Retroflex Versus Bunched in Treatment for Rhotic Misarticulation: Evidence From Ultrasound Biofeedback Intervention

Tara McAllister Byun[a], **Elaine R. Hitchcock**[b], and **Michelle T. Swartz**[b]

[a]New York University, New York

[b]Montclair State University, Montclair, NJ

Abstract

Purpose—To document the efficacy of ultrasound biofeedback treatment for misarticulation of the North American English rhotic in children. Because of limited progress in the first cohort, a series of two closely related studies was conducted in place of a single study. The studies differed primarily in the nature of tongue-shape targets (e.g., retroflex, bunched) cued during treatment.

Method—Eight participants received 8 weeks of individual ultrasound biofeedback treatment targeting rhotics. In Study 1, all 4 participants were cued to match a bunched tongue-shape target. In Study 2, participants received individualized cues aimed at eliciting the tongue shape most facilitative of perceptually correct rhotics.

Results—Participants in Study 1 showed only minimal treatment effects. In Study 2, all participants demonstrated improved production of rhotics in untreated words produced without biofeedback, with large to very large effect sizes.

Conclusions—The results of Study 2 indicate that with proper parameters of treatment, ultrasound biofeedback can be a highly effective intervention for children with persistent rhotic errors. In addition, qualitative comparison of Studies 1 and 2 suggests that treatment for the North American English rhotic should include opportunities to explore different tongue shapes, to find the most facilitative variant for each individual speaker.

Speech sound disorder in childhood poses a barrier to academic and social participation, with potentially lifelong consequences for educational and occupational outcomes (McCormack, McLeod, McAllister, & Harrison, 2009). Speech sound disorder is estimated to affect up to 10% of preschool and school-age children (National Institute on Deafness and Other Communication Disorders, 1994). Although most of these children go on to develop normal speech by ages 8 to 9 years, a subset of children show continuing errors, often despite months or years of intervention. In a survey of school-based practitioners, 91% of 98

Correspondence to Tara McAllister Byun: tara.byun@nyu.edu.

respondents reported encountering clients whose speech sound errors did not resolve in response to conventional intervention methods (Ruscello, 1995). Survey responses expressed a need for novel, improved intervention methods for persistent speech sound errors, particularly those involving late-developing rhotic and sibilant phonemes. A growing body of evidence suggests that treatment incorporating *visual biofeedback* could fill this need (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; McAllister Byun & Hitchcock, 2012; Modha, Bernhardt, Church, & Bacsfalvi, 2008; Preston, Brick, & Landi, 2013; Ruscello, 1995; Shuster, Ruscello, & Smith, 1992; Shuster, Ruscello, & Toth, 1995). The majority of this evidence comes from case studies, which are classified under Phase I, the lowest level of evidence in clinical outcomes research (Robey, 2004). However, there have been recent efforts to strengthen the evidence base supporting biofeedback intervention, notably through single-subject experimental designs that are classified under Phase II (Preston et al., 2013). In this article we report the results of Phase II clinical research documenting the effects of ultrasound biofeedback treatment for misarticulation of the North American English rhotic.

Characteristics of the North American English Rhotic

Clinicians and clinical researchers working with the North American English rhotic often make a distinction between *consonantal* and *vocalic* variants of the phoneme. Although there is some controversy surrounding this distinction (e.g., Ball, Müller, & Granese, 2013), most studies endorse the notion that consonantal and vocalic rhotics can pattern differently with respect to order of acquisition (e.g., Klein, McAllister Byun, Davidson, & Grigos, 2013) and generalization in treatment (e.g., Curtis & Hardy, 1959; McAllister Byun & Hitchcock, 2012; Preston et al., 2013). We will assume that prevocalic variants are consonantal and therefore use the symbol /ɹ/ in syllable onset position (e.g., *red*, [ɹɛd]; *tree*, [tɹi]). The syllabic variants in stressed and unstressed syllable nuclei are unambiguously vocalic and will be referred to with their appropriate IPA symbols (e.g., *her*, [hɝ]; *water*, [wɔɾɚ]). Finally, we will treat the postvocalic rhotic in words like *care* and *fear* as the vocalic offglide of a rhotic diphthong (e.g., /kɛɚ/, /fɪɚ/). The decision was based on acoustic and articulatory evidence that rhotics in post-vocalic position are more similar to syllabic than onset [ɹ] (McGowan, Nittrouer, & Manning, 2004).

The North American English rhotic is well-known for the challenge it poses in speech acquisition. This difficulty can be attributed at least in part to the complexity of the articulatory configuration used to produce the sound (Gick, Bernhardt, Bacsfalvi, & Wilson, 2008). For most English speech sounds, the tongue forms only one major constriction or narrowing of the vocal tract. However, articulatory descriptions of the North American English rhotic identify two major lingual constrictions: an anterior constriction in which the tongue approximates a point near the hard palate and a posterior constriction in which the tongue root retracts into the pharyngeal cavity (Adler-Bock et al., 2007; Klein et al., 2013). Many speakers also exhibit lateral bracing of the posterior tongue against the rear upper molars, forming a midline groove (Bacsfalvi, 2010). Lip rounding is additionally part of the articulatory configuration for most speakers (Bernhardt & Stemberger, 1998).

In intervention for rhotic misarticulation, the clinician's task is further complicated by the fact that the shape of the anterior constriction for the North American English rhotic varies across speakers (e.g., Delattre & Freeman, 1968). Tongue shapes for this phoneme are commonly divided into two major categories. In the *retroflex* variant, the tongue tip raises and may curl back slightly in the vicinity of the alveolar ridge. In the *bunched* variant, the tongue tip lowers while the tongue body raises to approximate the hard palate. However, it is now known that many adults produce a perceptually appropriate rhotic with tongue shapes that do not fit readily into either category (e.g., Tiede, Boyce, Holland, & Choe, 2004), and many speakers use different tongue shapes across different phonetic contexts (Mielke, Baker, & Archangeli, 2010; Stavness, Gick, Derrick, & Fels, 2012). These variants are perceptually equivalent and appear to be acoustically indistinguishable at the level of the first three formants (resonant frequencies of the vocal tract), although they may be differentiated by the fourth and fifth formants (Zhou et al., 2008).

At the present time, the developmental origin of the observed variation in tongue shapes for the North American English rhotic is not well understood. It is possible that speakers' individual vocal tract morphologies can predispose them to produce one variant or another, either across all phonetic contexts or in a specific subset of contexts. Alternatively, it may be that all variants are equally compatible with all vocal tracts, and as child speakers explore the range of mappings from vocal tract shapes to auditory-acoustic targets, they simply adopt whatever tongue shape they first happen on that achieves the desired auditory target in a particular context (Magloughlin, 2013). This study will provide indirect evidence on this unresolved question by comparing a treatment condition in which all participants were encouraged to adopt a single tongue shape versus a condition in which different tongue shapes were explored on an individualized basis.

Clinicians differ with respect to which tongue-shape variants they choose to cue in treatment for rhotic misarticulation. In an online survey of intervention practices for rhotics, Ball et al. (2013) found that 25% of 200 respondents reported cueing only the retroflex variant, 19% reported cueing only the bunched variant, and 55% reported cueing both types. Ball et al. (2013) additionally identified a number of factors that may be taken into consideration in the clinician's choice of which tongue shape to target. These factors include the relative ease with which different variants might be verbally described, the degree of difficulty child speakers are likely to experience in imitating different tongue postures, and the generalizability of different tongue shapes across phonetic contexts or communicative settings. At this time, however, there is a near-total lack of systematic evidence to indicate whether these factors favor retroflex, bunched, or other tongue shapes for the North American English rhotic. Clinicians choosing which tongue-shape variant to target might also consider the relative frequency with which different variants occur across speakers. In this case, there is evidence favoring bunched tongue shapes. In an ultrasound study of 27 American English speakers, Mielke et al. (2010) found that two produced only the retroflex variant, 11 produced only the bunched variant, and 14 used varying tongue shapes. Similar results were reported by Boyce et al. (2009) in a study of 47 male speakers from Cincinnati, OH.

Visual Biofeedback Intervention for Persistent Speech Errors

Previous research suggests that errors that have not responded to other forms of treatment can sometimes be eliminated through *visual biofeedback intervention*. Biofeedback involves the use of instrumentation to provide real-time information about aspects of speech that the speaker may find hard to perceive under ordinary circumstances, with the goal of bringing these processes under conscious control (Volin, 1998). For instance, a real-time spectrum or spectrogram can be used to provide visual information about the acoustic signal (e.g., McAllister Byun & Hitchcock, 2012; Shuster et al., 1992, 1995), and electropalatography can be used to represent regions of contact between the tongue and palate (e.g., Gibbon, Stewart, Hardcastle, & Crampin, 1999). This study focuses on biofeedback using ultrasound imaging used to reveal the shape and movements of the tongue during speech (e.g., Adler-Bock et al., 2007; Bernhardt et al., 2008; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Shawker & Sonies, 1985). In biofeedback treatment for speech, the clinician models the target sound and calls the client's attention to the sound's appearance on the feedback display. The client then attempts to produce the target sound. With cues from the clinician, the client modifies his or her own output in an effort to achieve a closer match with the visual model.

