AJSLP

Research Article

Investigating the Use of Traditional and Spectral Biofeedback Approaches to Intervention for /r/ Misarticulation

Tara McAllister Byun[a] and Elaine R. Hitchcock[a]

Purpose: Misarticulation of /r/ is among the most challenging developmental speech errors to remediate. Case studies suggest that visual biofeedback treatment can establish perceptually accurate /r/ in clients who have not responded to traditional treatments. This investigation studied the response of children with persistent /r/ misarticulation to a course of traditional treatment and a course of biofeedback treatment.
Method: Eleven children with /r/ misarticulation completed 10 weeks of individual treatment consisting of 4–6 weeks of traditional treatment followed by 4–6 weeks of biofeedback treatment. Progress was measured by tracking correct /r/ productions within treatment and probing /r/ in words at 3 time points.
Results: At the group level, there was no difference in independent judges' ratings of /r/ sounds produced by the children before and after traditional treatment. However, /r/ sounds produced after biofeedback treatment were significantly more likely to be rated by the judges as perceptually correct. Eight of the 11 children made measurable gains in the accuracy of isolated /r/ produced within treatment, with 4 showing significant generalization to untreated /r/ in words.
Conclusion: This descriptive study shows that treatment incorporating spectral biofeedback can facilitate accurate /r/ production in children with treatment-resistant errors. A follow-up period using traditional intervention methods may be necessary to encourage generalization.

Key Words: articulation disorders, residual speech errors, biofeedback, speech sound disorders, intervention

It is estimated that 10% of preschool and school-age children are affected by speech sound delays or disorders (National Institute on Deafness and Other Communication Disorders, 1994). Although speech production will ultimately normalize in most of these children, a subset will continue to show misarticulation even after years of intervention. Errors that have not been eliminated by 8 to 9 years of age are described as residual or persistent speech sound errors (Shriberg, Gruber, & Kwiatkowski, 1994). Even when minor, these errors can cause children and adolescents to be judged more negatively than their peers with age-typical articulation, creating a barrier to social and academic participation (Crowe Hall, 1991). The literature to date has not established a gold standard for intervention for persistent speech sound errors, which have been described as "one of the most neglected research areas in speech therapy" (Gibbon & Paterson, 2006, p. 275).

[a]Montclair State University, Montclair, NJ

Correspondence to Tara McAllister Byun, who is now at New York University: tara.byun@nyu.edu

Editor: Carol Scheffner Hammer
Associate Editor: Peter Flipsen Jr.

Received August 4, 2011
Revision received November 21, 2011
Accepted March 9, 2012
DOI: 10.1044/1058-0360(2012/11-0083)

In a 1995 survey of school-based speech-language pathologists (SLPs), 91% of 98 respondents reported encountering clients whose speech sound errors did not respond to conventional treatment methods (Ruscello, 1995). Forty-one percent of respondents indicated that they had discharged such clients before normalization. The clinician faces a difficult ethical decision in such cases. On one hand, it is undesirable to terminate treatment when evidence suggests that spontaneous resolution is unlikely after around 8 years of age (Gibbon & Paterson, 2006; Shriberg et al., 1994). On the other hand, clients who remain on the caseload year after year with little progress pose a disproportionate drain on clinician time and resources, and unsuccessful treatment is likely to become a source of frustration for both client and clinician. Clinicians in the abovementioned survey expressed acute awareness of these challenges and called for novel, improved intervention approaches for persistent speech sound errors.

Among the most problematic of treatment-resistant speech errors is misarticulation of the phoneme /r/. The /r/ sound can appear as the nucleus of a syllable, in which case it is transcribed /ɚ/ (as in *water*) or /ɝ/ (as in *bird*); this variant will be referred to as "vocalic /r/." It can also appear as a consonant in syllable-initial position in words like *rat* and *rope*. This article will use /r/ rather than /ɹ/ to transcribe the English consonantal rhotic. Finally, /r/ can occur in postvocalic position in words like *fear* and *hair*. Based on articulatory evidence (e.g., McGowan, Nittrouer, & Manning, 2004), this article will treat postvocalic /r/ as the vocalic offglide of a rhotic diphthong (e.g., [fɪɚ], [hɛɚ]). Only syllable-initial /r/ will be treated as a true consonant. Both vocalic and consonantal /r/ were treated and probed as part of the present study.

