AJSLP

Research Article

Efficacy of Electropalatography for Treating Misarticulation of /r/

Elaine R. Hitchcock,[a] Tara McAllister Byun,[b] Michelle Swartz,[a] and Roberta Lazarus[a]

Purpose: The purpose of the present study was to document the efficacy of electropalatography (EPG) for the treatment of rhotic errors in school-age children. Despite a growing body of literature using EPG for the treatment of speech sound errors, there is little systematic evidence about the relative efficacy of EPG for rhotic errors.

Method: Participants were 5 English-speaking children aged 6;10 to 9;10, who produced /r/ at the word level with < 30% accuracy but otherwise showed typical speech, language, and hearing abilities. Therapy was delivered in twice-weekly 30-min sessions for 8 weeks.

Results: Four out of 5 participants were successful in achieving perceptually and acoustically accurate /r/ productions during within-treatment trials. Two participants demonstrated generalization of /r/ productions to nontreated targets, per blinded listener ratings.

Conclusions: The present findings support the hypothesis that EPG can improve production accuracy in some children with rhotic errors. However, the utility of EPG is likely to remain variable across individuals. For rhotics, EPG training emphasizes one possible tongue configuration consistent with accurate rhotic production (lateral tongue contact). Although some speakers respond well to this cue, the narrow focus may limit lingual exploration of other acceptable tongue shapes known to facilitate rhotic productions.

It is commonly recognized that a major challenge in the field of speech-language pathology is the treatment of /r/ misarticulation. Rhotic errors occur with high frequency in children who present with speech sound disorder, and they are often resistant to traditional therapy techniques (Ruscello, 1995; Shuster, Ruscello, & Toth, 1995). The limited success rate of existing treatment programs for remediation of /r/ suggests an ongoing need for research exploring alternative methods to facilitate establishment and generalization of rhotic phonemes (Ruscello, 1995). Such research is necessary because unresolved speech errors can have a significant academic, social, and/or emotional impact that may continue through adulthood (Felsenfeld, Broen, & McGue, 1994; Hitchcock, Harel, & McAllister Byun, 2015; McCormack, McLeod, McAllister, & Harrison, 2009). Furthermore, in light of the many clients they serve, speech-language pathologists (SLPs) often struggle to justify keeping children with therapy-resistant errors on the caseload when progress seems to have plateaued. Thus, many individuals are discharged from treatment with persisting errors (Ruscello, 1995), in spite of evidence that spontaneous resolution is unlikely in older children (Gibbon & Paterson, 2006).

Visual Biofeedback for Speech Sound Errors

Enhanced interventions could improve outcomes for children with challenging speech sound errors and could also enable SLPs to move them off the caseload in a timely fashion, freeing up resources for other children with communication needs. A growing body of research suggests that cases of speech sound errors that have not responded to previous intervention can sometimes be eliminated through speech therapy incorporating *visual biofeedback,* in which a real-time visual representation of the user's speech is compared against a model representing correct production of a target sound (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Bacsfalvi, Bernhardt, & Gick, 2007; Gibbon & Paterson, 2006; McAllister Byun & Hitchcock, 2012; McAllister Byun, Hitchcock, & Swartz, 2014; Preston, Brick, & Landi, 2013; Schmidt, 2007; Shuster et al., 1995). Biofeedback uses some form of instrumentation to provide a visual aid for aspects of speech that are difficult to perceive under typical circumstances (Volin, 1998). Drawing on this external image, the learner can attempt to change a speech sound error by matching a visually accurate target

---

[a]Montclair State University, Bloomfield, NJ
[b]New York University

Correspondence to Elaine R. Hitchcock: hitchcocke@mail.montclair.edu

Editor: Krista Wilkinson
Associate Editor: Lynn Williams

Received June 30, 2016
Revision received December 16, 2016
Accepted May 10, 2017
https://doi.org/10.1044/2017_AJSLP-16-0122

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

instead of relying on internal self-perception. Thus, biofeedback can be thought of as a means to allow the learner to adopt an external focus of attention for speech movements (Maas et al., 2008).

Various technologies can be harnessed to provide biofeedback for speech. These include *visual–acoustic biofeedback,* in which the client views a computer-generated acoustic representation (e.g., real-time spectrogram) of his or her speech; *ultrasound biofeedback,* in which an ultrasound probe held beneath the chin generates an image of the client's tongue during speech; and *electropalatographic* (EPG) *biofeedback,* which uses a pseudopalate to register and display areas of contact between the client's tongue and palate. For each of these methods, small-scale studies have supported the hypothesis that biofeedback can facilitate acquisition of speech targets that have not responded to previous forms of intervention, (e.g., visual-acoustic: McAllister Byun & Hitchcock, 2012; Shuster, Ruscello, & Smith, 1992; Shuster et al., 1995; ultrasound: Adler-Bock et al., 2007; McAllister Byun et al., 2014; Preston et al., 2013; EPG: Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Dagenais, Critz-Crosby, & Adams, 1994; Fabus et al., 2015; Gibbon & Hardcastle, 1987; McAuliffe & Cornwell, 2008; Schmidt, 2007). The present study focuses on EPG, an application which is currently being marketed as a biofeedback tool for the treatment of rhotic errors in children (CompleteSpeech, n.d.).

EPG as a Biofeedback Tool

The ability to visualize lingua-palatal contacts has the potential to facilitate the remediation of therapy challenging speech sound errors. Individuals using EPG wear a pseudopalate within the mouth during treatment. The pseudopalate, which fits much like an orthodontic retainer, is embedded with electrodes designed to record the location and timing of tongue contacts with the hard palate (Cheng, Murdoch, Goozee, & Scott, 2007). The data are sent from the electrodes via a microprocessor that provides an interface between the pseudopalate and the computer. A software program displays the tongue-to-palate contacts on the computer screen. In some programs, a split-screen displays the output of pseudopalates worn by the clinician and the participant; the acoustic waveform of the individual's speech may also be displayed on the computer screen.

In this study, EPG intervention was provided using the CompleteSpeech Palatometer V1.0 system (SmartPalate International, Version V1.0). This system uses a custom-made pseudopalate (SmartPalate) and a microprocessor input/output device (DataLink) that is worn around the user's neck and interfaces between the SmartPalate and the computer, in this case an IBM Lenovo desktop (Model IZVPro). Figure 1 provides a grayscale example of an image captured with the study equipment.

Over two decades of research have documented the use of EPG for evaluation and treatment of speech error patterns resulting from a variety of diagnoses, including articulation and phonological disorders, cleft palate, apraxia, dysarthria, hearing impairment, Down syndrome, and cerebral palsy, among others (Bernhardt, Bacsfalvi, Gick, Radanov, & Williams, 2005; Carter & Edwards, 2004; Cleland, Timmins, Wood, Hardcastle, & Wishart, 2009; Dagenais et al., 1994; Gibbon & Paterson, 2006; Gibbon & Wood, 2003; Hardcastle, Gibbon, & Jones, 1991; Howard & Varley, 1995; Martin, Hirson, Herman, Thomas, & Pring, 2007; Schmidt, 2007). More specifically, past researchers have studied the tongue-to-palate contact patterns used by children with speech errors to better understand the nature of these errors and more effectively cue correct production. In addition, such information aids in developing targeted remediation programs using EPG as a visual biofeedback tool. Although these studies are small in scale, they indicate that EPG may be an effective enhancement to intervention in some individuals with speech sound errors.

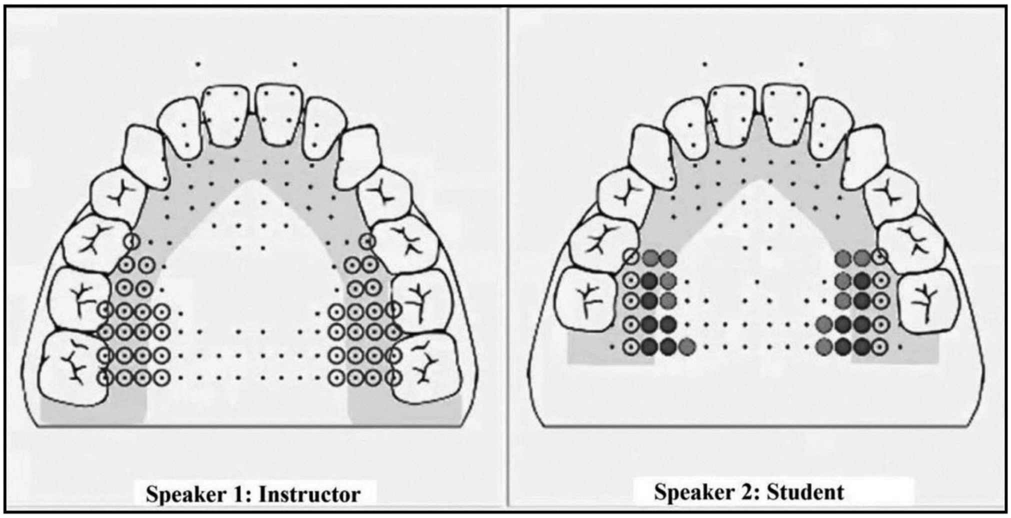
One noteworthy result reported by Dagenais et al. (1994) and McAuliffe and Cornwell (2008) was a finding that participants who made perceptual gains still showed limited changes in tongue-to-palate contacts. These findings suggest that some participants are able to modify an incorrect contact pattern enough to produce a perceptually and acoustically acceptable production without fully correcting the tongue-to-palate error pattern. This finding warrants further exploration in future studies.