Ultrasound Imaging of Speech

During ultrasound imaging of lingual articulation, an ultrasound probe is held in a medial position beneath the chin. When the high-frequency waves emitted by the probe encounter a change in density at the boundary between the tissues of the tongue and the air above the tongue, they reflect and are captured by the probe. The reflected sound energy is used to create an image of the surface of the tongue. With multiple images captured per second, a dynamic view of the movements of the tongue can be created. Rotating the probe by 90 degrees makes it possible to shift between sagittal and coronal views of the tongue.

In this study, ultrasound intervention was provided using an Interson SeeMore 7.5–15 MHz multifrequency linear probe. The Interson SeeMore probe is USB-powered, meaning that images are processed and displayed on a linked personal computer, in this case a Dell Latitude E6500 laptop. A scanning depth of 10 cm and a capture rate of 18 frames per second were used. Figure 1 provides examples of images captured with the study equipment.

Ultrasound Biofeedback Treatment: Previous Results

Although somewhat limited in number and scope, previous studies suggest that ultrasound biofeedback intervention can be effective in eliminating persistent errors affecting the North American English rhotic.[1] In a case study of two adolescents with persistent rhotic misarticulation, Adler-Bock et al. (2007) found substantial improvement in accuracy at the word level after 14 one-hour sessions of ultrasound biofeedback therapy. Modha et al. (2008) reported similar gains in a case study of one adolescent who received a combination

---

[1]Omitted in the interest of brevity are the results of several investigations documenting the efficacy of ultrasound biofeedback intervention for individuals with hearing impairment (Bacsfalvi, 2010; Bacsfalvi, Bernhardt, & Gick, 2007; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Shawker & Sonies, 1985) and Down Syndrome (Fawcett, Bacsfalvi, & Bernhardt, 2008).

of ultrasound biofeedback and traditional articulatory treatment. Bernhardt et al. (2008) documented the effects of a brief period of ultrasound biofeedback consultation between two extended intervals of traditional intervention on persistent rhotic errors in 13 children and adolescents ages 7–15 years. After the first phase (traditional treatment only), the perceptually rated accuracy of rhotic production did not differ from pretreatment levels, but a significant increase in accuracy was observed by the end of the study. Finally, Preston et al. (2013) conducted a single-subject experimental study of ultrasound biofeedback treatment in children ages 9–15 years with childhood apraxia of speech (CAS); all participants received treatment targeting rhotics as well as other phonemes. Using multiple baselines across behaviors, Preston et al. (2013) showed that children's progress on treatment targets was systematically linked to the introduction of ultrasound biofeedback. Five out of six participants showed a significant degree of improvement on generalization probes evaluating rhotic production.

The Present Study

Overview. Despite the promising nature of the preliminary results reviewed above, there is an ongoing need for systematic evidence documenting the effects of ultrasound biofeedback treatment, especially for the population of children/adolescents with residual errors affecting rhotics. Our study was designed to address this need by measuring the effects of a structured 8-week program of ultrasound biofeedback treatment using a single-subject experimental design with multiple baselines across participants. The original study design specified a single experimental protocol that would be administered to eight participants, who would be divided into two cohorts for scheduling purposes. However, the first cohort of four participants showed unexpectedly small and inconsistent treatment effects. Instead of continuing to test an unsuccessful treatment, a decision was made to redefine the two cohorts of the original study design as two separate studies (Study 1 and Study 2). Thus, a second cohort of four participants completed 8 weeks of ultrasound biofeedback intervention following a modified treatment protocol. The modification pertained to the tongue shapes cued during biofeedback treatment: In Study 1, all participants were cued to match the same bunched tongue shape, whereas in Study 2, participants were given the opportunity to explore different tongue-shape alternatives. This modification was undertaken based on qualitative observations during Study 1, described below. Although efforts were made to keep experimental conditions constant across the two studies apart from this minimal modification, the studies ultimately did differ in other respects. Thus, Study 1 and Study 2 cannot be regarded as a controlled comparison of ultrasound intervention with and without the option to select individualized tongue-shape targets. Nevertheless, we propose that qualitative comparison of these cases can offer insights that are relevant not only to the study of ultrasound biofeedback but also to broader questions about the acquisition of the North American English rhotic and treatment practices for rhotic misarticulation.

Study 1

Method

Participants—Participants were four monolingual native speakers of English, two males and two females, with an age range from 6;1 (years;months) to 10;3 (*M* = 8;0). Participants

were identified by referral from local speech-language pathologists (SLPs) or by inquiry from parents in response to flyers and postings on electronic distribution lists. Three out of four participants had previously received intervention for rhotic errors, with duration ranging from 7 months to 2.5 years. Two participants had previously received intervention targeting other speech sounds, specifically fricatives, affricates, and /l/. Detailed history data are reported in Table 1. All names reported here and henceforth are pseudonyms.

To be included in the study, participants were required to score within 1 SD of the age-level mean on the Auditory Comprehension subtest of the *Test of Auditory Processing Skills—Third Edition* (Martin & Brownell, 2005), pass a pure-tone hearing test (1000, 2000, and 4000 Hz at 20 dB HL), and show no gross structural or functional abnormality in a screening evaluation of the oral mechanism. Finally, to select participants whose speech was generally intact apart from rhotic errors, the Percentage of Consonants Correct—Revised (PCC–R; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997) was calculated from a 50-utterance spontaneous speech sample. The methodology described by Shriberg et al. (1997) was modified in that rhotic targets were excluded from the calculation of PCC–R, and participants were required to demonstrate a PCC–R of at least 95% after exclusion of rhotic targets.

A final set of inclusionary criteria was based on participants' accuracy in producing rhotics. Stimulability was evaluated with a standard protocol (Miccio, 2002) in which participants were prompted to imitate rhotic sounds in isolation and in syllable-initial, intervocalic, and syllable-final position in the vowel contexts /i/, /ɑ/, and /u/. Each target was elicited three times, and participants who were judged to produce a perceptually correct rhotic in more than 30% of trials were excluded from the study. Finally, participants were required to score below 30% accuracy on a 64-word rhotic probe, which used pictures and orthography to elicit familiar words with consonantal and vocalic rhotics in various phonetic contexts. Vocalic /ɚ/ was probed with four items representing each of the following: (a) stressed /ɝ/, (b) unstressed /ɚ/, (c) /ɑɚ/, (d) /ɛɚ/, (e) /ɔɚ/, and (f) /ɪɚ/. Consonantal /ɹ/ targets, which were elicited in front and back vowel contexts in equal numbers, included singleton /ɹ/ and /ɹ/ clusters featuring alveolar, velar, and labial consonant place. No feedback was provided during probe measures. The complete probe word list is provided in the Appendix.

Study Design—Experimental control was established using a multiple-baseline across-subjects design. Participants were randomly assigned to receive four, five, or six baseline sessions, in which they produced the same 64-word rhotic probe that was administered in the initial evaluation session. After the baseline period, participants completed sixteen 30- to 45-min individual treatment sessions over 8 weeks, including two introductory sessions and 14 biofeedback practice sessions. During treatment, a randomly selected 20-item subset of the rhotic word probe was administered at the start of every third session. Because the words elicited in this measure were never targeted in the context of intervention, these probes were used as a measure of generalization to untreated words over the course of the study. After the end of the treatment period, the full 64-word probe was re-administered to evaluate maintenance of any gains made in therapy. Three maintenance probes were collected over 1.5 weeks from the two female participants. Because of scheduling conflicts, the two male participants completed only two maintenance probes.

Treatment and testing sessions were recorded in a sound-shielded room using the Computerized Speech Lab (CSL) system (KayPentax, Model 4150B) with a 44.1 kHz sampling rate. Participants spoke into a Shure condenser microphone with a mouth-to-microphone distance of approximately 5 in. There was one exception to this recording protocol: Because of the abovementioned scheduling conflicts, the second and final maintenance probe for the two male participants was recorded in a quiet room at their school, using a Marantz PMD660 digital recorder and an Audio-Technica lavaliere microphone.