In a typical speaker, /r/ is produced with one anterior constriction in which the tongue is elevated to a point near the palate and also a posterior constriction in which the tongue root retracts to narrow the pharyngeal cavity (e.g., Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007). Most speakers also produce /r/ with rounding of the lips. Because most speech sounds feature only one major tongue constriction, /r/ has arguably the most complex articulatory configuration of all English speech sounds (Gick et al., 2008). Further complication arises from the fact that the anterior constriction for /r/ is subject to variability across speakers (Delattre & Freeman, 1968). In the *retroflex* variant of /r/, the tongue tip is raised and curled back slightly at a point near the palate. In the *bunched* variant of /r/, the tongue tip is lowered and the tongue body is raised to approximate the palate. The crucial tongue constrictions for /r/ are not externally visible, nor do they provide robust tactile or kinesthetic cues to the speaker. Clinicians thus report that they regard /r/ as one of the most challenging sounds to remediate (Shuster, Ruscello, & Toth, 1995).

Visual Biofeedback

Preliminary research evidence, detailed in the following paragraphs, suggests that intervention incorporating visual biofeedback can be successful in eliminating /r/ misarticulation in children who have not responded to traditional forms of treatment. In visual biofeedback intervention, instrumentation is used to provide information about aspects of speech that are subtle or difficult to perceive under ordinary circumstances. Visual biofeedback provides a visual display of the child's speech and a model of the correctly articulated sound, enabling the child to attempt to modify his or her production to achieve a closer match with the visual model. A variety of technologies can be used to provide feedback for speech, including ultrasound imaging (e.g., Adler-Bock et al., 2007), electromagnetic articulography (e.g., Katz, McNeil, & Garst, 2010), and electropalatography (e.g., Gibbon, Stewart, Hardcastle, & Crampin, 1999). The present study reports on a form of biofeedback in which the client views a visual representation of the acoustic signal of speech. This acoustic biofeedback can take the form of a spectrogram or a linear predictive coding (LPC) spectrum. Both spectrograms and LPC spectra depict the formants or resonant frequencies of the vocal tract, which appear as horizontal bars in the former and vertical peaks in the latter. Using acoustic biofeedback, clients can be taught to recognize and attempt to match the formant pattern that characterizes a target sound.

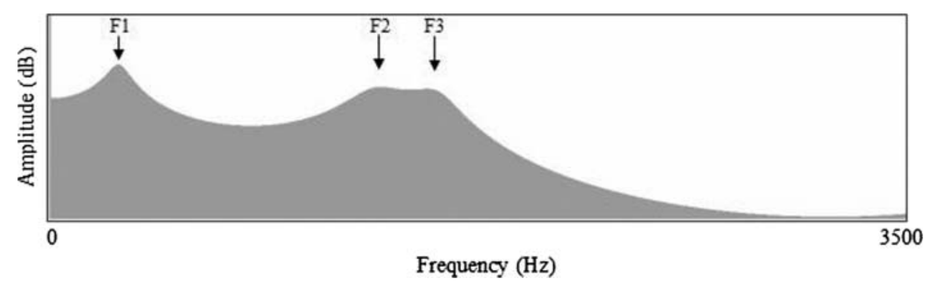
American English /r/ has a distinctive formant pattern that makes it particularly amenable to acoustic biofeedback intervention. A lowered third formant (F3), often so low that it appears to merge with the second formant (F2), is considered the hallmark of American English /r/ (e.g., Boyce & Espy-Wilson, 1997). Figure 1 shows the close spacing of F2 and F3 in an LPC spectrum of vocalic /r/ that was generated using the acoustic analysis software Sona-Match operating on the Computerized Speech Lab (CSL) system (KayPentax, Model 4150B). The acoustic properties of consonantal /r/ are similar to those of vocalic /r/, but the lowering of F3 is generally more extreme in the case of consonantal /r/ (McGowan et al., 2004).

