remains individual variation in treatment response for children with RSE. That is, children with RSE are a heterogeneous group that may include children with or without histories of earlier speech problems, as well as children

Figure 4. Changes in perceptual acuity along a continuum from /ɪ-/w/ for children treated with ultrasound visual feedback (UVF) and perception training plus ultrasound visual feedback (P + UVF). Error bars represent one standard error of the mean. Acuity (boundary width) is measured by the distance in continuum steps between the 25% and 75% probability points of rating a token as “rake” using a fitted logistic function for each participant. Larger acuity values represent broader perceptual boundaries.



with or without perceptual impairments that may be attributed to factors such as genetic causes, histories of otitis media, or psychosocial involvement (Shriberg et al., 2010). In this study, a portion of that variation could be explained by pretreatment auditory perceptual acuity. Children with better (sharper) pretreatment perceptual acuity in the /ɪ-/w/ identification task made better progress than children with poorer (broader) perceptual acuity, regardless of treatment assignment. These data support previous findings that there may be at least a subgroup of children with speech sound disorders for whom auditory perception of speech sounds plays an important role in their speech impairment (Hoffman et al., 1985; Rvachew et al., 2003; Shuster, 1998; Wolfe et al., 2003). Furthermore, because both treatments resulted in changes in the direction of sharper perceptual acuity, it might be the case that perception-production changes are bidirectional. Given the relative novelty of the perceptual task, however, it is unclear whether the changes in perceptual acuity are clinically significant.

Caveats and Limitations

Perceptual training, as delivered here, did not affect the amount of change in /ɪ/ production. This is despite previous findings indicating that category goodness judgment can enhance production treatment outcomes in preschoolers (Rvachew, 1994; Rvachew et al., 2004) and that self-monitoring training can facilitate improvement in /ɪ/ production accuracy in school-age children (L. K. Koegel et al., 1986; R. L. Koegel et al., 1988; Ruscello & Shelton, 1979). However, the perceptual training approaches used in this study did not influence the amount of change in /ɪ/ production. As there were differences in the number of production trials between the groups, this may have influenced the outcomes (n.b., the sessions were structured to adhere to a comparable time limit, as would be typical in most clinical settings, but the self-monitoring took added time, which resulted in fewer practice trials in the P + UVF group). Additionally, it is possible that speech perception deficits did not contribute to the production problems of these school-age children in the same manner as observed in younger children, making perceptual training unnecessary (Wolfe et al., 2003). That explanation is plausible, as preschool-age children may rely on auditory cues to tune their speech systems differently than school-age children. However, given the observed association between perceptual acuity and treatment progress, and because the children with RSE in this study demonstrated poorer acuity than a normative sample reported by McAllister Byun and Tiede (2017), it remains possible that speech perception differences continue to contribute to RSEs. It should be noted that following treatment, the mean perceptual acuity scores were somewhat higher (i.e., poorer) than those reported by McAllister Byun and Tiede for typically developing children; therefore, these children did not fully normalize their auditory perceptual acuity. Indeed, because we did not include a no-treatment

group, we cannot rule out the possibility that changes in auditory perception are due to maturation or practice effects.

It is conceivable that a different dose, timing, or type of perceptual training might influence outcomes. For example, different prepractice training stimuli may be useful to facilitate gains in speech perception. Furthermore, although self-monitoring was implemented at all levels of practice, it may be that self-monitoring is optimally delivered only in the later stages of treatment when children are practicing at phrase or sentence levels (e.g., R. L. Koegel et al., 1988). Alternatively, it may be that other individual factors, such as somatosensory abilities, play a child-specific role in speech learning (Guenther & Vladusich, 2012), which may contribute to variations in treatment response. Specifically, UVF focuses on articulator placement, which speakers ordinarily experience through somatosensory feedback. Individual weighting of auditory and articulatory information may be needed to maximize the effects of treatment. Therefore, there remains a critical need to identify the appropriate balance of skill-based training that can be tailored to a child's needs in order to maximize treatment outcomes.