Instructional Sessions—All study activities followed a standard script, implemented by the first author (a certified clinician with roughly five years of professional experience) or another certified SLP. A trained graduate student was also present to assist with scoring and feedback and to provide prompts to increase fidelity to the standard protocol. Treatment began with two instructional sessions intended to teach participants how to interpret ultrasound images of the tongue and recognize tongue shapes for the North American English rhotic. Training in the first two sessions emphasized two core components of rhotic articulation: tongue root retraction and anterior tongue elevation (Bacsfalvi, 2010). In the first session, the clinician used line drawings and child-friendly language to describe these two lingual constrictions as they appear in sagittal section for correctly produced rhotics. The second session provided an age-appropriate introduction to ultrasound imaging. The clinician presented ultrasound images for various speech sounds, explained that the white line depicts the surface of the tongue, and cued the participant to trace each tongue contour. The child was then cued to produce various speech sounds while viewing the ultrasound image of his or her own tongue. Finally, images and live demonstration were used to familiarize the participant with the ultrasound image of an appropriate tongue shape for rhotic sounds. Verbal instructions highlighted the need to produce distinct anterior and posterior constrictions in place of the undifferentiated or "humped" shape found to be prevalent in children with rhotic misarticulation (Boyce, Combs, & Rivera-Campos, 2011; Klein et al., 2013). As a memory aid, the tongue shape with two constrictions (see Figure 1A) was described as a "horse shape," in contrast to a single hump or "camel shape" (see Figure 1B).

A final component of tongue placement for the North American English rhotic, the midline groove, was introduced in the fourth week of treatment. This delay gave participants time to become familiar with the sagittal perspective before the introduction of the coronal section. The grooved tongue configuration was introduced with a line drawing of the tongue in coronal section with the margins elevated and the midline lowered, described as the "butterfly bite." Participants were also trained to interpret coronal ultrasound images depicting tongue shapes with and without midline grooving.

All models provided during this training and throughout Study 1 featured bunched tongue shapes. The decision to model only the bunched shape was in part a practical one because all of the treating clinicians involved in this study habitually produce only this variant. Moreover, the bunched shape was regarded as a good starting point based on evidence that bunched variants are more common than retroflex variants (Boyce et al., 2009; Mielke et al., 2010).

Prepractice—After the two initial instructional sessions, each treatment session began with a review of pictures and verbal descriptions of tongue placement for the North American English rhotic. To limit cognitive load, a single component of rhotic articulation (tongue root retraction, anterior tongue elevation, or midline grooving) was emphasized in each of the first 3 weeks of treatment. In subsequent weeks, cues for all components were integrated. This verbal review was followed by a 3–5-min "free play" period in which participants could try any manipulations to achieve a better rhotic sound while viewing the ultrasound feedback display. During free play, participants were free to vocalize or use silent tongue shape postures. The images representing correct and incorrect tongue shapes for rhotics (see Figure 1) were displayed as a reference. In the fourth week, when all components of rhotic articulation had been introduced, each child was prompted to silently sustain his or her best approximation of the bunched tongue shape, and an image of this tongue posture was captured. The image was traced onto a sheet protector, and participants were given the option to use this target as a guide during practice. All participants opted to keep the target in place in subsequent sessions.

Treatment Trials—After the prepractice period of each session (excluding the two instructional sessions), participants were cued to produce 30 trials of syllabic /ɝ/, followed by 10 trials each of the syllables /ɹɑ/, /ɹi/, and /ɹu/. Stimuli were elicited in constant order in blocks of five trials. Each block was preceded by a verbal cue reminding the child of one component of correct articulator placement for rhotic sounds. After each block, the clinician provided knowledge of performance feedback in the form of a qualitative comment on the client's speech movements (e.g., "Good job moving your tongue back").

Measurement—Three certified SLP listeners provided perceptual accuracy ratings for participants' productions of untreated rhotic words elicited without feedback in baseline, within-treatment, and maintenance-probe measures. Listeners were trained to rate rhotic sounds as fully accurate (1) or off-target (0), using a strict standard where even distorted sounds with some rhotic quality were rated 0. Before rating experimental stimuli, judges completed a sample set of 100 items that had been rated by an experienced clinician in a previous study. Only clinicians who demonstrated 80% agreement with the previous clinician's ratings were retained as raters.

All target words were isolated from audio recordings of baseline, within-treatment, and maintenance probes and pooled across participants. E-Prime 2.0 software (Psychology Software Tools) was used for randomized, de-identified stimulus presentation and response recording. The full set of items ($n = 2{,}385$) was subdivided into blocks of approximately 200 items. Raters completed all blocks in a self-paced fashion over the span of one or more weeks. Each unique stimulus item was ultimately rated by all three listeners. These three ratings were reduced to a single accuracy score (1 or 0) reflecting the mode across all three listeners for each item.

In the first-pass ratings, pairwise interrater reliability was 81% for Raters 1 and 2, 78% for Raters 1 and 3, and 75% for Raters 2 and 3. Rater 3 was provided with additional training using separate data and subsequently re-rated a 400-word subset of the full stimulus set, including all stimuli for which she had given a different response when Raters 1 and 2 were

in agreement ($n = 328$). After this process, pairwise agreement increased to 86% between Raters 1 and 3 and 83% between Raters 2 and 3.

The rhotic word probe represents a measure of generalization to an untreated context (word level, without biofeedback). It can also be informative to evaluate a participant's performance within the treatment setting, while biofeedback was provided. In our study, all treatment trials were scored online by the clinician delivering the intervention, but these unblinded ratings by a familiar listener are vulnerable to listener bias effects. For a more objective measure of progress within treatment, all trials elicited during treatment sessions ($n = 3,450$) were isolated and presented via E-Prime in a randomized, de-identified fashion for rating by a graduate student who had previous experience rating rhotic sounds but was not involved in treatment delivery for this study. A second student not involved in the study re-rated all items in a blinded fashion. Pairwise interrater agreement exceeded 88%. Because trials produced within treatment and rhotic word probes were rated by different types of listeners (graduate students vs. certified clinicians), no direct comparisons will be drawn between these two sets of ratings. Although it would be ideal to use the same type of listeners in both cases, the large number of within-treatment trials made it impractical to use certified clinician raters in that context.

Analyses—For visual inspection of treatment effects, the percentage of rhotic words rated correct was plotted across baseline, within-treatment, and maintenance probes for each participant. Standardized effect sizes were computed using d_2, Busk and Serlin's (1992) modification of Cohen's d statistic (Beeson & Robey, 2006). In d_2, standard deviations are pooled across baseline and maintenance intervals to reduce the number of cases where effect size cannot be calculated because of zero variance in the baseline period. Following Maas and Farinella (2012), a treatment effect was regarded as clinically meaningful if d_2 exceeded 1.0 (i.e., the difference between pre- and posttreatment means exceeded the pooled standard deviation). Unstandardized effect sizes (mean level difference, i.e., the raw difference between the mean percentage of items rated correct in maintenance vs. baseline intervals) were also calculated because standardized effect sizes can overestimate the magnitude of the effect in cases where variance is very low.

Fidelity—Across Studies 1 and 2, 20% of all sessions were reviewed to evaluate fidelity to the stated treatment protocol (Kaderavek & Justice, 2010). To measure fidelity, the audio record of a treatment session was reviewed by research assistants not involved in treatment delivery. The raters completed a checklist to verify the following aspects of study design: (a) each block of five trials was preceded by a reminder cue, (b) each block consisted of precisely five trials, (c) feedback or other interruptions did not occur within a block, and (d) qualitative knowledge of performance (KP) feedback was provided after each block. Results of the fidelity check for both Study 1 and Study 2 will be reported and discussed in the Results section for Study 2.

Results

Word Probes—The multiple-baseline graphs in Figure 2 depict baseline, treatment, and maintenance intervals; the treated interval is shaded gray. The *y*-axis represents the

percentage of items in each untreated rhotic word probe that were rated correct based on the mode across three blinded listeners. For baseline and maintenance probes, this percentage was calculated over 64 items; for within-treatment probes, this percentage was calculated over 20 items. Vocalic variants are represented with circles and a solid line, and consonantal variants are shown with asterisks and a dotted line.

All participants in Study 1 maintained an adequately stable baseline (<10% mean session-to-session variability over the baseline interval) for both vocalic and consonantal variants. Visual inspection reveals little change in performance on rhotic word probes across baseline, treatment, and maintenance phases. This impression is partially corroborated by the unstandardized and standardized effect sizes reported in Table 2. There are in fact four cases where d_2 exceeded the threshold value of 1.0: for vocalic variants produced by Neville and Mina and for consonantal variants produced by Mina and Gabby. In the cases involving vocalic rhotics, the unstandardized mean difference reveals a change of only 2–3 percentage points, indicating that the standardized effect size has been inflated by low variance. Thus, we will not interpret these changes as clinically meaningful. The two cases involving consonantal /ɹ/ showed larger unstandardized changes (7 and 21 percentage points for Gabby and Mina, respectively), which can potentially be viewed as clinically meaningful. On the other hand, Neville's production of consonantal /ɹ/ showed a decrease in accuracy of comparable magnitude (−10.6 percentage points), yielding a d_2 of −3.2. When standardized effect sizes were averaged across participants and variants, the mean did not exceed the threshold representing clinical significance (d_2 = .84). On balance, the perceptual ratings of word probes in Study 1 paint a disappointing picture of participants' ability to produce rhotic sounds without support after 8 weeks of ultrasound biofeedback treatment.