EPG and Rhotic Errors

Previous literature on speech sound errors has given particular attention to misarticulation of the phoneme /r/, a late-emerging sound that is described as one of the most likely to fail to respond to conventional therapies (Ruscello, 1995). To adequately address the literature regarding EPG and rhotic errors, we must first briefly review the unique characteristics of English rhotic production. Speech acquisition can be described as an individualized process in which speakers learn what configurations of their own vocal tract will map onto targets defined in auditory–acoustic space (e.g., Guenther, Hampson, & Johnson, 1998). For resonant sounds such as vowels and liquids, these targets can be thought of as the locations of formants or resonant frequencies of the vocal tract. The auditory target for /r/, a lowered third formant (F3), is known to require a particularly complex articulatory configuration (Boyce, 2015). Most articulatory descriptions identify virtually simultaneous anterior/palatal and posterior/pharyngeal lingual constrictions (Boyce, 2015; Klein, McAllister Byun, Davidson, & Grigos, 2013). The appropriate degree of lip constriction is also part of the articulatory configuration for most speakers (Bernhardt & Stemberger, 1998).

Articulatory imaging has shown that speakers produce English rhotics using a variety of different tongue shapes (Delattre & Freeman, 1968; Tiede, Boyce, Holland, & Choe, 2004; Zhou et al., 2008) and many speakers use different tongue shapes across different phonetic contexts (Mielke, Baker, & Archangeli, 2010; Stavness, Gick, Derrick, & Fels, 2012). The two most contrastive tongue shapes are

Figure 1. Electropalatographic image captured with the CompleteSpeech SmartPalate system. Speaker 1 represents the instructor palate image with preset rhotic targets (open circles) and no speaker input. Speaker 2 represents the student image with speaker input. Dark gray circles indicate a match between the target and the speaker's tongue-to-palate contact; light gray circles indicate the speaker's tongue-to-palate contact not in the target range; and open circles indicate unmatched targets.



commonly described as *retroflex* and *bunched,* although many intermediate variants exist. In the retroflex variant, the tongue tip raises and may approximate contact with the alveolar ridge while the tongue dorsum is lowered. In the bunched variant, the tongue tip lowers while the tongue dorsum raises to approximate the hard palate. The tongue shapes for retroflex and bunched /r/ are described as similar to the tongue configurations observed for /s/ and /ʃ/, respectively (Zhou et al., 2008). A grooved shape reflecting lowering of the middle of the tongue relative to raised lateral margins is observed in many speakers (Bacsfalvi, 2010). Zhou et al. (2008) reported that the tongue shape variants appear to show no differences perceptually or acoustically at the level of the first three formants; however, they may be differentiated by the fourth and fifth formants. In keeping with the notion of speech as a mapping from articulator configurations to auditory–acoustic outcomes, clients should be offered the opportunity to explore different tongue shapes to find the configuration that facilitates the most perceptually accurate rhotic sounds in their individual vocal tract (McAllister Byun et al., 2014).

The present study will adopt a division between *vocalic* /r/ (syllabic /r/ and rhotic diphthongs) and *consonantal* /r/ (onset position only) that is clinically widely used (but not universally accepted; see discussion in Lockenvitz, Kuecker, & Ball, 2015). Syllabic /r/ may act as the nucleus of a syllable as in /hɝ/-*her* or /fivɚ/-*fever*. When /r/ occurs in the postvocalic position such as /hɛɚ/-*hair* or /nɪɚ/-*near*, we will treat this variant as an off-glide of the rhotic diphthong, based on evidence that it is more closely identified with a syllabic /r/ from both acoustic and articulatory perspectives (McGowan, Nittrouer, & Manning, 2004). Last, /r/ can appear as a consonant in the syllable onset position in words such as *rip* and *grass*. We will adhere to the clinically common convention of using /r/ to transcribe the English consonantal rhotic.

EPG Biofeedback Treatment: Previous Results

Schmidt (2007) reported on EPG intervention for 13 children aged 7–12 years. All participants had previously received a minimum of 3 years of traditional articulation treatment. Nine participants demonstrated rhotic errors: four as their only error and five in conjunction with other speech sound errors. The group was heterogeneous with respect to the factors causing speech impairment, including hearing impairment, apraxia of speech, hypotonia, and cleft palate. Participants received a course of EPG intervention in two 30-min sessions per week (the total duration of treatment varied per participant, ranging from 12 to 30 sessions). Posttreatment, eight out of nine participants had successfully achieved stable, perceptually acceptable productions of their rhotic target sounds that had also generalized beyond treated words. Similar to Dagenais et al. (1994) and McAuliffe and Cornwell (2008), not all participants who exhibited improved perceptual/acoustic accuracy also showed significantly altered tongue-to-palate contacts, independent of the perceptual and acoustic outcomes reported.

In a recent case series, Fabus et al. (2015) documented treatment effects of EPG for /r/ errors for three participants aged 9–11 years, who were seen for one 45-min session per week for a duration of 10 weeks. Two children presented with only rhotic errors, whereas one child demonstrated rhotic and fricative errors. Pretreatment assessment revealed typical receptive and expressive language skills and hearing status within normal limits. Significant improvement in the production of rhotic phonemes in the treatment setting was observed for two of the three participants for whom /r/ was addressed, although gains were variable both within and across children. These results suggest that the addition of EPG as a biofeedback tool for the treatment of rhotic misarticulation can lead to positive outcomes.

Although previous results have been promising, the existing literature does not provide examples of controlled experimental studies measuring the effect of EPG biofeedback in a reasonably homogeneous group of individuals with rhotic misarticulation. At first glance, it seems logical to question how much added advantage can be expected from EPG as a supplement to treatment for rhotic errors. Sounds with easily identified tongue-to-palate contacts, such as /t/ or /k/, seem more appropriate for EPG remediation than rhotics, which have less clearly defined linguapalatal contacts and are subject to much greater variability across individuals. However, in light of the fact that rhotic misarticulation is one of the targets for which EPG is currently being marketed, it is important to conduct more research objectively assessing the efficacy of this application. The aim of the present study, using a single-subject, multiple-baseline research design and multiple methods of analysis, was to systematically measure the efficacy of EPG as a visual biofeedback tool for the treatment of rhotic productions in children with speech sound disorder.

Method

Participants

Participants were five children ranging in age from 6;10 to 9;10 (mean age = 7;8).[1] Demographically, four participants were White, and one was of Hispanic origin. All were monolingual native speakers of American English. Participant recruitment and retention are diagrammed in Figure 2. Four out of five participants had previously received 6 months to 2 years of traditional articulation therapy for /r/ errors. One male participant had not been a recipient of therapy prior to this study because his /r/ error was not judged to impact his academic work, and he was thus deemed ineligible for treatment in school.[2] Unlike the other participants, who had all received some duration of previous treatment without success, we do not know how this participant would have responded to a traditional articulation program. Only one participant had previously received therapy for other speech sounds (/s/, /z/, and /l/), all of which had resolved prior to the onset of the present treatment program. Participants' therapy histories are reported in Table 1.

All participants scored within the average range on a measure of receptive language (Test of Auditory Processing Skills–Third Edition, Auditory Comprehension subtest; Martin & Brownell, 2005). Participants also passed a pure-tone hearing test (1000, 2000, and 4000 Hz at 20 dB HL) and exhibited no gross structural or functional abnormality in an evaluation of the oral mechanism. To confirm that speech production skills were largely intact, all participants were assessed using the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000). For inclusion in the study, participants had to exhibit no more than three speech sounds in error, including /r/, in the Single Word and Storytelling subtests. Two final measures evaluated potential participants' ability to produce the /r/ sound. Stimulability was assessed by eliciting imitation of /r/ in isolation and in syllable-initial, intervocalic, and syllable-final positions in the vowel contexts /i, ɑ, u/ (Miccio, 2002). Participants who demonstrated ≥ 30% accuracy were not included in the study because children who are stimulable for a sound may be in the process of acquiring that sound in spontaneous production (Powell, 1993). Less than 30% accuracy was also required on a single-word /r/ probe task that was administered both as a criterion for inclusion in the study and as part of the baseline reflecting performance prior to the initiation of treatment. In the single-word probe measure, pictures and written words were used to elicit 64 familiar words containing /r/. These 64 items were selected to represent a full range of syllable positions and phonetic contexts because /r/ may be realized with differing accuracy in different environments (Elbert & McReynolds, 1975). Consonantal /r/ was probed in both singleton and cluster contexts. Because front vowels have, in some cases, been found to facilitate /r/ articulation (Kent, 1982), equal numbers of front and back vowel contexts were used when eliciting consonantal /r/. Vocalic /r/ was probed in the following forms: (a) stressed /ɝ/, (b) unstressed /ɚ/, (c) /ɑr/, (d) /ɛr/, (e) /ɔr/, and (f) /ɪr/. No feedback was provided during /r/ probe administration. Results of stimulability and probe findings are reported in Table 1.

Study Design

All study procedures were approved by the Institutional Review Board at Montclair State University. Prior to the initiation of treatment, participants met with an orthodontist who created a mold of their upper dentition. The dental mold was sent to the palate manufacturer, CompleteSpeech, who used the mold to create an individualized practice palate used for oral desensitization and a treatment palate embedded with electrodes.

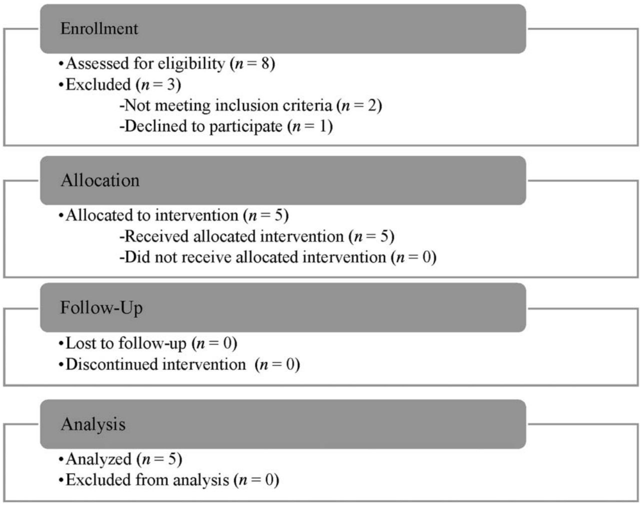
The study followed a multiple-baseline across-subjects design. Participants were randomly assigned to receive three, four, or five baseline sessions. Two probe measures, including the above-described 64-word /r/ probe and a probe eliciting /r/ at the sentence level (Schmidlin & Boyce, 2010), were

---

[1] Overall, the participants in the present study were slightly younger than what is considered typical for children diagnosed with residual speech errors. Therefore, further research is needed to evaluate how older children with residual /r/ errors would respond to the treatment investigated here.