Pretreatment auditory perceptual skills were related to the amount of change in /ɹ/ production accuracy in 14 sessions, with sharper perceptual acuity corresponding to greater improvement in /ɹ/ production. However, auditory perceptual skills changed in both groups following treatment. It is plausible that this change could be accounted for by children hearing themselves produce a more accurate /ɹ/ during treatment, which may have helped to improve perceptual acuity along the /ɹ/-/w/ continuum. Perception and discrimination of other speech sounds were not included, so it is unclear whether the auditory perceptual changes were limited to this contrast. Furthermore, it is possible that this identification task alone may not adequately capture all perceptual differences; for example, children's perception of speech may vary as a function of a number of factors such as the nature of the errors they perceive or produce or the word position. Auditory discrimination tasks may also be useful in determining individual differences in speech perception and should be considered in future research as part of a larger perceptual assessment battery. However, no task has yet been fully validated with this population, and therefore, specific perceptual assessments require further development. Clinically, the improvements in auditory perceptual acuity in both groups may indicate that the children who began the study with poorer acuity were beginning to show the necessary prerequisite auditory perceptual skills needed to change their production. Thus, additional treatment sessions may have resulted in improved speech sound production as perception skills were changing (Hitchcock et al., 2019). This remains speculative, however, and requires further investigation. Defining the relative sequence of production and perception changes with repeated assessments may eventually help to inform the timing of how or when clinicians address these skills in therapy.

Finally, this study included two active treatment conditions. Because we did not include a no-treatment control group, the effects of spontaneous improvement (e.g., due to maturation) cannot be fully ruled out. Given the persistent nature of RSEs, it is unlikely that the changes in /ɹ/ accuracy or in auditory perception are due simply to maturation, but the current design does not rule out this possibility.

Conclusions

Both UVF and P + UVF training are viable options to facilitate generalization in speech production in children with RSE-/ɹ/. Response to treatment was found to be related to perceptual acuity such that children with more acute auditory perception of rhotic sounds pretreatment achieved greater production gains in treatment compared to those with broader perceptual acuity pretreatment. Individual variation in auditory perception may help to explain some differences in treatment response and may therefore inform prognosis, at least within UVF treatments and perhaps among other treatment approaches as well.

Acknowledgments

This study was supported in part by Syracuse University College of Arts and Sciences, Syracuse University Gerber Auditory Science Grant (PI: J. Preston), and National Institutes of Health Grants R03DC013152 (PI: J. Preston), R15DC016426 (PI: J. Preston), and R01DC013668 (PI: D. Whalen). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Thanks to Tara McAllister and Jose Ortiz who, along with the second author, developed the Challenge Point Program software. Thanks to Tara McAllister for sharing the /ɹ/-/w/ continuum. Thanks to Tara McAllister and Sarah Hamilton-Dugan for sharing stimuli for category goodness training.

References

- Adler-Bock, M., Bernhardt, B., 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\(2007/017\)](https://doi.org/10.1044/1058-0360(2007/017))
- Bacsfalvi, P., Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *International Journal of Speech-Language Pathology*, 9(1), 36–45. <https://doi.org/10.1080/14417040601101037>
- Bernhardt, B., Bacsfalvi, P., Gick, B., Radanov, B., & Williams, R. (2005). Exploring the use of electropalatography and ultrasound in speech habilitation. *Journal of Speech-Language Pathology and Audiology*, 29(4), 169–182.
- Boersma, P., & Weenink, D. (2018). *Praat v 6.0.37*. <http://www.praat.org>
- Bowers, L., & Huisingsh, R. (2011). *LinguiSystems Articulation Test*. LinguiSystems, Inc.
- Cabbage, K. L., Hogan, T. P., & Carrell, T. D. (2016). Speech perception differences in children with dyslexia and persistent