Within-Treatment Accuracy—The word-probe scores reported in the previous section showed that Study 1 participants made only minimal gains in producing rhotics in untreated words elicited without feedback. However, it is possible that they had acquired some ability to produce correct rhotic sounds but had not yet generalized that skill to a context in which biofeedback was not available. Thus, additional analyses were conducted to examine accuracy during biofeedback trials within the treatment setting. Visual inspection of the longitudinal trajectories plotted in Figure 3 reveals extensive within- and across-participant variability. Only one child, Neville, failed to show progress within the treatment setting as well as on probe measures. Participants Gabby and Mina showed trajectories of increasing accuracy beginning in the 13th practice session. These participants also showed some degree of progress on word probes, but their gains in treatment were primarily observed on vocalic targets, whereas gains on the probe measures were meaningful only for consonantal variants. In addition, the magnitude of change within treatment was much greater for Gabby than for Mina, but the reverse was true with respect to their word-probe progress. The final participant, David, showed substantial increases in accuracy within the treatment setting, but these gains were not sustained over time and never generalized to a context in which biofeedback was not provided.

Discussion

The results of Study 1 provide mixed support for the effectiveness of ultrasound biofeedback in eliciting perceptually correct rhotics from children who have been unable to correctly produce these sounds under ordinary circumstances. Although three out of four participants were able to produce perceptually more accurate rhotic sounds while using ultrasound biofeedback, there was little carryover to correct production in the absence of biofeedback. This finding is consistent with previous research reporting that gains made through biofeedback treatment do not automatically generalize to contexts in which biofeedback is not available (e.g., Fletcher, Dagenais, & Critz-Crosby, 1991; Gibbon & Paterson, 2006; McAllister Byun & Hitchcock, 2012).

A particularly interesting observation emerging from Study 1 pertains to within-treatment gains by Gabby, who began to make sustained progress starting in her 13th practice session. The treating clinician reported a specific event that occurred in that session: In the course of her typical attempts to match the model representing a bunched tongue shape, Gabby happened to produce a retroflex shape, which yielded a perceptually accurate rhotic sound. The clinician then deviated from the standard cues for the bunched rhotic and instead reinforced the retroflex tongue shape. From this point on, Gabby's accuracy in the treatment context increased steadily, reaching a maximum of 48%. This sequence of events suggested that participants might make greater gains in treatment if they were given the opportunity to try a range of tongue shapes. To test this hypothesis, we conducted a second single-subject investigation of ultrasound biofeedback intervention. Study 2 was designed to track Study 1 as closely as possible, with the exception that the one-shape-fits-all articulatory target was replaced with an individualized approach in which tongue-shape targets could be retroflex, bunched, or other tongue shapes that typical adult speakers have been observed to use to produce the North American English rhotic (e.g., Tiede et al., 2004).

Study 2

Method

Participants—Participants were four monolingual native speakers of English, two male and two female individuals, with an age range from 7;8 to 15;8 (*M* = 10;10). Criteria for inclusion in Study 2 were the same as described for Study 1. All participants had previously received intervention targeting rhotics before this study, with the duration of treatment ranging from 1 to 8 years. Three participants had previously been treated for other speech errors, including vowel distortions, /l/, /s/, /z/, /θ/, and /ð/. Detailed characteristics of these participants are reported in Table 3.

Study Design—In Study 2, we followed the same multiple-baseline across-subjects design as Study 1. Participants completed 3, 4, or 5 pretreatment baseline sessions, followed by 17 individual 30-min treatment sessions. The setting and equipment used were the same as in Study 1. Treatment was administered by the second author (a certified clinician with over 19 years of professional experience) or another certified SLP, assisted by a trained graduate student. Treatment sessions in which practice trials were elicited had the same number, duration, and structure across Studies 1 and 2. However, the nature of the preliminary

instructional sessions differed. Whereas Study 1 featured two introductory sessions before the initiation of rhotic practice trials, we added a third session in Study 2. The additional session was used to discuss the range of tongue shapes that can be associated with perceptually acceptable production of the North American English rhotic (e.g., Tiede et al., 2004) and to allow participants to try out different candidate tongue shapes.

Instructional Sessions—In the first session, we provided participants with an introduction to the ultrasound and tongue shapes for rhotics using the same script and materials presented in Study 1. In a new modification, a contextual rhotic probe (Schmidlin & Boyce, 2010)[2] was elicited while the participant's tongue movements were ultrasound recorded, with the screen facing away from the child during recording. This video was collected with the goal of identifying candidate tongue-shape targets for each participant. The first and second authors separately reviewed the videos and identified phonetic contexts in which participants most closely approximated a perceptually accurate rhotic sound. They then compared these approximations against magnetic resonance (MR) images of adult tongue shapes for /ɝ/ (Tiede et al., 2004) and selected three potential target shapes from among the MR images. Two targets were selected to be as similar as possible to the participant's best perceptual approximation of rhotic quality, while the third was chosen as a highly distinct alternative. This exploratory option was included with the rationale that some participants might achieve perceptually correct rhotic production with a tongue posture they had not previously attempted or approximated. After following this selection process independently, the authors conferred and reached consensus on the three MR imaging targets that would be used for each participant. The authors also identified specific strategies and verbal cues that were expected to be most successful in shaping the child's current productions into perceptually accurate rhotics (Schmidlin & Boyce, 2010).

In the second session, participants were familiarized with images of the North American English rhotic in sagittal section. Unlike Study 1, where all participants heard a standard script, this session was dynamic and featured the three tongue shapes individually targeted for each child. The child attempted to match these targets in the context of producing rhotics while viewing his or her own ultrasound image. The clinician offered feedback on perceptual accuracy and tried to elicit more accurate approximations using the child's selected cues. However, there was no requirement to use only the tongue shapes and strategies identified originally; the clinician was free to incorporate any targets or cues found to be facilitative.

In the third instructional session, the coronal view and the midline groove component of rhotic articulation were introduced. This contrasted with Study 1, where this information was deferred until the fourth week of treatment as a means of limiting cognitive load. However, in Study 2 the primary goal was to identify and reinforce any cues that brought a given participant to his or her closest approximation of a perceptually correct rhotic sound, and for some participants, these cues might involve tongue postures seen in coronal section.

---

[2]The contextual probe features both vocalic and consonantal rhotics in various vowel contexts at syllable, monosyllabic word, and multisyllabic word levels. Production of rhotics in the context of potentially competing consonants such as /l/, /w/, and /j/ is probed. Finally, consonantal /ɹ/ is probed in various vowel contexts in initial consonant clusters (/bɹ, fɹ, gɹ, pɹ, kɹ, dɹ, tɹ, stɹ/), which some sources characterize as facilitative of correct rhotic production (e.g., Hoffman, 1983).

The materials and verbal cues used to introduce the coronal view and midline groove were the same as those used in Study 1.

Prepractice—As in Study 1, each treatment session after the initial instructional period began with a review of images and verbal descriptions of appropriate tongue-placement options for rhotics. Cues were still provided in a semistructured fashion, with a single cue (e.g., tongue root retraction) serving as the primary focus in a given week of treatment. However, the free-play period at the start of each session differed in that pre-practice activities were tailored to the individual child, incorporating the cues and tongue-shape targets identified as most facilitative in the evaluation and previous treatment sessions.

Treatment Trials—After the first three instructional sessions, treatment sessions elicited 30 trials of syllabic /ɝ/ and 10 trials each of the syllables /ɹɑ/, /ɹi/, and /ɹu/. The structure of treatment sessions was the same as Study 1 with respect to the order of target elicitation and the nature and schedule of feedback. Practice also incorporated an individualized tongue-shape target traced onto a sheet protector, elicited by prompting each child to sustain the tongue shape judged to yield his or her most accurate approximation of the rhotic sound. For equivalence with Study 1, this target was not elicited until the fourth week of treatment. In following weeks, this target could be used at the child's discretion. As in Study 1, all participants elected to use the target line for the duration of the treatment program. A summary of similarities and differences in the methods and materials adopted in Studies 1 and 2 is provided in Table 4.