[2] It is not an uncommon practice in U.S. schools to make decisions about treatment allocation on the basis of whether the child's speech errors affect his or her academic performance. For a more detailed discussion of this issue, see Hitchcock et al. (2015).

Figure 2. Participant recruitment and retention.



Enrollment
- Assessed for eligibility ($n = 8$)
- Excluded ($n = 3$)
  - -Not meeting inclusion criteria ($n = 2$)
  - -Declined to participate ($n = 1$)

Allocation
- Allocated to intervention ($n = 5$)
  - -Received allocated intervention ($n = 5$)
  - -Did not receive allocated intervention ($n = 0$)

Follow-Up
- Lost to follow-up ($n = 0$)
- Discontinued intervention ($n = 0$)

Analysis
- Analyzed ($n = 5$)
- Excluded from analysis ($n = 0$)

elicited in each baseline session. The sentence probe consisted of five sentences containing both vocalic and consonantal /r/ in various phonetic contexts, with multiple /r/ words per sentence. A criterion of baseline stability, defined as < 10% mean session-to-session variability over the baseline interval, was adopted to establish the absence of maturational gains prior to the initiation of treatment (Kratochwill & Levin, 2014). After the baseline period, participants completed sixteen 30- to 45-min individual treatment sessions over 8–10 weeks, including two introductory sessions and 16 biofeedback practice sessions. A randomly selected 20-item subset of the rhotic word probe was administered at the start of every third session during treatment. Words elicited in this measure were never targeted in the context of intervention; thus, these probes were used as a measure of generalization to untreated words over the duration of the study. To evaluate maintenance of any gains made in therapy, the full 64-word and sentence-level probes were collected in three sessions over a period of 2 weeks after the end of the treatment phase. Last, a 1-month follow-up maintenance probe was collected from the three participants. Due to scheduling conflicts, two participants did not complete the follow-up probe.

All baseline, treatment, and maintenance sessions were recorded in a sound-shielded room using the Computerized Speech Lab system (Model 4500, Pentax Medical) with a 44.1-kHz sampling rate. Participants spoke into a Shure condenser microphone with a mouth-to-microphone distance of approximately 5 in. for a favorable signal-to-noise ratio.

All study sessions were implemented following a standardized protocol. Treatment was carried out by either the first author or the fourth author, both certified SLPs. A trained graduate research assistant was present to provide support with equipment management, fidelity to the stated treatment protocol, and data collection.

Introductory Sessions

The study began with two initial sessions designed to familiarize the participant with the palate and instruct him/her in the skill of interpreting the visual display of tongue-to-palate contacts. The treating clinician followed standard scripts to maintain consistency across introductory sessions. The first introductory session consisted of a fitting of the practice palate and pseudopalate with electrodes and an orientation to the palate image using a split-screen display of the clinician's and participant's computerized palate representations. This was followed by a 10-min exploration period using the pseudopalate to orient the participant to the equipment and the associated visual image. The practice palate was sent home with instructions to wear the palate for 5–10 min for 7 days to desensitize the participant to wearing the pseudopalate. Per parent report, all participants followed through with the instructions with the exception of Ethan, whose practice palate was misplaced after 2 days in the home setting.

The second introductory session included an explanation of the tongue-to-palate contacts for two sounds that participants could readily produce, namely, lingua-alveolar and velar phonemes. The child was then cued to produce these phonemes and observe the associated contact patterns. Finally, a preset tongue-to-palate contact pattern consistent with a potential rhotic configuration was added to the computer palate image while the child attempted /r/ productions. The target encouraged bracing of the posterior tongue against the lateral margins of the palate. Verbal articulator placement cues were provided to encourage participants to match their tongue-to-palate contacts to the target region while producing their most perceptually accurate rhotic sound. If a participant made an articulatory modification that did not match the tongue contact target

Table 1. Participant history.

Pseudonym	Age	Therapy history	Previous TX targets	Stimulability	/r/ Word probe
Desiree	9;10	1 year	/r/	16% (5/30)	5% (3/64)
Brianna	8;2	1 year	/r/	13% (4/30)	0% (0/64)
Ethan (1st grade)	7;1	2 years	Multiple targets + /r/	0% (0/30)	0% (0/64)
Derek (2nd grade)	7;0	No previous TX[a]		0% (0/30)	0% (0/64)
Jenna (1st grade)	6;10	6 months	/r/	0% (0/30)	2% (1/64)

Note. TX = treatment.

[a]Derek's /r/ error was reportedly not impacting his academic work; thus, he was considered not eligible for treatment in school.

but did result in a perceptually improved rhotic, perceptual accuracy was favored (McAllister Byun et al., 2014); participants were thus cued to continue to match the tongue-to-palate contact pattern associated with their best perceptual production. Last, cues for elements of articulator placement that are not visible via EPG (e.g., tongue root retraction, jaw aperture, and degree of lip constriction) were provided to help the participant achieve accurate production of rhotic targets.

Treatment Trials

Treatment trials commenced after the first two introductory sessions. All treatment sessions began with a 5- to 10-min period of "free play" designed to review the cues identified as most facilitative in the introductory session and/or previous treatment sessions. The pseudopalate was generally used during the free-play period, although it was not reintroduced for participants who had advanced to a high level in the adaptive complexity hierarchy described below. Next, the participant produced 60 trials targeting both vocalic and consonantal /r/ productions (30 /ɝ/, 10 trials each of the syllables /rɑ/, /ri/, and /ru/). Stimuli were elicited in constant order in blocks of five trials. A verbal reminder of one articulator placement cue was provided directly prior to each block of trials. Participants attempted to match the clinician's live model and/or a preset model reflecting lateral tongue bracing. After each block, the clinician provided the participant with 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"). Treatment targets were initiated at the syllable level. Participants who reached 80% correct within a session were advanced through a hierarchy of phonological complexity (e.g., consonant–vowel nonwords, consonant–vowel real words, consonant–vowel–consonant nonwords, and consonant–vowel–consonant real words). Biofeedback was gradually faded from 100% to 50% to 0% for those participants who reached 80% or greater accuracy within a session. In the 50% biofeedback condition, EPG biofeedback was withdrawn in blocks (five trials on, five trials off) to minimize the challenges of inserting and removing the palate, which was sometimes held in place with dental adhesive. All /r/ trials during therapy sessions were scored online by the treating clinician.

Measurement

Progress was tracked using the /r/ word probes administered before, during, and after biofeedback therapy. Three certified SLPs were trained to use standard criteria to score children's /r/ sounds in each word as on-target ("1") or off-target ("0"). They were instructed that distorted sounds with some rhotic quality should be rated "0." Each item was rated by all three listeners in a randomized, de-identified fashion with computerized response recording. Before rating experimental stimuli, raters completed a sample set of 86 items that had been rated by an experienced clinician in a previous study. Only clinicians who demonstrated ≥ 80% rater 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. Praat (Version 5) software (Boersma & Weenink, 2010) was used for randomized, de-identified stimulus presentation and response recording. The full set of items (n = 2,932) was subdivided into blocks of approximately 200 items. Raters completed all blocks in a self-paced fashion over the span of 2 weeks. The raters were allowed to replay each token a maximum of 3 times. 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. Pairwise interrater reliability was 91% for Raters 1 and 2, 89% for Raters 1 and 3, and 84% for 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. Due to the lack of blinding in this context, clinician ratings are potentially influenced by bias, such as an expectation of improvement over the course of treatment. We can attempt to estimate the magnitude of this bias by comparing the clinician's unblinded ratings against the mode across blinded listeners for a set of tokens that were rated in both ways. The full set of probe items (baseline, maintenance, and within treatment) was used for this purpose. Interrater reliability was 93%, indicating a high level of agreement between the online judgments of the treating clinician and the mode across three blinded expert listeners. Blinded expert ratings were not available for within-treatment productions due to the large number of trials (N = 4,800) elicited in this context. However, Harel, Hitchcock, Szeredi, Ortiz, and McAllister Byun (2016) found that reliability measures estimated from a small subset of data were highly correlated with the same measure as derived from a larger subset of the data from the same individuals. Such findings suggest that one set of data can be a useful predictor of future ratings and support the possibility that the high level of agreement observed to hold between the clinician and the blinded raters on baseline and maintenance probes might also hold for within-session data. However, no direct comparisons will be drawn between these two sets of ratings because the trials produced within treatment and rhotic word probes were rated by different listeners and occurred under different production conditions (i.e., within-treatment trials could include palate use).

Analyses

Consistent with our past work (Hitchcock et al., 2015; McAllister Byun & Hitchcock, 2012; McAllister Byun et al., 2014), the percentage of rhotic words rated correct was plotted across baseline, within-treatment, and maintenance probes for each participant in order to visually examine the effects of treatment over time. Standardized effect sizes (ESs) were computed using d_2, which is Busk and Serlin's (1992) modification of Cohen's *d* statistic (Beeson & Robey, 2006). Values for d_2 were calculated by pooling the standard deviations across baseline and maintenance intervals to limit the number of cases where ES cannot be calculated due to zero variance in the baseline period. In the present study, a treatment effect was considered clinically meaningful if d_2 exceeded 1.0, indicating that the difference between pre- and posttreatment means exceeded the pooled standard deviation (Maas & Farinella, 2012). Because d_2 can overestimate the magnitude of an effect in instances where variance is extremely low, unstandardized ESs were also calculated. The unstandardized ES, or mean level difference, is defined as the raw difference between the mean percentage of items rated correct in maintenance versus baseline intervals.