- speech delay. *Speech Communication*, 82, 14–25. <https://doi.org/10.1016/j.specom.2016.05.002>
- Cleland, J., Scobbie, J. M., Roxburgh, Z., Heyde, C., & Wrench, A.** (2019). Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback. *Journal of Speech, Language, and Hearing Research*, 62(2), 229–246. https://doi.org/10.1044/2018_JSLHR-S-17-0360
- Cleland, J., Scobbie, J. M., & Wrench, A. A.** (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, 29(8–10), 575–597. <https://doi.org/10.3109/02699206.2015.1016188>
- Dunn, L. M., & Dunn, D. M.** (2007). *Peabody Picture Vocabulary Test—Fourth Edition*. Pearson.
- Fawcett, S., Bacsfalvi, P., & Bernhardt, B.** (2008). Ultrasound as visual feedback in speech therapy for /t/ with adults with Down syndrome. *Down Syndrome Quarterly*, 10(1), 4–12.
- Flipsen, P.** (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.** (2000). *Goldman-Fristoe Test of Articulation—Second Edition*. AGS.
- 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>
- Hearnshaw, S., Baker, E., & Munro, N.** (2018). The speech perception skills of children with and without speech sound disorder. *Journal of Communication Disorders*, 71, 61–71. <https://doi.org/10.1016/j.jcomdis.2017.12.004>
- Hitchcock, E. R., Swartz, M. T., & Lopez, M.** (2019). Speech sound disorder and visual biofeedback intervention: A preliminary investigation of treatment intensity. In *Seminars in speech and language* (Vol. 40, pp. 124–137). Thieme Medical Publishers.
- Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. H.** (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders*, 50(1), 46–53. <https://doi.org/10.1044/jshd.5001.46>
- IBM Corp.** (2016). *SPSS Statistics for Windows* (Version 24.0).
- Jamieson, D. G., & Rvachew, S.** (1992). Remediating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology*, 16(3), 201–210.
- Karlsson, H. B., Shriberg, L. D., Flipsen, P., Jr., & McSweeney, J. L.** (2002). Acoustic phenotypes for speech-genetics studies: Toward an acoustic marker for residual /s/ distortions. *Clinical Linguistics & Phonetics*, 16(6), 403–424. <https://doi.org/10.1080/02699200210128954>
- Koegel, L. K., Koegel, R. L., & Ingham, J. C.** (1986). Programming rapid generalization of correct articulation through self-monitoring procedures. *Journal of Speech and Hearing Disorders*, 51(1), 24–32. <https://doi.org/10.1044/jshd.5101.24>
- Koegel, R. L., Koegel, L. K., Ingham, J. C., & Van Voy, K.** (1988). Within-clinic versus outside-of-clinic self-monitoring of articulation to promote generalization. *Journal of Speech and Hearing Disorders*, 53(4), 392–399. <https://doi.org/10.1044/jshd.5304.392>
- 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](https://doi.org/10.1044/1058-0360(2008)025)
- 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
- McAllister Byun, T., Inkelas, S., & Rose, Y.** (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language*, 92(1), 141–178. <https://doi.org/10.1353/lan.2016.0000>
- McAllister Byun, T., & Tiede, M.** (2017). Perception-production relations in later development of American English rhotics. *PLOS ONE*, 12(2), e0172022. <https://doi.org/10.1371/journal.pone.0172022>
- Modha, G., Bernhardt, B. M., Church, R., & Bacsfalvi, P.** (2008). Ultrasound in treatment of /r/: A case study. *International Journal of Language & Communication Disorders*, 43(3), 323–329.
- Nijland, L.** (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics*, 23(3), 222–239. <https://doi.org/10.1080/02699200802399947>
- O'Connell, N. S., Dai, L., Jiang, Y., Speiser, J. L., Ward, R., Wei, W., Carroll, R., & Gebregziabher, M.** (2017). Methods for analysis of pre-post data in clinical research: A comparison of five common methods. *Journal of Biometrics & Biostatistics*, 8(1). <https://doi.org/10.4172/2155-6180.1000334>
- Ohde, R. N., & Sharf, D. J.** (1988). Perceptual categorization and consistency of synthesized /r-w/ continua by adults, normal children and /r/-misarticulating children. *Journal of Speech and Hearing Research*, 31(4), 556–568. <https://doi.org/10.1044/jshr.3104.556>