Rationale for Biofeedback

Studies of motor skill learning provide two main reasons to hypothesize that treatment using visual biofeedback may succeed in establishing correct /r/ in children with residual speech errors. First, to master a skilled movement pattern, the learner must associate a particular set of motor commands with a specific set of sensory consequences (Maas et al., 2008; Schmidt, 1975, 2003). In speech, correct production of a sound requires an accurate mental representation of the auditory target. Shuster (1998) found that children with /r/ misarticulation showed a decreased ability to discriminate correct versus distorted /r/ sounds in their own output. She hypothesized that these speakers possessed an overly broad underlying representation that encompassed misarticulated as well as accurate /r/. A child who does not reliably perceive the difference between correct and incorrect /r/ is unlikely to benefit from treatment in which the clinician supplies an auditory model of the target sound and prompts the child to match it. Biofeedback can confer an advantage by offering the alternative of an accurate, well-defined visual target.

The second rationale supporting the use of biofeedback to establish correct /r/ pertains to the difference between externally and internally directed attention during motor skill acquisition. In a motor task such as learning to hit a baseball,

FIGURE 1. Linear predictive coding (LPC) spectrum of vocalic /r/ produced by a typical adult female.



a person who focuses on the movements of the bat would be practicing externally directed attention, whereas a person who focuses on the movements of his or her arms is demonstrating internally directed attention. Studies of nonspeech motor learning have reported that performance is more accurate and less variable when an external focus of attention is adopted (Freedman, Maas, Caligiuri, Wulf, & Robin, 2007). The external direction of attentional focus may also enhance retention of learned motor skills (Wulf, 2007). Visual acoustic biofeedback allows the learner to adopt an external focus of attention for speech movements, which is difficult to accomplish under ordinary circumstances (Maas et al., 2008).

Previous Results

Although previous investigations of visual acoustic biofeedback treatment for persistent /r/ misarticulation have been limited in number and scope, existing results are promising. Two case studies have documented successful application of spectrographic biofeedback to remediate treatment-resistant /r/ misarticulation in three clients ages 10, 14, and 18 (Shuster, Ruscello, & Smith, 1992; Shuster et al., 1995). Before the start of visual biofeedback intervention, all participants demonstrated 0% accuracy in /r/ production, even though all had previously received at least 2 years of conventional treatment. After two to six sessions of spectrographic biofeedback treatment, all participants had attained a benchmark of 70% correct production of isolated vocalic /r/. After 10 to 11 sessions, all participants were producing /r/ in isolation and rhotic diphthongs with 80%–100% accuracy. At the time that Shuster et al. (1995) published their findings, the 10-year-old participant in their study had transferred correct /r/ production to spontaneous conversation and had been discharged from treatment, and the 14-year-old had achieved correct production at the sentence level and was working on transfer to conversation.

Study Goals

Despite the promising nature of the case study results described in the previous paragraph, few clinicians currently use biofeedback for speech sound intervention. Barriers to the adoption of biofeedback include the cost of the associated equipment and the need for additional training to use these technologies. Clinicians are unlikely to invest the time and money needed to overcome these barriers unless biofeedback methods are supported by a strong evidence base. The case studies described above do not reach the requisite level of evidence. The present study was undertaken with the goal of systematically investigating the hypothesis that biofeedback can facilitate correct /r/ production in children with treatment-resistant /r/ errors. This descriptive study documents the gains made by 11 children with /r/ misarticulation as they completed two phases of intervention. In the first phase, all participants received a standardized program of traditional treatment. Immediately afterward, participants completed a similarly structured program using visual acoustic biofeedback in conjunction with the cues trained in traditional intervention. It was hypothesized that children who did not make progress in response to traditional treatment would initiate gains in /r/ production accuracy following the introduction of biofeedback.

Method

Participants

Participants were 11 monolingual native speakers of English ranging in age from 6;0 (years;months) to 11;9 (M_{age} = 9;0). Demographically, all participants were White and were exposed predominantly to a mainstream dialect of American English. Participants were identified primarily by referral from local SLPs. Because most children referred were male, and four out of five females evaluated did not meet all study criteria, 10 out of the 11 participants enrolled in the study were male.[1] The same 10 participants had previously received intervention for /r/ errors, but the one female participant had not been a recipient of treatment before this study. Participants with a previous history of /r/ treatment had received intervention for between 1 and 4 years. A traditional articulatory method of intervention was reported for all of these participants. Nine participants had previously received treatment for other speech sounds, especially /s/, /z/, and /l/. Participants' treatment histories are reported in Table 1.