Acoustic analysis, specifically measurement of the F3–F2 distance, was also performed as a supplemental assessment of /r/ production accuracy for all word probes. Acoustic measurements were conducted by four graduate students who were trained to use Praat acoustic software (Boersma & Weenink, 2010) in a one-to-one instructional module. The graduate students were taught to manipulate settings within Praat's automated linear predictive coding (LPC) formant tracking function until the automatically calculated formants were clearly paired with the visible areas of energy concentration in the spectrogram. After identifying a region representing the rhotic portion of the utterance, the students were shown how to select a point representing the minimum height of F3 while avoiding values that appeared as outliers relative to adjacent points in the automatic LPC formant track. Next, the first three formants (values in hertz) were calculated with a 14-ms Hamming window around the selected point using Burg's method of computing LPC coefficients. The present study only reports measurements calculated from the F3 and F2 data.

Research has shown that a low height for F3 is representative of a rhotic production (Delattre & Freeman, 1968; Hagiwara, 1995). Moreover, the second formant is often found to be high, which suggests that an accurate rhotic production is characterized by a relatively small distance between F3 and F2 (Boyce & Espy-Wilson, 1997; Dalston, 1975). In keeping with past work, the present study will use the F3–F2 distance as the primary acoustic measure of degree of rhoticity (Flipsen, Shriberg, Weismer, Karlsson, & McSweeny, 2001; Shriberg, Flipsen, Karlsson, & McSweeny, 2001).

Fidelity

Twenty percent of all sessions were reviewed to evaluate fidelity to the stated treatment protocol (Kaderavek & Justice, 2010). Each rater 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 feedback was provided after each block. Results of the fidelity check will be reported and discussed in the Results section.

Results

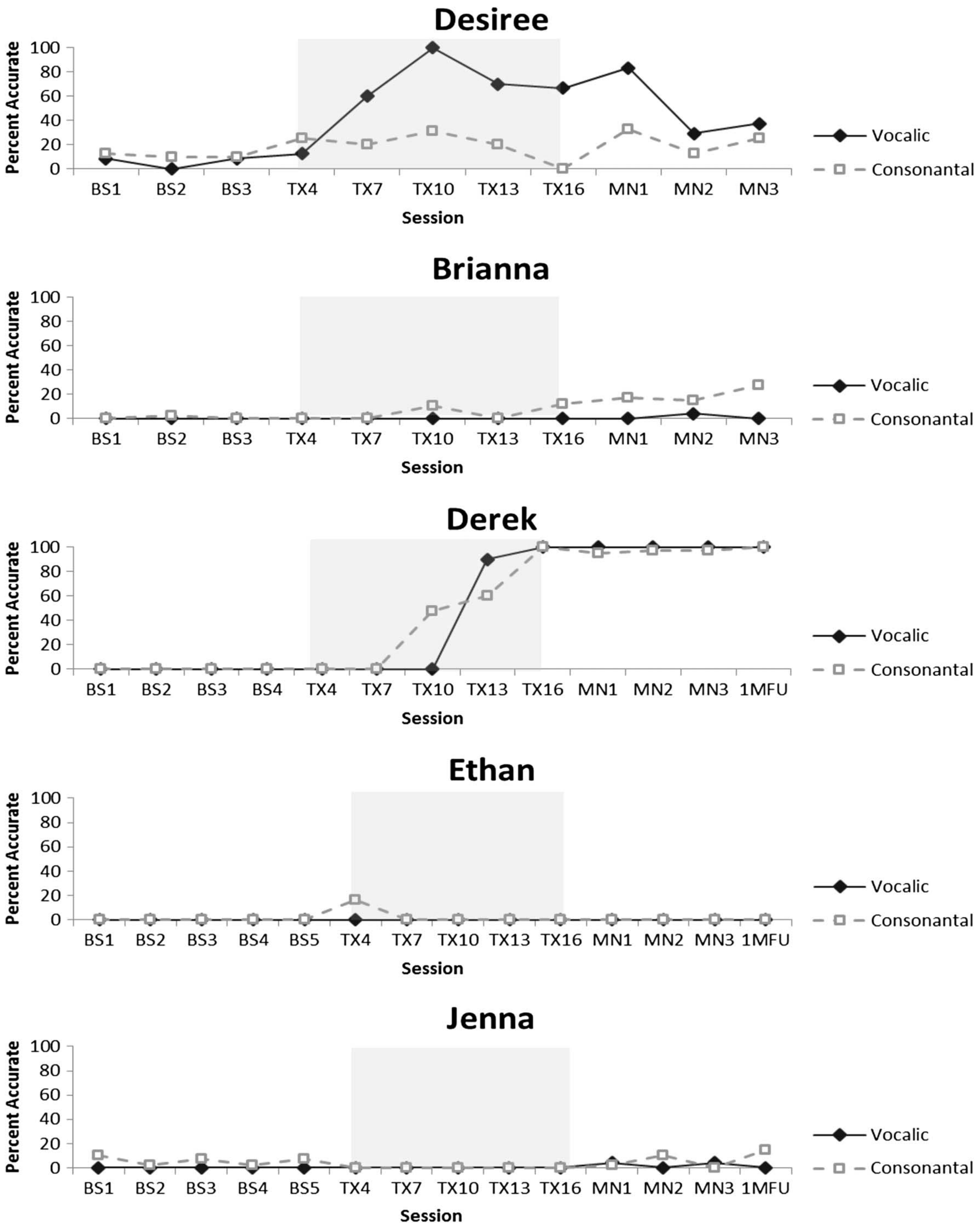
Word Probes: Perceptual Ratings

The multiple-baseline graphs in Figure 3 depict baseline, treatment, and maintenance intervals, with the treated interval 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. Consonantal and vocalic variants are plotted separately, with consonantal variants represented with squares and a dotted line and vocalic variants represented with diamonds and a solid line. For baseline and maintenance probes, the percentage scores plotted were calculated over 32 items each for vocalic and consonantal trials; for within-treatment probes, percentages were calculated over 10 vocalic and 10 consonantal items.

The multiple-baseline graphs in Figure 3 visually reflect the trajectory of change over the course of the study for each participant. These findings can be corroborated with the magnitude of the standard ES and mean level differences found in Table 2. All participants maintained an adequately stable baseline for both vocalic and consonantal variants. One participant, Derek, who had previously not received formal speech therapy, showed a strong response to treatment, as evidenced by a significant increase in perceptually rated accuracy across both vocalic and consonantal contexts. The majority of this increase occurred abruptly, with accuracy increasing from 0% to ≥ 90% over the course of a 3-week window beginning at Session 8 for consonantal variants and Session 11 for vocalic variants. In the three maintenance probes and the follow-up probe, Derek's accuracy for vocalic and consonantal /r/ variants at the word level remained effectively at ceiling (mean = 100% and 97.5% correct, respectively). Although the standardized ES could not be calculated due to zero variance in baseline and maintenance sessions, the mean level difference of 97.5% clearly represents a very large ES, supporting the substantial perceptual gains evident from visual inspection.

Visual inspection for the second participant, Desiree, showed a mixed response to treatment, with moderate improvement on vocalic targets and somewhat smaller gains for consonantal targets. In the case of vocalic targets, visual evidence of improvement was supported by an unstandardized mean level difference of 45.8 percentage points and a standardized ES of 2.1. Considerably higher levels of accuracy were observed for vocalic variants on generalization probes elicited during the treatment period, but accuracy fell off after the first maintenance session, suggesting that Desiree had some difficulty retaining the gains that she made in treatment. Consonantal variants were associated with a

Figure 3. Word probe performance. The *y*-axis represents the percentage of tokens rated perceptually correct based on the mode across three blinded clinician listeners. BS = baseline; TX = treatment; MN = maintenance; 1MFU = 1-month follow-up.



modest mean level difference of 8.33; the standardized ES came out to 1.7, which does pass the threshold to be viewed as clinically meaningful (Maas & Farinella, 2012). In contrast with Desiree, Brianna showed stronger progress on consonantal than vocalic variants of /r/. For consonantal targets, visual evidence of minimal improvement was noted by an unstandardized mean level difference of 32.47 percentage points and an ES of 3.7, which was likely inflated by low variance across probe measures. She showed minimal improvement for vocalic targets, with a mean level

Table 2. ES for word probes per participant.

	ES category			
	Vocalic		Consonantal	
Participant	**Mean level difference**	**d_2**	**Mean level difference**	**d_2**
Desiree	45.8	2.1	8.33	1.7
Brianna	1.2	.83	32.47	3.7
Ethan	0	0.0	0	0.0
Derek	100	0.0[a]	97.5	105.9
Jenna	1.7	2.0	–0.5	–0.45

Note. ES = effect size.

[a] Technically, a standardized ES could not be calculated for the targets due to zero variance. However, an ES of 0.0 clearly captures the lack of progress or complete accuracy of the target.

difference of 1.2 and an ES of .83, which does not meet the threshold to be considered clinically meaningful.

The two remaining participants, Ethan and Jenna, showed minimal change in performance on rhotic word probes for either consonantal or vocalic targets. Jenna's productions involving vocalic rhotics show an unstandardized mean difference of only 1.7 percentage points, indicating that the standardized ES of 2.0 percentage points has been inflated by low variance in the baseline and maintenance phases. Thus, we will not interpret this change as clinically meaningful. Consonantal differences were also minimal, as evidenced by a mean level difference of −0.5 and an ES of −0.45. Ethan's scores were unchanged at 0% correct throughout the baseline and maintenance intervals. Therefore, a standardized ES cannot technically be calculated; however, it is clear that his lack of progress can be reasonably represented with an ES of 0.0.