Measurement—The measurement protocols adopted in Study 1 were also followed in Study 2, with three exceptions. First, because participants in Study 2 made substantial gains on the probe measures evaluating untreated words produced without feedback, measurement of their progress in individual treatment sessions was judged to be unnecessary; only word-probe results will be reported below. Second, in both studies, binary accuracy ratings were assigned by certified clinicians who listened to individual rhotic words in a blinded, randomized fashion. However, the method of stimulus delivery differed: in Study 1, E-prime was used, whereas in Study 2, stimuli were presented using the online experiment presentation platform Experigen (Becker & Levine, 2010). The latter approach is more convenient because it requires no special software licenses, allowing listeners to rate stimuli from their home computers. The basic mechanism of randomized stimulus presentation and response collection is equivalent across these two platforms. Third, in Study 1 the same three raters listened to every stimulus item. In Study 2, blocks of stimuli were distributed across four raters in such a way that every block was rated by three unique individuals. For every pair of raters, interrater reliability was calculated over the blocks shared between those two individuals. Pairwise interrater reliability ranged from 80.4% to 86.7%.

Analyses—As in Study 1, results of perceptually rated rhotic word-probe measures were interpreted through a combination of visual inspection and calculation of unstandardized and standardized effect sizes. Effect sizes were calculated and interpreted as described above for Study 1.

In Study 2, analyses of perceptual accuracy ratings were supplemented with qualitative inspection of tongue shapes produced before and after ultrasound biofeedback treatment. The contextual probe for ultrasound recording (Schmidlin & Boyce, 2010) was re-elicited at the end of the study, and findings were compared to evaluate whether the tongue shapes judged to be most facilitative for a given participant remained the same from pre- to posttreatment or changed over the course of treatment. This analysis was not included in Study 1 because no contextual probe videos were collected.

Results

Tongue Shapes—Before treatment, two participants (Lilianne and Jordan) were judged to produce their most accurate rhotic approximations with a tongue shape that more closely resembled a bunched versus retroflex variant. A third participant, Philip, produced his best rhotic sounds with an approximation of a retroflex shape. For the final participant, Autumn, retroflex and bunched shapes were judged to be equally facilitative. At the end of the study, three of four participants continued to produce their best approximations using a variant of the tongue shape that was initially judged most facilitative. Only Autumn made a notable change over the course of the study, shifting from free variation between bunched and retroflex shapes to a stable preference for a retroflex tongue shape.

Word Probes—All participants in Study 2 maintained an adequately stable baseline for both vocalic and consonantal targets. In contrast with Study 1, visual inspection of the multiple-baseline graphs in Figure 4 shows a clear and sustained response to treatment in all participants. Participants differed in the relative magnitude of gains on vocalic and consonantal variants. Participants Philip and Lilianne showed greater improvement on vocalic than consonantal targets, whereas Jordan showed the reverse pattern, and Autumn demonstrated an equivalent degree of improvement across consonantal and vocalic targets. Participants also varied in the rate at which treatment gains became evident, with participants Lilianne and Autumn making more immediate progress than Philip and Jordan. Autumn showed a notable decline in accuracy on the word-probe measure administered in Treatment Session 16, but she returned to ceiling levels of accuracy during the maintenance phase. The temporary reversal was informally attributed to a 2-week absence from treatment because of an extended weather emergency.

The unstandardized and standardized effect sizes reported in Table 5 are consistent with the impressions derived from visual inspection. In Study 2, all participants showed standardized mean differences equal to or greater than 1.0 for both vocalic and consonantal targets. Except for the relatively modest gains observed for Jordan's progress on vocalic variants and Philip's progress on consonantal variants ($d_2 = 1.0$ and 1.7, respectively), the observed effect sizes were large to very large (range = 4.0–16.7). Averaging across all participants and targets yielded a mean d_2 of 7.3. Unstandardized effect sizes were also indicative of robust improvement for all participants.

Discussion

Study 2 featured large treatment gains that were replicated across all four participants. These gains were observed on untreated words elicited without biofeedback, indicating that speech

skills acquired through biofeedback treatment can generalize to a broader context. Study 2 thus offers systematic evidence that ultrasound biofeedback treatment, with appropriate parameters of implementation (e.g., with the flexibility to target a tongue shape that is facilitative for the specific speaker), can be a highly effective form of intervention for children with treatment-resistant rhotic misarticulation. This supports previous Phase II research on the efficacy of ultrasound biofeedback treatment for rhotic errors in children with CAS (Preston et al., 2013), as well as earlier Phase I studies documenting the effects of ultrasound intervention for residual rhotic errors (Adler-Bock et al., 2007; Bernhardt et al., 2008; Modha et al., 2008). The present result is particularly striking in light of the fact that participants in Study 2 had received treatment for 1 to 8 years without progress before their success in ultrasound biofeedback intervention.

Fidelity—In Study 1, the primary deviation from the stated protocol involved interruptions during a block, which occurred in roughly 11% of blocks. The studies differed in that most blocks in Study 1 were preceded and followed by qualitative cues and feedback (95% and 90%, respectively), whereas in Study 2, this verbal input was provided in less than half of blocks (40% and 46%, respectively). This difference was attributed to participants' greater accuracy in Study 2 (see Results), which led the clinician to scale back her input. (For full details of the fidelity check, see online supplementary materials.)

General Discussion

The effects produced by ultrasound biofeedback treatment differed strikingly across Studies 1 and 2. By comparing these two studies, we can draw preliminary inferences regarding the optimal parameterization of ultrasound treatment methods. In Study 1, all children attempted to match the same bunched tongue-shape target and received the same standardized placement cues. In Study 2, both tongue-shape targets and cues were tailored to individual participants. Our results offer evidence that, within the context of ultrasound intervention, an individualized program is more effective than a one-size-fits-all approach. We further suggest that the same principle can be expected to apply in the context of rhotic treatment without ultrasound biofeedback. As discussed above, clinicians may consider a variety of factors when making this decision, but to date there has been a lack of evidence to support one alternative or the other (Ball et al., 2013). This study provides empirical evidence that it is not optimal to target a single tongue shape for all clients; instead, clients should be offered opportunities to explore different tongue shapes to find the configuration that is most facilitative of perceptually accurate rhotic sounds. In this respect, our findings are compatible with theoretical work arguing that the targets of speech production are not directly articulatory in nature but may instead reflect individually learned mappings from vocal-tract gestures onto targets defined in auditory-acoustic space (e.g., Guenther, Hampson, & Johnson, 1998). Our finding that outcomes were enhanced when participants were encouraged to use whatever tongue shape was most facilitative of a perceptually correct rhotic quality also resonates with recent treatment research reporting that speech production can be improved through intervention emphasizing the acoustic or auditory properties of a target sound, even in the absence of explicit articulator placement cues (e.g., Rvachew & Brosseau-Lapré, 2012). Our findings also agree with the conclusions reached by

Klein et al. (2013) in an ultrasound study of tongue shapes produced by children acquiring rhotics over a course of traditional treatment.

Finally, these results have implications for our broader understanding of the acquisition of the North American English rhotic. As noted in the introduction, it is not known whether individuals are predisposed by their vocal-tract morphology to adopt a particular tongue-shape variant or whether all variants are equally compatible with a wide range of vocal-tract shapes. Our results suggest that for at least some speakers, tongue shapes for rhotics are not interchangeable. The case of Gabby (Study 1) is particularly suggestive: Although she was unable to produce perceptually accurate rhotics over weeks of treatment targeting a bunched tongue shape, she made striking gains within the treatment setting immediately after the introduction of a retroflex tongue-shape target.

However, several factors limit the strength of the conclusions that can be drawn from the findings reported here. It is not possible to treat Study 1 and Study 2 as a controlled comparison of ultrasound treatment with and without the option to select individualized tongue shapes, because the two studies differed along other parameters. In particular, participants' average age was greater in Study 2 than in Study 1, and older participants may be better able to benefit from biofeedback intervention than younger individuals (McAllister Byun, Maas, & Swartz, 2013). Second, although the same clinician was immediately responsible for most treatment delivery in both Studies 1 and 2, the researcher who supervised the intervention and guided clinical decisions differed across the studies. In Study 1, supervision was provided by the first author, who holds clinical certification but has worked primarily in a research setting, whereas in Study 2, primary supervision was provided by the second author, who is active in research but also has over 19 years of direct clinical experience. This study does not allow us to tease out the relative contributions of these various factors in producing the observed contrast in treatment outcomes. Thus, before strong conclusions can be drawn about the importance of individualized tongue-shape targets in the acquisition and remediation of the North American English rhotic, it will be necessary to follow up on this research in a more systematic fashion. Nevertheless, we maintain that qualitative comparison of Studies 1 and 2 constitutes a useful first step toward evidence-based guidelines for the selection of tongue-shape targets for rhotic intervention, whether in the context of ultrasound biofeedback or in a more traditional treatment approach.