All participants scored within the average range on a measure of receptive language, the Auditory Comprehension subtest of the Test of Auditory Processing Skills—3 (Martin & Brownell, 2005). Participants also passed a pure-tone hearing test and exhibited no gross structural or functional abnormality in an evaluation of the oral mechanism. However, minor deviations were noted in eight participants, as reported in Table 1. To ensure that speech production skills were largely intact apart from /r/ misarticulation, a 50-utterance spontaneous speech sample was elicited from all participants, and the percentage of consonants correct—revised (PCC–R) was calculated (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). The PCC–R differs from the conventional PCC in that phonetic distortion of a consonant target is not counted as an error. Using PCC–R rather than PCC to define the criterion for inclusion in the study meant that participants would not be ruled out based on the presence of only minor phonetic errors. The methodology described by Shriberg et al. (1997) was modified in that /r/ targets were excluded from the calculation of PCC–R for the current participants.

To be included in the study, participants had to meet at least one of the following conditions: (a) PCC–R after exclusion of /r/ >95%, or (b) PCC–R after exclusion of /r/ within the average range of PCC–R values reported by Shriberg et al. (1997) for the child's age group. Two final

[1]The fact that more males were referred than females is consistent with previous characterizations of the gender breakdown of persistent /r/ misarticulation (Shriberg, 2009). The fact that four out of the five female candidates who were referred did not meet inclusionary criteria is, to the best of our knowledge, the product of random chance. One female candidate was ruled out due to a nonpassing score on the auditory comprehension screening, and the other three were ruled out because they produced other speech sound errors, yielding a percentage of consonants correct–revised (PCC–R; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997) below the 95% cutoff.

TABLE 1. Study participant characteristics and treatment history.

Pseudonym	Age at study onset	Oral mechanism findings	Treatment history	Previous treatment targets
Bob	8;4	Within normal limits	Began treatment at age 2; /r/ targeted from age 7	Multiple errors; specific sounds not known
Charlotte	6;10	Mild cross bite; slightly narrow palatal vault; slight asymmetry of velar elevation	No previous treatment	None
Derek	9;9	Mild open bite	Began treatment at age 5; /r/ targeted from age 6	Multiple errors; specific sounds not known
Jack	9;10	Slight asymmetry of velar elevation	Began treatment at age 7; /r/ targeted from age 7	None
Joe	9;1	Mild class II malocclusion; difficulty dissociating tongue/jaw	Began treatment at age 6; /r/ targeted from age 7	/l/, /s/, /z/, /ð/, /θ/
Leo	10;1	Class II malocclusion; wide spaces between dentition; slightly short frenum; enlarged tonsils	Began treatment at age 5; /r/ targeted from age 8	/l/, /s/, /z/
Maurice	11;9	Within normal limits	Began treatment at age 1;9; /r/ targeted from age 9	Multiple errors; specific sounds not known
Michael	7;9	Close bite; difficulty dissociating tongue/jaw	Began treatment at age 6; /r/ targeted from age 6	/l/, /s/, /z/, /t/, /d/, /n/
Owen	6;0	Wide spaces between dentition	Began treatment at age 5; /r/ targeted from age 5	/l/, /s/
Percy	11;9	Class II malocclusion; minor groping during nonspeech tasks; narrow palate	Began treatment at age 5; /r/ targeted from age 8	/l/, /s/, /z/, all consonant blends
Randy	8;0	Slightly anterior frenum attachment	Began treatment at age 3; /r/ targeted from age 6	/s/, /z/, multiple errors in earlier childhood

Note. Age is reported in years;months.

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 position 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 a baseline of performance before 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: stressed /ɝ/, unstressed /ɚ/, /ɑr/, /ɛr/, /ɔr/, and /ɪr/. No feedback was provided during /r/ probe administration.

Procedure

The goal of this study was to examine whether children who fail to respond to traditional forms of treatment for /r/ can benefit from acoustic biofeedback intervention. Although all but one of the participants in the study had previously received traditional intervention targeting /r/, the nature, duration, and intensity of treatment varied widely across participants. This left open the possibility that some participants' errors could still resolve through conventional methods when a different intensity or technique was used. To ascertain whether participants' /r/ misarticulation could genuinely be characterized as treatment resistant, a period of traditional intervention was provided to all participants before the introduction of biofeedback. The drawback of this design is that, because biofeedback always followed traditional treatment and never vice versa, gains observed during the biofeedback treatment period cannot be characterized as a specific response to biofeedback; the cumulative effect of traditional treatment may also be playing a role. To offset this concern, participants were randomly assigned to switch to biofeedback treatment after 4, 5, or 6 weeks of traditional treatment. If gains were observed only after the switch to biofeedback, independent of the duration of the traditional treatment period, the hypothesis that progress was specifically linked to biofeedback would be supported.