Word Probes: Acoustic Measurements

In order to corroborate the perceptual ratings of the expert clinical raters, acoustic ratings were collected in the form of F3–F2 distance values; recall that a lower F3–F2 distance corresponds with a more accurate /r/ sound. Mean F3–F2 distances and standard deviations are reported in Table 3. A paired-samples two-tailed *t* test was used to compare pre- and posttreatment F3–F2 distances for consonantal and vocalic /r/ for each participant. These findings (Table 4) revealed a statistically significant difference between F3–F2 distances for pre- versus posttreatment consonantal productions for three of the five participants (Desiree, Derek, and Jenna) and three of the five participants for vocalic productions (Desiree, Derek, and Brianna).

For Derek, statistically significant acoustic changes for both consonantal and vocalic rhotic productions were observed, consistent with the perceptual judgments reported previously. Desiree also showed significant acoustic changes for both consonantal and vocalic variants, although the magnitude of change was smaller than for Derek, in keeping with the smaller ES calculated on the basis of perceptual ratings of her speech. Brianna showed a small but significant change in the mean F3–F2 distance for vocalic but not consonantal variants, but in this case, blinded listeners' ratings did not indicate a clinically meaningful change in vocalic accuracy. Meanwhile, Jenna showed a small but significant decrease in F3–F2 distances that affected only consonantal targets, which was also not identified by blinded listeners' ratings of a clinically meaningful difference for the perceptual accuracy of either vocalic or consonantal variants. Finally, no statistically significant differences in F3–F2 distances were observed for Ethan, which aligns with the raters' perceptual judgments of his probe productions.

Both the perceptual ratings and acoustic measures of word probes reported above suggest large gains for Derek, moderate gains for Desiree, questionable mild gains for Brianna, and no meaningful gains for Ethan and Jenna. Participants differed in whether gains affected vocalic /r/, consonantal /r/, or both. Taken in total, the word-level probes suggest that participants receiving EPG biofeedback treatment for misarticulation of /r/ take away differing levels of generalization gains.

Within-Treatment Trials: Perceptual Measurements

Additional analyses of within-treatment data were conducted to explore the possibility that gains were in progress but had not yet generalized beyond the treatment setting. Recall that within-treatment scores were generated by the treating clinician in an unblinded fashion. Moreover, different participants could practice at different levels of complexity, depending on what level they had reached in the treatment hierarchy. For these reasons, results in this section must be interpreted with more caution than the probe results reported previously. On the other hand, recall that the treating clinician's unblinded ratings of probe measures showed strong agreement with ratings assigned by blinded experts, suggesting that the influence of bias on her ratings was relatively minor.

Visual inspection of the within-treatment trajectories of progress shown in Figure 4 reveals extensive within- and across-participant variability, including gains among the participants who did not show progress on the word probe measures. As in the generalization probes, Derek showed

Table 3. Mean and standard deviation of F3–F2 difference measured at pre- and posttest.

Participant	Variant	Time of probe	*n*	Average F3–F2 difference	Average session *SD*
Desiree	Consonantal	PreTX	120	1,634.89	329.90
	Consonantal	PostTX	120	1,334.34	331.83
	Vocalic	PreTX	72	1,450.66	175.74
	Vocalic	PostTX	72	1,092.63	278.65
Brianna	Consonantal	PreTX	120	1,350.76	275.04
	Consonantal	PostTX	120	1,323.60	361.21
	Vocalic	PreTX	72	1,702.59	186.38
	Vocalic	PostTX	72	1,447.59	181.93
Ethan	Consonantal	PreTX	160	1,923.29	361.29
	Consonantal	PostTX	160	1,967.76	369.49
	Vocalic	PreTX	96	2,053.36	317.81
	Vocalic	PostTX	96	1,999.24	297.87
Derek	Consonantal	PreTX	160	2,203.82	387.98
	Consonantal	PostTX	160	469.50	173.70
	Vocalic	PreTX	96	2,004.37	332.09
	Vocalic	PostTX	96	474.25	162.33
Jenna	Consonantal	PreTX	160	1,840.50	422.75
	Consonantal	PostTX	160	1,690.02	509.27
	Vocalic	PreTX	95	1,668.08	424.56
	Vocalic	PostTX	95	1,720.42	366.88

Note. PreTX = pretreatment; PostTX = posttreatment.

the strongest response to treatment, exhibiting 83.3% and 80% accuracy at the sentence level for vocalic and consonantal variants, respectively, by the end of treatment. For Desiree, the clinician's ratings suggested that her performance in the treatment sessions exceeded her accuracy on generalization probes. By the final treatment session, she was judged to produce both consonantal and vocalic variants with greater than 50% accuracy at the sentence level. Even participants who showed minimal or no meaningful gains on generalization probes were judged to exhibit at least a moderate response within treatment for either vocalic or consonantal variants. In her final treatment session, Jenna was judged to exhibit 90% accuracy in producing vocalic /r/ and 30% accuracy in producing consonantal /r/ at the single-word level. In their final treatment sessions, Ethan and Brianna, respectively, were judged to exhibit accurate consonantal /r/ productions in 45% and 50% of their total attempts at the syllable level, although limited gains were observed for vocalic variants. Overall, the within-treatment findings show that, according to the treating clinician's judgments, all participants produced a perceptually acceptable /r/ in at least 50% of trials at some point over the course of treatment. For multiple participants, there were noteworthy differences between performance within the treatment setting and performance on generalization probes; we return to this topic in the Discussion section.

Table 4. Paired *t*-test results per participant for F3–F2 measures pre- and posttreatment.

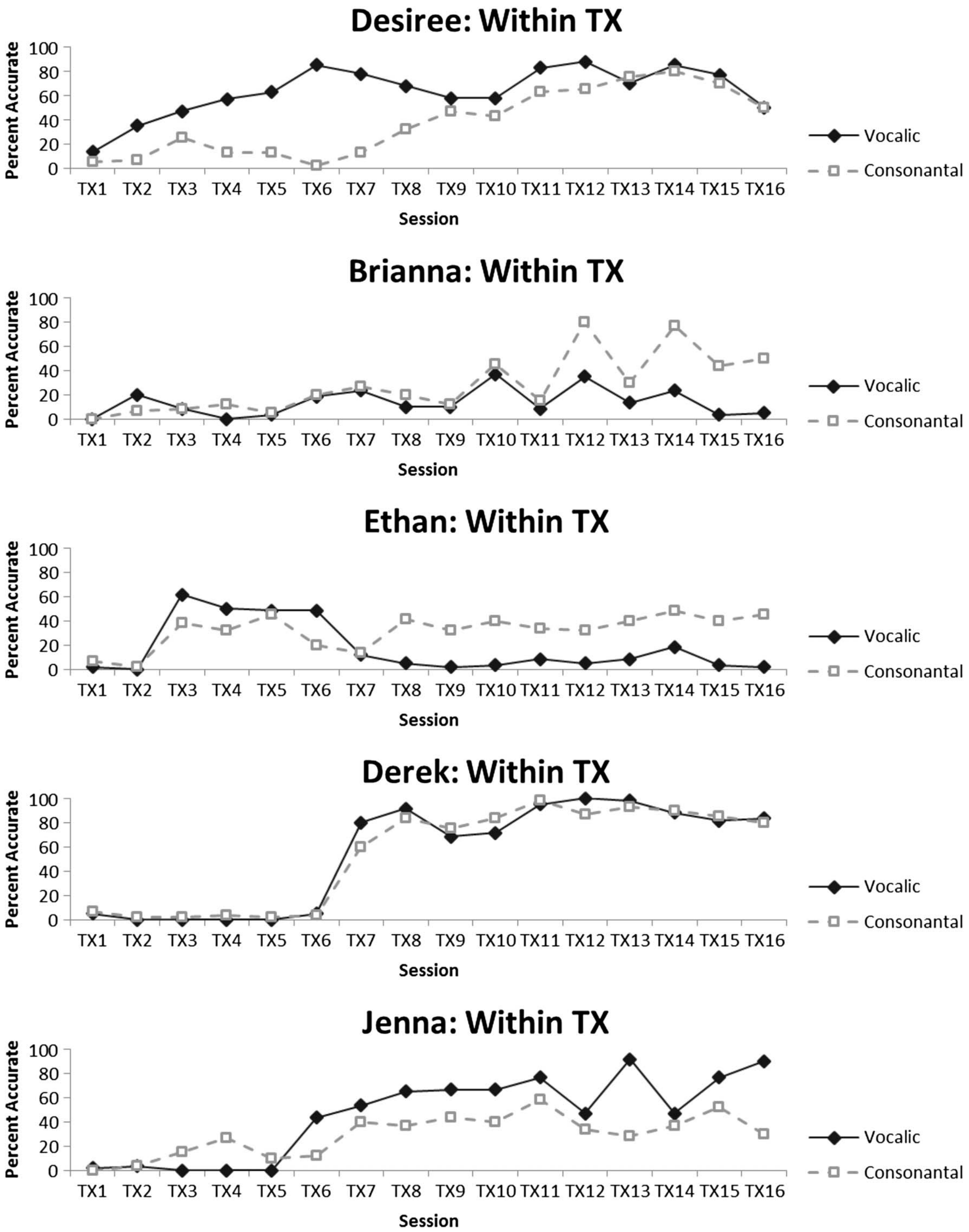
Participant	Variant	Mean difference	*SD*	*t*	*df*	*p*
Desiree	Consonantal	300.55	323.90	10.17	119	<.0001
	Vocalic	358.02	290.76	10.45	71	<.0001
Brianna	Consonantal	27.17	324.12	0.92	119	.360
	Vocalic	255.00	208.14	10.40	71	<.0001
Ethan	Consonantal	−44.47	481.27	−1.17	159	.244
	Vocalic	54.13	460.53	1.15	95	.252
Derek	Consonantal	1,734.32	411.49	53.31	159	<.0001
	Vocalic	1,530.12	360.04	41.64	95	<.001
Jenna	Consonantal	150.48	738.60	2.58	159	.011
	Vocalic	52.34	510.08	−1.00	94	.320

Note. Values were judged to be statistically significant using a standard criterion of $p = .05$ and a conservative criterion of $p \leq .0001$.