Conclusion

This project was undertaken with the primary aim of using single-subject experimental methods to collect systematic evidence of the efficacy of ultrasound biofeedback intervention for rhotic misarticulation. Because of unexpectedly poor outcomes in the first cohort of four participants, a series of two closely related studies was conducted in place of a single study. The results of Study 2 indicated that ultrasound biofeedback, with appropriate parameters of treatment, can be a highly effective intervention for children whose rhotic errors have not responded to other forms of treatment. As technological advances continue to lower the cost of access to ultrasound imaging, clinicians in educational and private practice settings can reasonably begin to view ultrasound as a feasible option to address the

challenge presented by treatment-resistant speech errors. In addition, qualitative comparison of the results of Study 1 and Study 2 provides evidence that intervention for rhotic misarticulation should include opportunities to explore different tongue shapes to find the most facilitative variant for the individual speaker. Although there is a need for more systematic follow-up research, this insight has potential relevance for all treatment of rhotic misarticulation, both with and without biofeedback.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

Appendix

64-word rhotic probe administered in evaluation, baseline, and maintenance sessions.

| bar	far	her	scrub
bear	fear	pear	shower
bread	four	pray	sir
broom	friend	purr	sore
brown	fries	rain	star
car	frog	read	straw
core	fruit	red	string
crab	fur	rip	strong
crack	gear	rock	tear
crow	grape	root	tiger
deer	green	row	train
door	grew	run	trash
draw	group	scrape	trip
dream	grow	scratch	troll
drum	hair	scream	truck
fair	hammer	screw	water

References

Adler-Bock M, Bernhardt BM, Gick B, Bacsfalvi P. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. American Journal of Speech-Language Pathology. 2007; 16:128–139. [PubMed: 17456891]

Bacsfalvi P. Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. Revue canadienne d'orthophonie et d'audiologie. 2010; 34:206–217.

Bacsfalvi P, Bernhardt BM, Gick B. Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. Advances in Speech–Language Pathology. 2007; 9:36–45.

Ball MJ, Müller N, Granese A. Towards an evidence-base for /r/-therapy in English. Journal of Clinical Speech and Language Studies. 2013; 20:1–23.

Becker, M.; Levine, J. Experigen: An online experiment platform. 2010. Retrieved from https:// github.com/tlozoot/experigen

Beeson PM, Robey RR. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. Neuropsychology Review. 2006; 16:161–169. [PubMed: 17151940]

Bernhardt BM, Bacsfalvi P, Adler-Bock M, Shimizu R, Cheney A, Giesbrecht N, Radanov B. Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. Clinical Linguistics & Phonetics. 2008; 22:149–162. [PubMed: 18253873]

Bernhardt BM, Gick B, Bacsfalvi P, Ashdown J. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. Clinical Linguistics & Phonetics. 2003; 17:199–216. [PubMed: 12858839]

Bernhardt, B.; Stemberger, J. Handbook of phonological development: From the perspective of constraint-based nonlinear phonology. Academic Press; San Diego, CA: 1998.

Boyce SE, Combs S, Rivera-Campos A. Acoustic and articulatory characteristics of clinically resistant /r/. The Journal of the Acoustical Society of America. 2011; 129:2625.

Boyce SE, Tiede MK, Groves-Wright K, Espy-Wilson C, Holland C, Choe K. Dialect distribution of North American English /r/. The Journal of the Acoustical Society of America. 2009; 125:2575.

Busk, PL.; Serlin, RC. Meta-analysis for single-case research.. In: Kratochwill, TR.; Levin, JR., editors. Single-case research design and analysis: New directions for psychology and education. Erlbaum; Hillsdale, NJ: 1992.

Curtis JF, Hardy JC. A phonetic study of misarticulation of /r/. Journal of Speech and Hearing Research. 1959; 2:244–257.

Delattre P, Freeman DC. A dialect study of American r's by x-ray motion picture. Linguistics. 1968; 6:29–68.

Fawcett S, Bacsfalvi P, Bernhardt BM. Ultrasound as visual feedback in speech therapy for /r/ with adults with Down Syndrome. Down Syndrome Quarterly. 2008; 10:4–12.

Fletcher SG, Dagenais PA, Critz-Crosby P. Teaching consonants to profoundly hearing-impaired speakers using palatometry. Journal of Speech, Language, and Hearing Research. 1991; 34:929–943.

Gibbon F, Paterson L. A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. Child Language Teaching and Therapy. 2006; 22:275–292.

Gibbon F, Stewart F, Hardcastle WJ, Crampin L. Widening access to electropalatography for children with persistent sound system disorders. American Journal of Speech-Language Pathology. 1999; 8:319–334.

Gick, B.; Bernhardt, B.; Bacsfalvi, P.; Wilson, I. Ultrasound imaging applications in second language acquisition.. In: Hansen, J.; Zampini, M., editors. Phonology and second language acquisition. John Benjamins; Amsterdam, the Netherlands: 2008. p. 309-322.

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review. 1998; 105:611–633. [PubMed: 9830375]

Hoffman PR. Interallophonic generalization of /r/ training. Journal of Speech and Hearing Disorders. 1983; 48:215–221. [PubMed: 6621014]

Kaderavek JN, Justice LM. Fidelity: An essential component of evidence-based practice in speech-language pathology. American Journal of Speech-Language Pathology. 2010; 19:369–379. [PubMed: 20601624]

Klein HB, McAllister Byun T, Davidson L, Grigos MI. A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. American Journal of Speech-Language Pathology. 2013; 22:540–553. [PubMed: 23813195]

Maas E, Farinella KA. Random versus blocked practice in treatment for childhood apraxia of speech. Journal of Speech, Language, and Hearing Research. 2012; 55:561–578.

Magloughlin LV. An ultrasound study of the acquisition of North American English /r/. In Proceedings of Meetings on Acoustics. 2013; 19:060071.

Martin, NA.; Brownell, R. Test of auditory processing skills. Academic Therapy Publications; Novato, CA: 2005.

McAllister Byun T, Hitchcock ER. Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. American Journal of Speech-Language Pathology. 2012; 21:207–221. [PubMed: 22442281]

McAllister Byun, T.; Maas, E.; Swartz, MT. Motor learning principles in /r/ treatment: Direction of attentional focus. Talk presented at the American Speech-Language-Hearing Association; Chicago, IL.: 2013.

McCormack J, McLeod S, McAllister L, Harrison LJ. A systematic review of the association between childhood speech impairment and participation across the lifespan. International Journal of Speech-Language Pathology. 2009; 11:155–170.

McGowan RS, Nittrouer S, Manning CJ. Development of [a] in young, Midwestern, American children. The Journal of the Acoustical Society of America. 2004; 115:871–884. [PubMed: 15000198]

Miccio AW. Clinical problem solving: Assessment of phonological disorders. American Journal of Speech-Language Pathology. 2002; 11:221–229.

Mielke J, Baker A, Archangeli D. Variability and homogeneity in American English /r/ allophony and /s/ retraction. Laboratory Phonology. 2010; 10:699–730.

Modha G, Bernhardt B, Church R, Bacsfalvi P. Case study using ultrasound to treat /a/. International Journal of Language & Communication Disorders. 2008; 43:323–329. [PubMed: 17852527]

National Institute on Deafness and Other Communication Disorders. National strategic research plan. Department of Health and Human Services; Bethesda, MD: 1994.

Preston JL, Brick N, Landi N. Ultrasound biofeedback treatment for persisting childhood apraxia of speech. American Journal of Speech-Language Pathology. 2013; 22:627–643. [PubMed: 23813207]

Robey RR. A five-phase model for clinical-outcome research. Journal of Communication Disorders. 2004; 37:401–411. [PubMed: 15231420]

Ruscello DM. Visual feedback in treatment of residual phono-logical disorders. Journal of Communication Disorders. 1995; 28:279–302. [PubMed: 8576411]

Rvachew S, Brosseau-Lapré F. An input-focused intervention for children with developmental phonological disorders. Perspectives on Language Learning and Education. 2012; 19:31–35.

Schmidlin, S.; Boyce, SE. Using ultrasound biofeedback for articulation therapy: Selected cases. Poster presented at the Annual Convention of the American Speech-Language Hearing Association; Philadelphia, PA.: 2010.

Shawker TH, Sonies BC. Ultrasound biofeedback for speech training. Instrumentation and preliminary results. Investigative Radiology. 1985; 20:90–93. [PubMed: 3980183]

Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL. The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. Journal of Speech, Language, and Hearing Research. 1997; 40:708–722.

Shuster LI, Ruscello DM, Smith KD. Evoking [r] using visual feedback. American Journal of Speech-Language Pathology. 1992; 1:29–34.

Shuster LI, Ruscello DM, Toth AR. The use of visual feedback to elicit correct /r/. American Journal of Speech-Language Pathology. 1995; 4:37–44.