Participants received a total of 10 hr of individual treatment for /r/ misarticulation in two 30-min sessions per week. All treatment was administered by trained graduate students under the supervision of a certified SLP. Student clinicians followed a standard script for treatment delivery. In both traditional and biofeedback conditions, each session elicited 30 attempts at /ɝ/ and 30 attempts at /ɝ/ plus a vowel (10 trials for each of the vowels /i/, /ɑ/, and /u/). The /ɝ/-vowel combinations were intended to encourage consonantal /r/ through successive approximation from vocalic /r/, which emerges earlier than consonantal /r/ in some children (McGowan et al., 2004). To encourage automaticity of the motor plan for /r/, children were instructed to practice producing /r/ for 10 min per day, 5 days per week. Participants' parents maintained a record of compliance in a home practice log. According to this log, the children completed an average of 39 min of home practice per week (range = 18.5 to 50 min).

In both treatment conditions, the clinician provided feedback after every five /r/ trials in the form of a qualitative comment on the client's speech movements (e.g., "I like the way you kept your jaw high"). By commenting on the nature of the movement, clinicians were providing knowledge of performance (KP) feedback, whereas feedback indicating the accuracy of the production would constitute knowledge of results (KR). Verbal KP feedback was used in order to maximize consistency across traditional and biofeedback conditions because visual biofeedback is a form of KP (Volin, 1998).

Participants who produced either vocalic or consonantal /r/ with ≥80% accuracy within a treatment session were advanced to the next step in a traditional hierarchy of phonological complexity. For vocalic /r/, the first step was to produce simple consonant-vowel (CV) and VC nonwords with vocalic /r/ (e.g., *mer, erd*), followed by CV and VC real words. Nonword stimuli were introduced first to encourage participants to try out a new motor plan for /r/, because the motor plan associated with a real word may be more deeply ingrained. Children who achieved ≥80% accuracy with CV/VC real-word stimuli were advanced to CVC nonwords and then CVC real words. For consonantal /r/, participants who reached 80% accuracy for /ɝ/+vowel combinations were prompted to blend vocalic /r/ with the vowel to produce syllables with true consonantal /r/ (e.g., /rɑ/, /ri/, /ru/). Subsequent levels of the hierarchy targeted consonantal /r/ in CV real words, CVC nonwords, and CVC real words. Standard lists of targets were used at all levels.

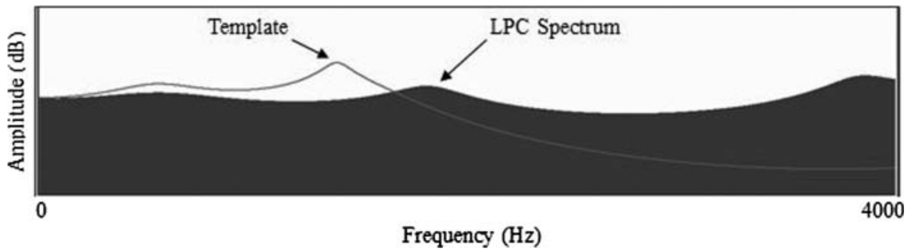
Traditional intervention. Participants received 4, 5, or 6 consecutive weeks of traditional articulatory intervention intended to enhance the accuracy of /r/ production by teaching four components of correctly articulated /r/. Traditional treatment durations of 4, 5, or 6 weeks were randomly assigned to participants such that four participants received 4 weeks of traditional intervention, four participants received 5 weeks, and three participants received 6 weeks. To limit cognitive load, one component of /r/ articulation was introduced as the focus in each week of treatment.[2] In week 1, participants were cued to produce /r/ with an appropriate degree of lip rounding. This cue was provided because North American /r/ is typically produced with rounded lips, especially in prevocalic contexts (Bernhardt & Stemberger, 1998). Week 2 focused on tongue tip placement using Shriberg's (1975) method for eliciting /r/ through successive approximation from /l/. Participants were cued to sustain an /l/ sound and then drag the tongue tip backward along the alveolar ridge until it became necessary to drop the tongue. In week 3, participants were cued to maintain a high, stable jaw position during /r/ articulation. A straw held between the back molars was used to cue high jaw position. Jaw position was targeted because many children with /r/ misarticulation exhibit exaggerated jaw movements (Shriberg, 1980). In the fourth week of treatment, participants were cued to produce /r/ with a high level of tension throughout the entire tongue. This cue was intended to encourage the formation of a constriction with the tongue root, which is essential to normal /r/ articulation (Delattre & Freeman, 1968). Clasping the hands and pushing them against one another was used as a cue to encourage overall muscular tension, including tension in the lingual muscles. Participants who received a fifth or sixth week of traditional treatment were instructed to integrate all four cues from the preceding weeks.