Within-Treatment Trials: Acoustic Measurements

Within-treatment F3–F2 measures were collected to validate the treating clinician's judgments of within-treatment changes that may not have generalized to untreated words elicited without biofeedback. Due to the large volume of data collected within treatment, it was not possible to measure every session; evaluating the first and last treatment sessions for each participant would thus seem to be the logical choice. However, because participants advanced through a hierarchy of complexity at different rates within treatment, simply comparing the first and last sessions would mean that different participants would be assessed at different levels of stimulus and task complexity. Given the motoric and linguistic differences associated with productions at different levels of complexity, we did not want to compare syllable- to word- or sentence-level productions. Thus, F3–F2 measures of syllable-level trials in the initial treatment session were compared with measures from the final session elicited at the syllable level of difficulty. Based on their respective patterns of progress in treatment, the selected sessions were TX7 for Desiree and

Figure 4. Within-treatment performance. The *y*-axis represents the percentage of tokens rated perceptually correct by the treating clinician. The *x*-axis represents the treatment session. TX = treatment.



Derek, TX13 for Jenna, and TX16 for Brianna and Ethan. These findings are reported in Table 5.

As with the F3–F2 measures for word probe targets, paired-samples two-tailed *t* tests were used to assess the significance of F3–F2 changes for within-treatment productions for each participant. Mean F3–F2 distances and standard deviations for two within-treatment sessions are reported in Table 5. The paired-samples *t* tests reported

Table 5. Mean and standard deviation of the F3–F2 distance measured within treatment.

Participant	Variant	Time of probe	*n*	Average F3–F2 difference	Average session *SD*
Desiree	Consonantal	Within TX1	30	1,483.36	198.12
	Consonantal	Within TX7	30	1,326.01	486.80
	Vocalic	Within TX1	30	1,426.59	227.43
	Vocalic	Within TX7	30	594.31	269.17
Brianna	Consonantal	Within TX1	30	1,545.68	279.07
	Consonantal	Within TX16	30	1,341.63	357.45
	Vocalic	Within TX1	30	1,627.35	322.35
	Vocalic	Within TX16	30	1,506.22	126.19
Ethan	Consonantal	Within TX1	30	1,909.89	473.32
	Consonantal	Within TX16	30	1,954.33	497.54
	Vocalic	Within TX1	30	1,781.06	474.22
	Vocalic	Within TX16	30	2,051.24	560.64
Derek	Consonantal	Within TX1	28	2,407.03	418.60
	Consonantal	Within TX7	28	1,069.35	346.16
	Vocalic	Within TX1	30	2,105.63	223.17
	Vocalic	Within TX7	30	786.20	235.74
Jenna	Consonantal	Within TX1	30	1,566.92	258.58
	Consonantal	Within TX13	30	1,338.94	731.78
	Vocalic	Within TX1	30	1,506.11	321.19
	Vocalic	Within TX13	30	582.91	177.27

Note. TX = treatment.

in Table 6 revealed statistically significant differences in F3–F2 distances from the initial and final sessions of syllable-level practice for both vocalic and consonantal productions for Derek; these findings align with the treating clinician's perceptual ratings. Statistically significant decreases in F3–F2 distances were also observed for vocalic productions for Desiree and Jenna and for consonantal variants for Brianna. Again, these findings are consistent with the clinician's within-treatment perceptual assessments. No statistically significant or clinically meaningful differences in F3–F2 distances were observed for Ethan. This contrasts with the clinician's perceptual ratings, which indicated gains on consonantal targets. We return to this issue in the discussion.

EPG Tongue-to-Palate Contacts

In the present study, treatment trials were scored as correct if they were perceived as accurate /r/ sounds, independent of whether the child's tongue-to-palate contact pattern matched the target, which emphasized lateral contacts. This decision was made with the rationale that multiple tongue shapes can yield a perceptually accurate /r/, and it was judged to be more functional for communicative purposes to favor perceptual accuracy over articulatory target matching. The present study did not systematically track changes in tongue-to-palate contacts with improved /r/ accuracy. Anecdotally, however, some participants were able to modify an incorrect contact pattern enough to produce a perceptually acceptable new production without matching a specific, canonical pattern of tongue-to-palate contacts (Hitchcock, Lazarus, & Swartz, 2016). We will briefly revisit this issue in the discussion, with a more thorough investigation to follow in a companion paper.

Fidelity

Results of the fidelity check are reported in Table 7. The primary deviation from the stated protocol involved interruptions during a block, which were reported in roughly 11% of blocks. Adequate levels of adherence to the stated protocol were observed for qualitative cues, feedback, and number of trials (91%, 100%, and 93%, respectively).

Discussion

The results of this investigation provide mixed support for the effectiveness of EPG biofeedback treatment for children affecting rhotic speech sounds. All participants, excluding Ethan, showed clear, although variable, perceptual and acoustic gains within the treatment setting, whereas only two (Derek, Desiree) showed word probe gains indicative of generalization beyond the treated targets. The mixed nature of Brianna's gains across perceptual and acoustic probe measures suggests a more questionable response to treatment with respect to generalization. These findings align with past 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; McAllister Byun et al., 2014).

For more robust results, participants' progress in treatment was evaluated using multiple methods: acoustic measures and blinded listeners' perceptual ratings of probe words, as well as acoustic measures and the treating clinician's ratings of within-treatment trials. In general, these different outcome measures painted a consistent picture of

Table 6. Paired *t*-test results per participant for within-treatment F3–F2 measures.

Participant	Variant	Mean difference	*SD*	*t*	*df*	*p*
Desiree	Consonantal	157.35	548.82	1.57	29	.127
	Vocalic	832.28	303.84	15.00	29	<.0001
Brianna	Consonantal	204.05	353.03	3.17	29	.004
	Vocalic	121.13	366.54	1.81	29	.081
Ethan	Consonantal	–44.43	796.12	0.31	29	.762
	Vocalic	–270.18	843.27	1.76	29	.090
Derek	Consonantal	1,337.68	575.46	120	27	<.0001
	Vocalic	1,319.43	300.90	242	29	<.0001
Jenna	Consonantal	227.98	742.93	1.68	29	.104
	Vocalic	923.20	414.83	12.19	29	<.0001

Note. Values were judged to be statistically significant using a standard criterion of $p = .05$ and a conservative criterion of $p \leq .0001$.

participants' progress over the course of therapy. The main exceptions were as follows: (a) The finding that blinded listeners' perceptual ratings for Brianna showed clinically meaningful perceptual changes for consonantal variants not supported by acoustic measures, as well as only significant acoustic changes for vocalic rhotics; (b) the finding that Jenna showed a significant acoustic change for consonantal rhotics that was not corroborated by blinded listeners' perceptual ratings; and (c) the finding that the treating clinician reported progress in the perceptual accuracy of Ethan's syllable-level consonantal productions in the treatment setting, but these perceived gains were not substantiated by acoustic measures.

For both Brianna and Jenna, the significant acoustic changes that were not reflected in listeners' perceptual ratings may reflect a covert contrast, that is, a shift toward a more phonetically accurate production that does not cross the boundary to the target phoneme category and thus may not be perceptually detected (Gibbon, 1999; Munson, Johnson, & Edwards, 2012; Scobbie et al., 2000; Tyler, Edwards, & Saxman, 1990; Tyler, Figurski, & Langsdale, 1993). Covert contrast is often seen as an intermediate stage leading to acquisition of an overt contrast. However, more data tracking of Brianna's and Jenna's progress over time would need to be collected to determine if the significant acoustic differences were a reflection of progress rather than sampling error. In Ethan's case, the simplest interpretation of the discrepancy may be that the clinician's desire to bring about a positive change in participants' speech biased her to perceive a greater degree of progress than Ethan objectively made. Another possibility is that Ethan's consonantal productions during this time were highly variable and the relatively accurate productions that the clinician perceived as correct /r/ sounds were balanced out or even outweighed by highly inaccurate productions in the same session. Similarly, the same could be said for Brianna, whose consonantal productions in word probes were clinically meaningful in blinded listeners' perceptual ratings but not statistically significantly different acoustically. This is consistent with evidence that children engage in highly variable exploratory productions during the early stages of acquisition of a new motor plan (Goffman, Ertmer, & Erdle, 2002; Wu, Miyamoto, Castro, Ölveczky, & Smith, 2014). This possibility is supported by the observation that Ethan and Brianna showed relatively unchanged mean F3–F2 differences with slightly elevated mean standard deviations of F3–F2 measures at posttreatment. Finally, the use of summary statistics to judge discrete gains in phonological acquisition, such as those used to assess formant changes in the present study, have been shown to mask discrete acquisitional trends (Hitchcock & Koenig, 2013). Thus, further exploration of token-by-token formant measures is warranted.

These outcomes raise questions about what factors might account for the observation that treatment gains generalize for some but not all participants. The relevant factors may pertain to the structure of practice, such as the level of scaffolding provided and the dosage of treatment. It is also possible that the likelihood of generalization is influenced by individual characteristics, such as differing levels of tolerance for the palate (either due to oral sensitivity or palate fit). We explore several of these possibilities in more detail below.