Stavness I, Gick B, Derrick D, Fels SS. Biomechanical modeling of English /r/ variants. The Journal of the Acoustical Society of America. 2012; 131:EL355–360. [PubMed: 22559452]

Tiede MK, Boyce SE, Holland CK, Choe KA. A new taxonomy of American English /r/ using MRI and ultrasound. The Journal of the Acoustical Society of America. 2004; 115:2633–2634.

Volin RA. A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task. American Journal of Speech-Language Pathology. 1998; 7:81–90.

Zhou X, Espy-Wilson CY, Boyce S, Tiede M, Holland C, Choe A. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. The Journal of the Acoustical Society of America. 2008; 123:4466–4481. [PubMed: 18537397]
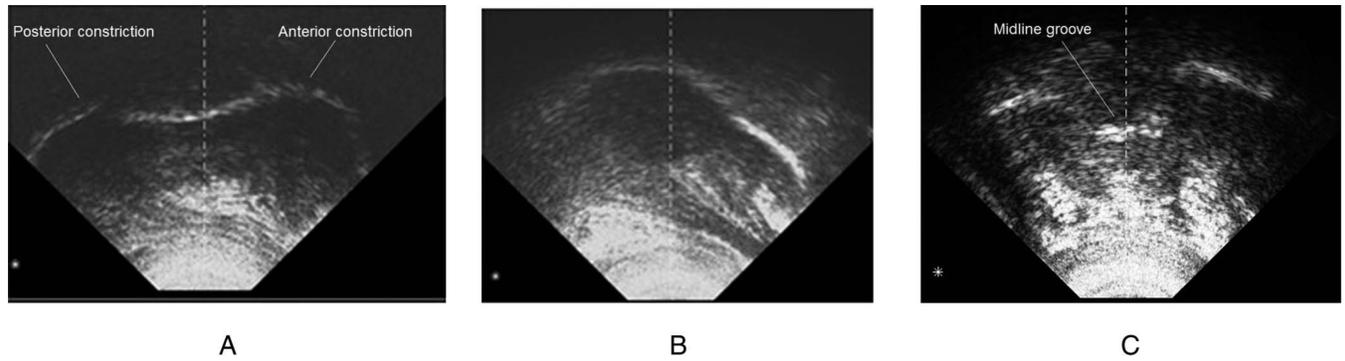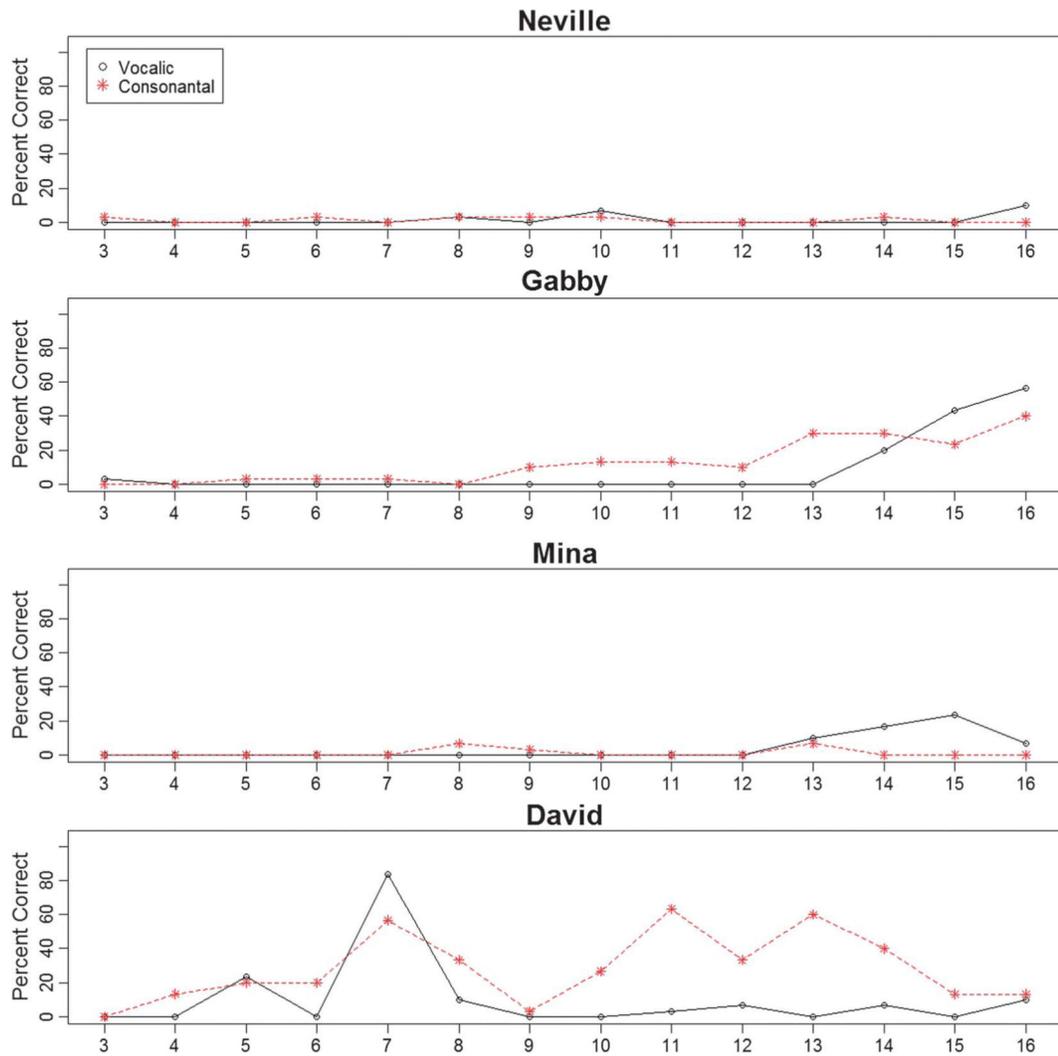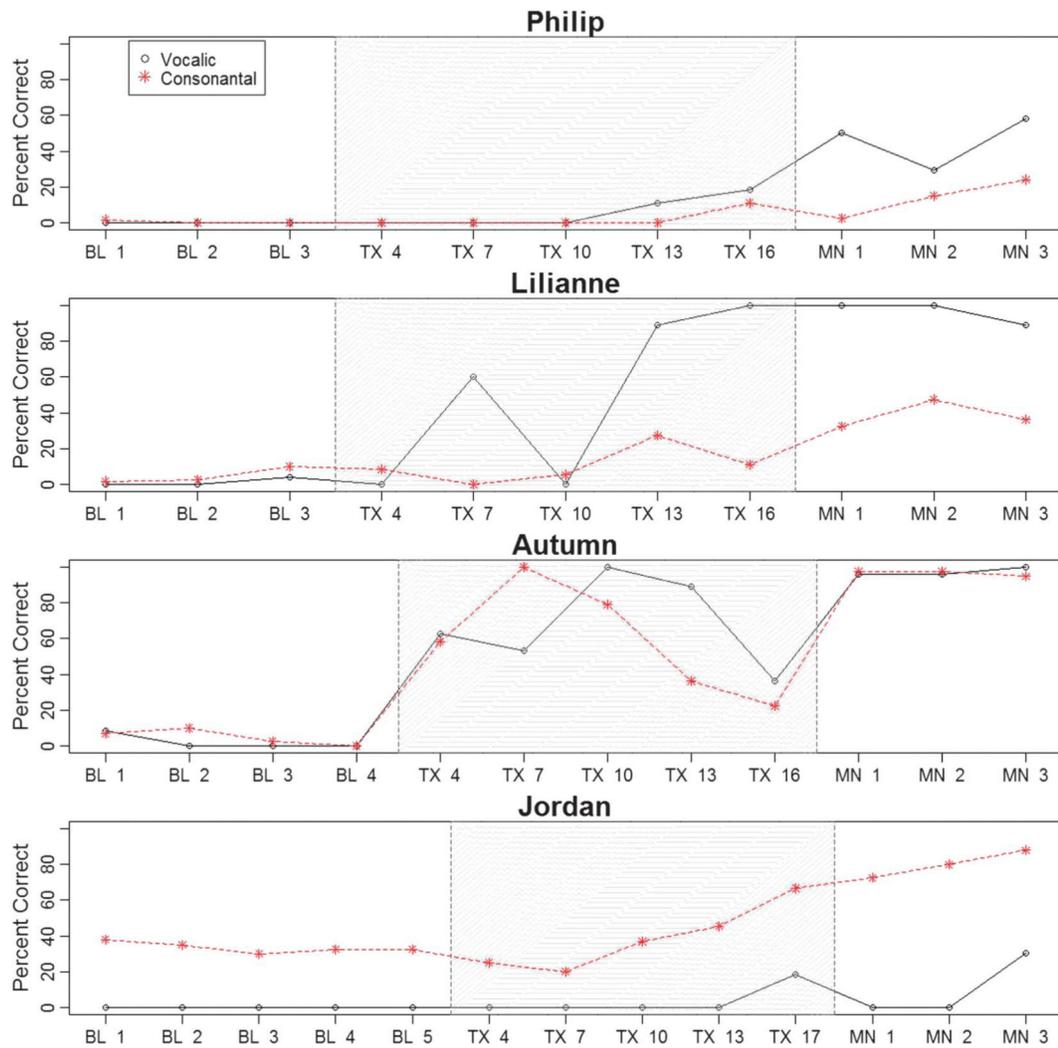
Figure 1.
Ultrasound images captured with the Interson SeeMore probe. The surface of the tongue appears as the white line. Images A and B are in sagittal section; the right side of the image is anterior. Image C is in coronal section. A: Typical adult speaker's bunched tongue shape for /ɝ/. B: Typical adult speaker's tongue shape for /u/, used in treatment to represent an undifferentiated (incorrect) tongue shape for rhotics. C: Typical adult speaker's tongue shape for /ɝ/ in coronal section, revealing midline groove and lateral bracing.
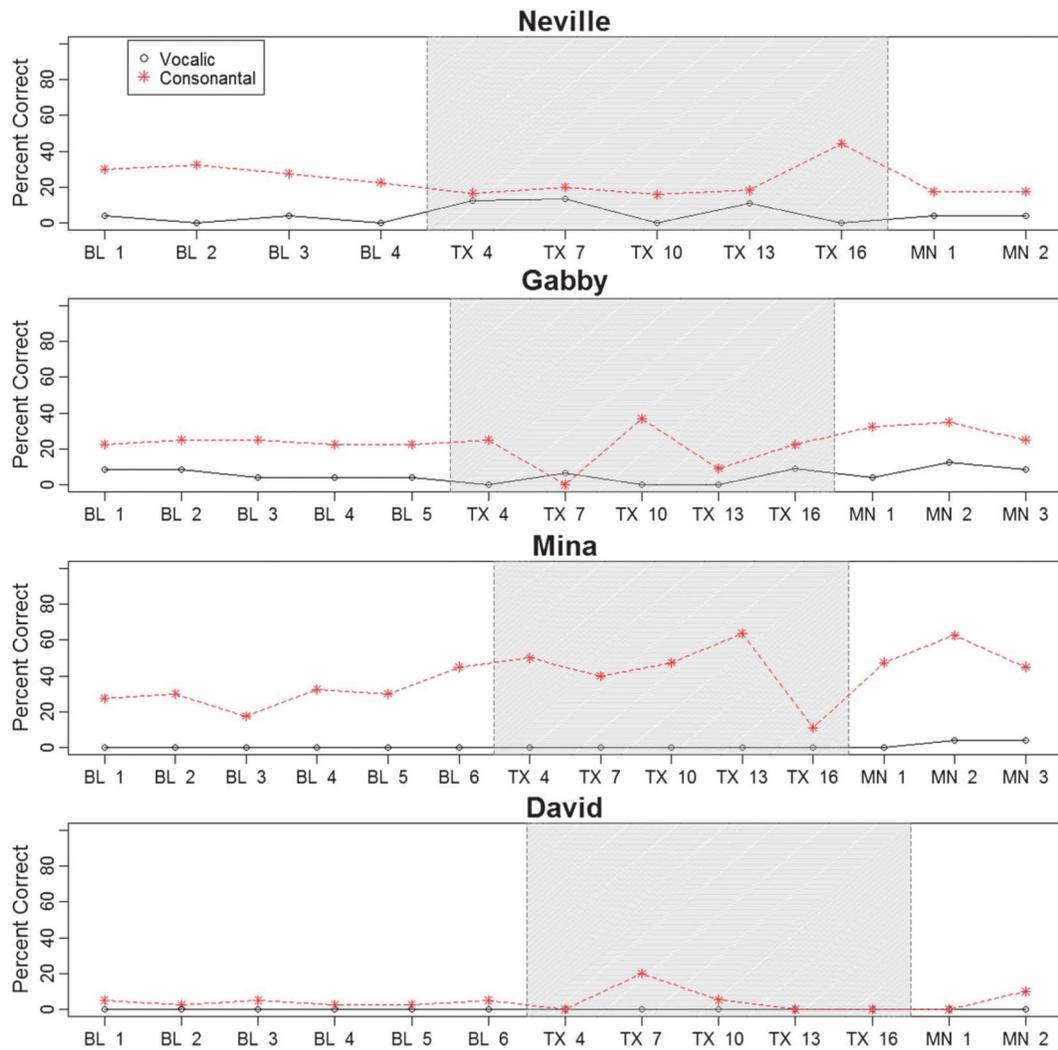
Figure 2.