Biofeedback intervention. After completing the traditional treatment phase, participants were again required to complete the stimulability measure and 64-word /r/ probe. Biofeedback intervention was initiated immediately thereafter, with no withdrawal of treatment between traditional and biofeedback phases. The duration of the biofeedback treatment phase was complementary to the duration of the traditional treatment phase, such that a participant who received 4 weeks of traditional intervention received 6 weeks of biofeedback, and vice versa. At the start of the biofeedback condition, participants were introduced to the CSL Sona-Match software program (KayPentax), which displays a real-time LPC spectrum of speech sounds. LPC spectra were selected as the means of providing biofeedback because the display is visually simpler than a spectrogram. The Sona-Match program allows the clinician to load a template representing an appropriate pattern of formant heights for a particular sound. This template can be superimposed over the dynamic LPC spectrum of the client's speech, and the client can modify his or her output to try to make his or her spectrum line up with the template (see Figure 2). Because formant heights are influenced by vocal tract size, several templates of perceptually accurate /r/ were collected from typically developing children of different ages, and participants were provided with a template from the child who represented the closest match for the participant's age, size, and sex.

In the first biofeedback session, participants were encouraged to produce a variety of sounds and observe how the formants ("peaks" or "bumps") move when different sounds are produced. Participants were then familiarized with the concept of matching formant templates in an exercise involving sounds that they could articulate correctly. The clinician presented the template for a vowel and offered the participant a choice among three vowels (e.g., "This sound is /u/, /i/, or /o/"). The participant was then directed to guess which sound it was by trying each of the choices and looking for the closest match. After three successful matches, the target formant configuration for /r/ was introduced. The clinician used models and age-appropriate verbal explanation to show the child that F2 and F3 were far apart in an

[2]Although a single component of /r/ production was emphasized in each session of intervention, treatment trials always targeted the full phoneme /r/, never a single gestural component. This practice was adopted based on evidence that practicing one component of a motor skill typically does not generalize to accurate execution of the entire motor program, and therefore complex movements should be practiced as a whole (Maas et al., 2008).

FIGURE 2. Sona-Match display with LPC formant tracking and correct /r/ template.



incorrect /r/ sound, but they moved close together or merged in correct /r/ production. When a participant could verbally describe the properties of the /r/ target in the LPC spectrum, biofeedback treatment was initiated.

Each biofeedback session began with a period of free-play in which participants were encouraged to use a variety of manipulations to try to make their spectrum match the template. During this period, clinicians provided general encouragement and occasional specific cues (e.g., "Try moving your tongue back and watch what happens to the wave"). A total of 30 tokens of vocalic /r/ were then elicited in sets of five trials. Before each set, the clinician provided one verbal cue describing correct articulator placement for /r/. For the first four sets, cues corresponded to the four components of /r/ production that were trained during traditional intervention. The fifth and sixth sets featured a prompt to integrate all of the preceding cues. Note that because traditional articulatory cues were incorporated into practice with the CSL Sona-Match, this intervention represents a hybrid of traditional and biofeedback treatment rather than a pure form of biofeedback. The decision to integrate traditional methods and biofeedback was made in response to evidence that some participants may require articulator placement cues to benefit from biofeedback (Shuster et al., 1992). After each set of trials, the clinician provided verbal KP feedback (e.g., "I like the way you made the third bump move over"). In addition, the visual and auditory record of each set of five trials was played back, and the participant was asked to identify the token that most closely approximated the /r/ target. As in the traditional treatment condition, 30 trials of /ɝ/ were followed by 30 trials of /ɝ/+vowel.