- Preston, J. L., Brick, N., & Landi, N.** (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 22(4), 627–643. [https://doi.org/10.1044/1058-0360\(2013\)12-0139](https://doi.org/10.1044/1058-0360(2013)12-0139)
- 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 & Communication Disorders*, 52(1), 80–94. <https://doi.org/10.1111/1460-6984.12259>
- Preston, J. L., Leece, M. C., McNamara, K., & Maas, E.** (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology*, 26(3), 840–852. https://doi.org/10.1044/2017_AJSLP-16-0155
- 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*, 119, e55123. <https://doi.org/10.3791/55123>
- 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., 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
- Ruscello, D. M., & Shelton, R. L. (1979). Planning and self-assessment in articulatory training. *Journal of Speech and Hearing Disorders, 44*(4), 504–512. <https://doi.org/10.1044/jshd.4404.504>
- Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*(2), 347–357. <https://doi.org/10.1044/jshr.3702.347>
- Rvachew, S., & Brosseau-Lapr e, F. (2012). *Developmental phonological disorders: Foundations of clinical practice*. Plural.
- Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology, 13*(3), 250–263. [https://doi.org/10.1044/1058-0360\(2004\)026](https://doi.org/10.1044/1058-0360(2004)026)
- Rvachew, S., Ohberg, A., Grawburg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech-Language Pathology, 12*(4), 463–471. [https://doi.org/10.1044/1058-0360\(2003\)092](https://doi.org/10.1044/1058-0360(2003)092)
- Rvachew, S., Rafaat, S., & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology, 8*(1), 33–43. <https://doi.org/10.1044/1058-0360.0801.33>
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8*(1), 18. <https://doi.org/10.1186/1741-7015-8-18>
- S en echal, 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.
- Shiller, D. M., Rvachew, S., & Brosseau-Lapr e, F. (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech-Language Pathology and Audiology, 34*(3), 181–192.
- Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36*(1), 105–140. <https://doi.org/10.1044/jshr.3601.105>
- Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4*(1), 38–53.
- Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeney, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilken, 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., & Kwiatkowski, J. (1994). Developmental phonological disorders I. *Journal of Speech and Hearing Research, 37*(5), 1100–1126. <https://doi.org/10.1044/jshr.3705.1100>
- Shriberg, L. D., Lewis, B. A., Tomblin, J. B., McSweeney, J. L., Karlsson, H. B., & Scheer, A. R. (2005). Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of Speech, Language, and Hearing Research, 48*(4), 834–852. [https://doi.org/10.1044/1092-4388\(2005\)058](https://doi.org/10.1044/1092-4388(2005)058)
- Shuster, L. I. (1998). The perception of correctly and incorrectly produced /t/. *Journal of Speech, Language, and Hearing Research, 41*(4), 941–950. <https://doi.org/10.1044/jslhr.4104.941>
- 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>
- Van Riper, C., & Erickson, R. L. (Eds.). (1996). *Speech correction: An introduction to speech pathology and audiology*. Allyn & Bacon.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. R. (2013). *Comprehensive Test of Phonological Processing—Second Edition*. Pro-Ed.
- Wechsler, D. (2011). *Wechsler Abbreviated Scales of Intelligence—Second Edition*. Pearson.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition*. Pearson.
- Wolfe, V., Presley, C., & Mesaris, J. (2003). The importance of sound identification training in phonological intervention. *American Journal of Speech-Language Pathology, 12*(3), 282–288. [https://doi.org/10.1044/1058-0360\(2003\)074](https://doi.org/10.1044/1058-0360(2003)074)