Levels of Scaffolding

Generalization learning is thought to be maximized when the learner is challenged to precisely the right degree, such that the task is neither too hard nor too easy (Guadagnoli & Lee, 2004). If a task is too easy and accuracy within practice is very high, there is no opportunity for learning. Conversely, if the task is too hard and accuracy is very low, the learner may become overwhelmed and unable to process new learning. Rvachew and Brosseau-Lapré (2012) adopted Guadagnoli and Lee's principles from the context of general motor learning to the specific

Table 7. Results of fidelity check.

Participant	Session checked (%)	Qualitative cue presented prior to block (%)	Consisted of precisely five trials (%)	Interruptions during a block (%)	Qualitative feedback presented after block (%)
Brianna	6	100	83	0	100
Derek	31	95	95	7	100
Desiree	13	67	95	21	100
Ethan	25	98	92	17	100
Jenna	25	94	98	10	100
Overall mean	20	91	93	11	100

context of speech. They suggest that practice structured according to this challenge point framework could enhance speech-motor learning, specifically with respect to generalization of learned skills, in children with speech disorders. Specific elements that combine to determine the functional difficulty of a speech performance task include the child's skill level, difficulty of the target, level of clinician support, practice schedule, and type of feedback. In the present study, most of the parameters suggested by Rvachew and Brosseau-Lapré (2012) were fixed at the simplest level for most of the treatment period. Task difficulty was increased when a participant demonstrated > 80% success across all within-treatment trials for a specific parameter, but many participants did not meet this criterion until late in the course of treatment. With this high level of support creating a relatively low level of task difficulty, it is not unexpected that participants exhibited a high level of accuracy within practice while showing limited generalization to other contexts.

One possibility is that greater generalization of gains could be achieved with alterations to the protocol for fading of biofeedback. Participants were required to achieve > 80% within a treatment session before EPG biofeedback was faded to 50% of trials, and again for a further reduction to 0% of the within-treatment trials. Advancing through a hierarchy of phonological complexity required the same criterion of at least 80% overall accuracy. A more faithful implementation of the challenge point framework might be achieved if advancement, in the form of reduced frequency of biofeedback or increased target complexity, had been identified in smaller blocks rather than across the summed accuracy for all trials in a session. Modifying the protocol accordingly might have maximized the likelihood that the learner would be challenged to exactly the right degree. For example, Ethan was perceptually judged to make significant gains in production of the vocalic variant across Sessions 3–6. Within this period, certain blocks were observed to have greater than 80% accuracy, but the sum total of the session never reached 80% accuracy. Thus, he was held in the 100% biofeedback mode at a constant level of simple target complexity. We also noted that Ethan demonstrated difficulty with attentional focus throughout the treatment session, which increased as treatment progressed over time. Although we cannot determine definitively what contributed to Ethan's lack of progress in treatment, it is possible that he was underchallenged initially, which could result in diminished investment in the treatment program, as suggested by Rvachew and Brosseau-Lapré (2012). Given the high degree of success observed with Derek, it is possible he was the only participant challenged at the ideal level for learning.

Treatment Dosage

Warren, Fey, and Yoder (2007) suggest that dosage is an important factor in achieving optimal progress in speech and language intervention. They defined dosage as "the number of properly administered teaching episodes during a single intervention session" (Warren et al., 2007, p. 71). It is well documented that changes in speech behaviors are influenced by dosage, with higher dosages providing additional opportunities for practice and thus for improvement. The 60 trials per session elicited in the present study may have been too low a dose to facilitate lasting gains. Other small-scale visual biofeedback studies show from 60 (McAllister Byun & Hitchcock, 2012; McAllister Byun et al., 2014) to 150–228 (Preston et al., 2013) or 175 (Shuster et al., 1995) trials per session. Variable outcomes across these studies further suggest that success in treatment may not depend exclusively on the number of trials within a treatment session but may also be influenced by the total number and scheduling of treatment sessions. It is possible that the duration of the present study, 16 sessions over a course of 8–10 weeks, is not a substantial enough time frame to offset a speech pattern that has been habituated over many years.

Tongue-to-Palate Observations

As noted above, previous literature has reported a number of instances in which participants who made perceptual gains over the course of EPG treatment still showed limited changes in tongue-to-palate contacts (Dagenais et al., 1994; McAuliffe & Cornwell, 2008). These findings suggest that some participants are able to modify an incorrect contact pattern enough to produce a perceptually and acoustically acceptable new production without matching a specific, canonical pattern of tongue-to-palate contacts. Similar considerations arose in the present study, but a thorough examination of these factors warrants a paper in its own right. Thus, a companion paper will examine similarities and discrepancies between perceived accuracy and the typicality of EPG tongue-to-palate contacts in the treatment data elicited in this study.

Considerations for EPG Treatment of Rhotics

It is well known that typical adult speakers realize /r/ with a wide range of tongue shapes (e.g., Delattre & Freeman, 1968). This is thought to reflect a process of finding a configuration that produces the desired acoustics in the context of a speaker's unique vocal tract morphology. Previous clinical research incorporating ultrasound biofeedback suggested 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" (McAllister Byun et al., 2014, p. 2128). EPG provides information about lateral bracing of the tongue, but it yields little information about other features of rhotic articulation, such as the relationship of the tongue tip to the palate, the proximity of the tongue dorsum to the palate, or the degree of pharyngeal constriction. Participants whose patterns of misarticulation are heavily influenced by these other components of rhotic articulation may derive little benefit from the cueing of lateral bracing offered by EPG biofeedback. It is also possible that lateral contacts are more characteristic of one type of accurate rhotic tongue shape—specifically, the bunched /r/ variant, which is produced with a raised tongue dorsum—and is not as well defined or consistently observed for other configurations such as retroflex tongue shapes. Thus, EPG may be a less effective treatment for rhotic misarticulation when compared with other visual biofeedback types because it offers limited options for lingual exploration. Specific comparison of EPG versus other types of biofeedback for rhotic misarticulation is needed to substantiate this hypothesis.

Limitations

Several factors limit the strength of the conclusions that can be drawn from the findings reported here. First, it is possible that using a design that meets What Works Clearinghouse Standards in full (minimum 5 data points per phase), rather than with reservations (minimum 3 data points per phase), might strengthen the validity of our conclusions. Second, the relatively young age and small number of participants makes it impossible to make strong inferences about generalizability to the broader population. Furthermore, the strongest responder, Derek, had not previously received any treatment for rhotic misarticulation, meaning that the possibility exists that he might have responded to traditional treatment. Given the strong within-treatment response across the majority of the participants in the present study, with a subset showing significant generalization gains, there is sufficient evidence to encourage additional studies focusing on the use of EPG for the treatment of rhotic errors. Last, additional single-case research studies could help optimize treatment parameters such as the dosage and frequency with which biofeedback is provided. Moreover, future studies using EPG or other forms of visual biofeedback should consider systematic manipulation of biofeedback within a structured, organized hierarchy such as the challenge point framework (Rvachew & Brosseau-Lapré, 2012). In the longer term, the present findings support the need for a randomized controlled clinical trial evaluating the efficacy of EPG biofeedback, as suggested by Lee, Law, and Gibbon (2009).

In the future, research is warranted to systematically compare the efficacy of EPG versus other types of biofeedback interventions for speech, such as visual–acoustic or ultrasound biofeedback. Such research should compare the different technologies used with respect to treatment outcomes, cost, and ease of use, as well as identify individual patient characteristics that might influence the relative impact of different biofeedback types. It is well known that the use of EPG has certain financial constraints (i.e., cost of computer, software, and palate) and logistical/time-sensitive challenges (i.e., a visit to the dentist's office to take an impression of the child's palate, shipping the mold to the manufacturer, and waiting for the palate via postal service delivery). Furthermore, additional considerations when using EPG include the following: (a) Children may exhibit different levels of tolerance for the palate, which could affect the clinician's ability to deliver the intended treatment dosage; and (b) the palate may not fit comfortably over time due to frequent changes in children's dentition. Although these factors were manageable in the present study, it is possible that some participants may experience more significant problems. Conversely, technological advances have yielded EPG software programs that may be loaded on laptop computers and, thus, are easily portable, making implementation across clinical settings a possibility.

In light of these limitations, research exploring the relative efficacy of different biofeedback types is essential so that clinicians can make an evidence-based selection of the biofeedback option that will represent the most effective and resource-efficient solution for their clients.

Conclusions

The present study was undertaken to evaluate the use of EPG as a visual biofeedback tool to enhance treatment outcomes in children with rhotic misarticulation. The results suggest that EPG biofeedback can be an effective form of treatment for some children whose errors have not responded to traditional intervention strategies, as well as those who have not been previously treated. However, the amount of generalization beyond treated targets was highly variable across participants in the current study. Given these outcomes, it is clear that much remains to be explored about the specific parameters that will yield optimal gains in EPG treatment. Such research should also examine how individual characteristics such as age influence the rate and magnitude of treatment gains achieved by individual participants. Finally, additional research is needed to compare the relative efficacy of EPG versus different visual biofeedback methods. This is particularly important in the context of rhotic misarticulation because EPG offers information about lateral lingua-palatal contacts but does not provide information about other necessary components of rhotic production, including anterior/palatal and posterior/pharyngeal constrictions. This gap in the knowledge base is important to recognize because the cost of biofeedback technologies has declined steadily over the last 20 years, with the consequence that some schools or private practices can realistically consider investing in a device to provide biofeedback. Findings from a comparative biofeedback research study have the potential for significant clinical impact by offering evidence to guide clinicians in choosing a biofeedback technology that is both efficacious and cost-effective.

Acknowledgments

This research was supported by an internal separately budgeted research grant from Montclair State University to the first author. The authors gratefully acknowledge the contribution of the following individuals: for manufacturing participant dental molds free of charge, Drs. Bruce Fox and Sonia Abraham; for stimulus rating, Meghan Hemmer, Sarah Granquist, Olivia Bell, and Diana Barral; and for data collection and analysis, Kurt Keena, Julie Irwin, Lauren Dioguardi, and Melissa Lopez. We also thank our participants and their families for their ongoing cooperation throughout the study. Aspects of this research were presented at the Annual Convention of the American Speech-Language-Hearing Association in Orlando, FL (2014), and the Annual Convention of the New Jersey Speech-Language Hearing Association in Long Branch, NJ (2014).