Data Collection

Previous investigations of the efficacy of biofeedback intervention have used a combination of perceptual rating by SLPs or everyday listeners and acoustic analyses (Bernhardt et al., 2008). In the present study, progress in treatment was assessed through three measures: (a) perceptual accuracy of /r/ in words elicited in pre-, mid-, and posttreatment probes; (b) acoustic measurements of /r/ in words elicited in pre-, mid-, and posttreatment probes; and (c) perceptual accuracy of /ɝ/ and /ɝ/+vowel combinations elicited during intervention. Both performance in the treatment setting and performance on measures of generalization to untreated items were taken into consideration because these are independent aspects of motor learning; research has shown that it is not possible to predict one from the other (Maas et al., 2008).

Perceptual accuracy ratings for /r/ sounds produced in pre-, mid-, and posttest probes were provided by three independent judges who held clinical certification in speech-language pathology. Sound files isolated from all participants' pre-, mid-, and posttest probes were pooled and were presented in random order using E-Prime 2.0 software (Psychology Software Tools). These judges were asked to rate /r/ productions with scores of 1 (*fully incorrect*), 2 (*partially correct but distorted*), or 3 (*fully correct*). They were instructed to be strict in their allocation of the 3 rating. Before rating actual participant stimuli, judges were provided with examples of /r/ productions that were rated 1, 2, or 3 by consensus between two experienced clinicians. However, interrater reliability when using this 3-point rating scale was unacceptably low. There was particular lack of agreement as to which error sounds should be considered distortions versus true substitutions. The rating categories for distortions and substitutions were therefore collapsed into a single "off-target" category, retroactively mapping the 3-point rating scale onto a binary scale. With this change, pairwise interrater agreement increased to an acceptable level of 81%.

As an additional index of /r/ production accuracy, the height of the third formant was measured in each /r/ target that was elicited in the 64-word pre-, mid-, and posttest probes. Acoustic measurements were carried out using Praat acoustic software (Boersma & Weenink, 2010) by graduate students who had prior experience measuring speech stimuli and received additional training for the method used here. Speech samples were digitized at 11025 Hz with a cutoff frequency of 5500 Hz. The students manipulated settings within Praat's automated LPC formant tracking function until the automatically calculated formants were well matched to visible areas of energy concentration in the spectrogram. They were instructed to select a point representing the minimum height of F3 in the /r/ target interval while avoiding points that appeared as outliers relative to adjacent points in the automatic LPC formant track. The height in Hertz of the first three formants was then calculated for a 14-ms Hamming window around the selected point using Burg's method of computing LPC coefficients. If the formant heights thus calculated were judged not to represent a good fit for the visible formant structure, formants were recalculated at a different point or with a different filter order. Although the first three formants were measured through this procedure, only F3 data were used in the present study. Outliers were trimmed by excluding measurements that fell >2*SD*s above the group mean of F3 (n = 27).[3]

The graduate students who took these acoustic measurements were the same students who administered treatment to study participants. We acknowledge that their familiarity with the study introduces some potential for bias. Although student clinicians were blinded to the elicitation condition (pre-, mid-, or posttest) when measuring tokens from the second cohort of five participants, tokens elicited from the first six participants were measured without blinding. To test for bias in these measurements, 15% of the files from the first cohort were remeasured by the same rater in a blinded fashion. An intraclass correlation coefficient (ICC) of .79 was calculated, indicating adequate intrarater agreement across blinded and unblinded conditions. In addition, 10% of the files were remeasured in a blinded fashion by the second student rater. Again, an ICC indicated a sufficient level of agreement (ICC = .81).

Perceptual accuracy of /ɝ/ and /ɝ/+vowel targets elicited within treatment sessions was evaluated by the same two trained graduate students. The students were trained to rate /r/ sounds as *fully accurate* (1) or *off target* (0). Students were required to achieve ≥80% agreement with the ratings of an experienced clinician (the second author) on a sample set of 180 stimuli. Students then rated /r/ trials that were extracted from the recordings of treatment sessions. E-Prime 2.0 software was used to present stimuli in a randomized, de-identified fashion and record responses. Again, we recognize the potential for bias to be introduced by assigning the rating task to the same student clinicians who collected the data and were thus familiar with the voices of some of the participants. However, the large number of stimuli made it impractical to solicit outside raters for this task. To test for an impact of listener bias on rating decisions, 10% of the stimuli were rescored by an independent data rater with no knowledge of the study design or participants. Agreement between the independent data rater and the student raters was 84% for the first student and 85% for the second student. This level of agreement with an unbiased listener supports the validity of the student clinicians' ratings.