Word probe performance, Study 1. *Y*-axis represents percentage of tokens rated perceptually correct based on mode across three blinded clinician listeners. BL = baseline; TX = treatment; MN = maintenance.

Figure 3.
Within-treatment performance, Study 1. *Y*-axis represents percent of tokens rated
perceptually correct by a blinded listener. *X*-axis represents treatment session. Session
numbers begin at 3 because no trials were elicited during instructional Sessions 1–2.

Figure 4.
Word probe performance, Study 2. *Y*-axis represents percentage of tokens rated perceptually correct based on mode across three blinded clinician listeners.

Table 1

Participant characteristics and treatment history.

Pseudonym	Age at study onset (years;months)	Previous treatment duration	Previous treatment targets
Neville	7;7	7 months	rhotics
Gabby	10;3	2.5 years	/ʃ/, /tʃ/, /l/, /θ/, /v/, rhotics
Mina	6;1	No previous treatment	na
David	7;8	1.5 years	/s/, rhotics

Note. na = not applicable.

Table 2

Standardized and unstandardized effect sizes for all participants in Study I. Changes in standardized effect size considered clinically meaningful are in boldface type.

Pseudonym	Vocalic variants		Consonantal variants	
	Mean level difference	d_2	Mean level difference	d_2
Neville	2.1	**1.1**	−10.6	**−3.2**
Gabby	2.5	0.8	7.3	**2.3**
Mina	2.8	**2.2**	21.3	**2.4**
David	0.0	0.0[a]	1.3	0.3

[a]Technically, a standardized effect size could not be calculated for this target due to zero variance. However, an effect size of 0.0 clearly captures the lack of progress on this target.

Table 3

Participant characteristics and treatment history: Study 2

Pseudonym	Age at study onset (years;months)	Previous treatment duration	Previous treatment targets
Philip	9;3	4 years	/l/, vowel distortions, rhotics
Lilianne	10;9	2.5 years	/s/, /z/, rhotics
Autumn	7;8	1 year	/θ/, /ð/, rhotics
Jordan	15;8	8 years	rhotics

Table 4

Comparison of protocol across Study 1 and Study 2

Context	Study 1	Study 2
Treatment duration	Two instructional sessions (45 min each); 14 practice sessions (30 min each) over 8 weeks	Three instructional sessions (45 min each); 14 practice sessions (30 min each) over 8.5 weeks
Treatment delivery	Individually delivered by speech-language pathologist (SLP) and student assistant	Individually delivered by SLP and student assistant
	Treatment decisions directed by first author (certified SLP, 5+ years experience)	Treatment decisions directed by second author (certified SLP, 19+ years experience)
Instructional sessions	*Schedule*: Two initial sessions introducing sagittal targets; session introducing coronal target in Week 4	*Schedule*: Three initial sessions introducing sagittal and coronal targets
	Materials: Line drawings of major lingual constrictions; ultrasound images of bunched rhotic	*Materials:* Line drawings of major lingual constrictions; ultrasound images of bunched rhotic; MR images of various rhotic tongue shapes
	Script: Standard script describing bunched tongue shape	*Script:* Customized script describing tongue shapes judged most facilitative for speaker
Treatment sessions	Standard cues during prepractice and treatment trials 30 trials /ɝ/, 30 trials /ɹɑ, ɹi, ɹu/	Individualized cues during prepractice and treatment trials 30 trials /ɝ/, 30 trials /ɹɑ, ɹi, ɹu/

Note. MR = magnetic resonance.

Table 5

Standardized and unstandardized effect sizes for all participants in Study 2. Changes in standardized effect size considered clinically meaningful are in boldface type.

Pseudonym	Vocalic variants		Consonantal variants	
	Mean level difference	d_2	Mean level difference	d_2
Philip	45.8	**4.3**	13.3	**1.7**
Lilianne	94.9	**16.7**	34.0	**4.0**
Autumn	95.1	**12.9**	91.8	**11.8**
Jordan	10.2	1.0	46.6	**6.2**