References

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

Bacsfalvi, P. (2010). Établissement des composantes linguales du son /r/ à l'aide d'ultrasons chez trois adolescents avec un implant cochléaire [Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants]. *Revue canadienne d'orthophonie et d'audiologie, 34*(3), 206–217.

Bacsfalvi, P., Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology, 9*(1), 36–45.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*(4), 161–169.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

Bernhardt, B. M., 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.

Bernhardt, B. M., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics, 17*(3), 199–216.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer* (Version 5) [Computer program]. Retrieved from: http://www.praat.org/

Boyce, S. E. (2015, November). The articulatory phonetics of /r/ for residual speech errors. *Seminars in Speech and Language, 36*(4), 257–270.

Boyce, S. E., & Espy-Wilson, C. Y. (1997). Coarticulatory stability in American English /r/. *The Journal of the Acoustical Society of America, 101*(6), 3741–3753.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Erlbaum.

Carter, P., & Edwards, S. (2004). EPG therapy for children with long-standing speech disorders: Predictions and outcomes. *Clinical Linguistics & Phonetics, 18*(6–8), 359–372.

Cheng, H. Y., Murdoch, B. E., Goozee, J. V., & Scott, D. (2007). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research, 50,* 375–392.

Cleland, J., Timmins, C., Wood, S. E., Hardcastle, W. J., & Wishart, J. G. (2009). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Clinical Linguistics & Phonetics, 23*(12), 926–939.

CompleteSpeech. (n.d). *SmartPalate system orientation guide*. Retrieved from: https://completespeech.com/wpcontent/uploads/SP_OpGuide.pdf

CompleteSpeech Palatometer [Apparatus and software]. (Version V1.0). (2013). SmartPalate International, Orem, UT. Retrieved from: http://www.completespeech.com/smartpalate/

Computerized Speech Lab: Model 4500 [Apparatus and software]. (2012). Montvale, NJ: Pentax Medical

Dagenais, P. A., Critz-Crosby, P., & Adams, J. B. (1994). Defining and remediating persistent lateral lisps in children using electropalatography: Preliminary findings. *American Journal of Speech-Language Pathology, 3*(3), 67–76.

Dalston, R. M. (1975). Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults. *The Journal of the Acoustical Society of America, 57*(2), 462–469.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics, 6*(44), 29–68.

Elbert, M., & McReynolds, L. V. (1975). Transfer of /r/ across contexts. *Journal of Speech and Hearing Disorders, 40*(3), 380–387.

Fabus, R., Raphael, L., Gatzonis, S., Dondorf, K., Giardina, K., Cron, S., & Badke, B. (2015). Preliminary case studies investigating the use of electropalatography (EPG) manufactured by CompleteSpeech® as a biofeedback tool in intervention. *International Journal, 3*(1), 11–23.

Felsenfeld, S., Broen, P. A., & McGue, M. (1994). A 28-year follow-up of adults with a history of moderate phonological disorder educational and occupational results. *Journal of Speech, Language, and Hearing Research, 37*(6), 1341–1353.

Fletcher, S. G., Dagenais, P. A., & Critz-Crosby, P. (1991). Teaching consonants to profoundly hearing-impaired speakers using palatometry. *Journal of Speech, Language, and Hearing Research, 34*(4), 929–943.

Flipsen, P., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: Reference data for residual /ɜ/ distortions. *Clinical Linguistics & Phonetics, 15*(8), 603–630.

Gibbon, F. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*(2), 382–397.

Gibbon, F., & Hardcastle, W. J. (1987). Articulatory description and treatment of "lateral /s/" using electropalatography: A case study. *British Journal of Disorders of Communication, 22*(3), 203–217.

Gibbon, F., & Paterson, L. (2006). A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. *Child Language Teaching and Therapy, 22*(3), 275–292.

Gibbon, F., & Wood, S. E. (2003). Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: A case study. *Clinical Linguistics & Phonetics, 17*(4–5), 365–374.

Goffman, L., Ertmer, D. J., & Erdle, C. (2002). Changes in speech production in a child with a cochlear implant acoustic and kinematic evidence. *Journal of Speech, Language, and Hearing Research, 45*(5), 891–901.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition*. Circle Pines, MN: American Guidance Service.

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.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611–633.

Hagiwara, R. (1995). Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics, 90,* 1–187.

Hardcastle, W. J., Gibbon, F., & Jones, W. (1991). Visual display of tongue–palate contact: Electropalatography in the assessment and remediation of speech disorders. *British Journal of Disorders of Communication, 26*(1), 41–74.

Harel, D., Hitchcock, E. R., Szeredi, D., Ortiz, J., & McAllister Byun, T. (2016). Finding the experts in the crowd: Validity and reliability of crowdsourced measures of children's gradient speech contrasts. *Clinical Linguistics & Phonetics, 31*(1), 1–14.

Hitchcock, E. R., Harel, D., & McAllister Byun, T. (2015, November). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in Speech and Language, 36*(4), 283–294.

Hitchcock, E. R., & Koenig, L. L. (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research, 56*(2), 441–457.

Hitchcock, E. R., Lazarus, R., & Swartz, M. (2016, November). *Assessing tongue-to-palate contact patterns in /r/ remediation using electropalatography*. Poster session presented at the annual convention of the American Speech-Language-Hearing Association, Philadelphia, PA.

Howard, S., & Varley, R. (1995). III: EPG in therapy using electropalatography to treat severe acquired apraxia of speech. *European Journal of Disorders of Communication, 30*(2), 246–255.

Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology, 19*(4), 369–379.

Kent, R. D. (1982). Contextual facilitation of correct sound production. *Language, Speech, and Hearing Services in Schools, 13*(2), 66–76.

Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology, 22*(3), 540–553.

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.

Lee, A. S. Y., Law, J., & Gibbon, F. E. (2009). Electropalatography for articulation disorders associated with cleft palate. *The Cochrane Library*. https://doi.org/10.1002/14651858.CD006854.pub2

Lockenvitz, S., Kuecker, K., & Ball, M. J. (2015). Evidence for the distinction between "consonantal-/r/" and "vocalic-/r/" in American English. *Clinical Linguistics & Phonetics, 29*(8–10), 613–622.

Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(2), 561–578.

Maas, E., Robin, D. A., Hula, S. N. A., 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.

Martin, K. L., Hirson, A., Herman, R., Thomas, J., & Pring, T. (2007). The efficacy of speech intervention using electropalatography with an 18-year-old deaf client: A single case study. *Advances in Speech Language Pathology, 9*(1), 46–56.

Martin, N. A., & Brownell, R. (2005). *Test of Auditory Processing Skills–Third Edition*. Novato, CA: Academic Therapy Publications.

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

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.

McAuliffe, M. J., & Cornwell, P. L. (2008). Intervention for lateral /s/ using electropalatography (EPG) biofeedback and an intensive motor learning approach: A case report. *International Journal of Language & Communication Disorders, 43*(2), 219–229.

McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology, 11*(2), 155–170.

McGowan, R. S., Nittrouer, S., & Manning, C. J. (2004). Development of [ɹ] in young, Midwestern, American children. *The Journal of the Acoustical Society of America, 115,* 871–884.

Miccio, A. W. (2002). Clinical problem solving: Assessment of phonological disorders. *American Journal of Speech-Language Pathology, 11*(3), 221–229.

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

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: A comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology, 21*(2), 124–139.

Powell, T. W. (1993). Phonetic inventory constraints in young children: Factors affecting acquisition patterns during treatment. *Clinical Linguistics & Phonetics, 7,* 45–57.

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.

Ruscello, D. M. (1995). Visual feedback in treatment of residual phonological disorders. *Journal of Communication Disorders, 28*(4), 279–302.

Rvachew, S., & Brosseau-Lapré, F. (2012). *Developmental phonological disorders: Foundations of clinical practice*. San Diego, CA: Plural Pub.

Schmidlin, S., & Boyce, S. E. (2010). */r/ pre and post test*. Unpublished manuscript, University of Cincinnati, Ohio.

Schmidt, A. M. (2007). Evaluating a new clinical palatometry system. *Advances in Speech Language Pathology, 9*(1), 73–81.

Scobbie, J. M., Gibbon, F., & William, J. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. *Papers in Laboratory Phonology V: Acquisition and the Lexicon, 5,* 194.

Shriberg, L. D., Flipsen, P., Karlsson, H. B., & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: An acoustic marker for residual /Z/ distortions. *Clinical Linguistics & Phonetics, 15*(8), 631–650.

Shuster, L. I., Ruscello, D. M., & Smith, K. D. (1992). Evoking [r] using visual feedback. *American Journal of Speech-Language Pathology, 1*(3), 29–34.

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

Stavness, I., Gick, B., Derrick, D., & Fels, S. (2012). Biomechanical modeling of English /r/ variants. *The Journal of the Acoustical Society of America, 131*(5), EL355–EL360.

Tiede, M. K., Boyce, S. E., Holland, C. K., & Choe, K. A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America, 115,* 2633–2634.

Tyler, A. A., Edwards, M. L., & Saxman, J. H. (1990). Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders, 55*(2), 251–261.

Tyler, A. A., Figurski, G. R., & Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech, Language, and Hearing Research, 36*(4), 746–759.

Volin, R. A. (1998). A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task. *American Journal of Speech-Language Pathology, 7*(1), 81–90.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 70–77.

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience, 17*(2), 312–321.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *The Journal of the Acoustical Society of America, 123,* 4466–4481.