Results

Group Results: Perceptual Ratings of /r/ Probe Measures

Independent clinicians' ratings of /r/ in words elicited in pre-, mid-, and posttest probes were analyzed using a generalized linear mixed-effects model (Baayen, 2008). A mixed-effects model includes both fixed (i.e., repeatable) and random (i.e., nonrepeatable) variables. The dependent variable was the perceptual accuracy rating assigned to /r/ in words by blinded clinicians. Because perceptual judgments had been transformed into 1 or 0 ratings, the logit link function for binary data was used. Two separate mixed logit models were created. The first evaluated the influence of traditional treatment on /r/ production accuracy by analyzing the ratings of /r/ productions that were collected at pretest and midtest. The second evaluated the influence of biofeedback intervention on /r/ production accuracy by analyzing the ratings of /r/ productions that were collected at midtest and posttest. Each model included fixed effects of treatment (two levels: before or after treatment) and treatment duration (three levels: 4, 5, or 6 weeks).

Two additional fixed factors were included to examine how properties of the stimuli influenced /r/ production accuracy. Based on evidence that some children acquire vocalic /r/ earlier than consonantal /r/ (McGowan et al., 2004), a fixed factor termed "vocalic" was used to compare vocalic targets against consonantal /r/ in onset position. Likewise, given evidence that some children acquire /r/ in clusters before other contexts (Hoffman, 1983), a fixed factor termed "cluster" was used to compare /r/ in an onset cluster against other types of /r/. Random variables of participant and rater were also included in each model. The random factor of participant, with one level for each child, was included to adjust for the fact that individual children could have different initial levels of accuracy or differing rates of response to treatment. The random effect of rater was included because binary ratings of perceptual accuracy cannot be averaged across judges. Instead, all three judges' ratings of each token were entered into the model and the random effect of rater was included, meaning that the model was adjusted to accommodate the fact that individual raters apply slightly different standards when evaluating the accuracy of /r/ sounds.

In the mixed logit model analyzing the perceptual ratings of /r/ sounds produced before and after the period of traditional treatment, only the factor "cluster" emerged as a significant predictor of accuracy ($\beta = .96, z = 5.58, p < .0001$). The positive coefficient β associated with this factor indicates that /r/ sounds in cluster contexts were significantly more likely to be rated perceptually correct than /r/ sounds in other contexts. This finding will be discussed in greater detail in the following paragraphs. The effect of treatment (pretest vs. midtest) was not a significant predictor of accuracy in this model.

In the mixed logit model analyzing the perceptual ratings of /r/ sounds produced before and after biofeedback intervention (midtest vs. posttest), the factor of treatment was significant ($\beta = .82, z = 8.57, p < .0001$). The positive coefficient indicates that participants' /r/ attempts were more likely to be rated perceptually accurate after the period of biofeedback treatment than before it. The factor "cluster" was again significant ($\beta = .48, z = 3.28, p = .001$), indicating that /r/ sounds were more likely to be rated perceptually accurate in cluster contexts than elsewhere. The factor "vocalic" was also significant ($\beta = -.46, z = -2.90, p = .004$). The negative coefficient indicates that vocalic /r/ was less likely to be rated perceptually accurate than consonantal /r/. The advantage for consonantal over vocalic /r/ was not merely a by-product of the advantage for cluster over noncluster

[3]Because acoustically correct /r/ was relatively infrequent in this sample, F3 measured in correct /r/ sounds could fall >2*SD*s below the group mean. To avoid incorrectly classifying these values as measurement errors, lower outliers were not removed from the sample before analysis.

contexts because the factor "vocalic" remained significant when the model was re-fitted with all instances of /r/ in cluster contexts excluded ($\beta = -.57$, $z = -3.24$, $p = .001$). The finding that consonantal and cluster /r/ were more accurate than vocalic /r/ runs counter to previous findings that children master vocalic /r/ earlier than consonantal /r/ (e.g., McGowan et al., 2004). However, the task of perceptually rating children's speech sounds is more challenging for sounds of short duration than for prolonged sounds. Vocalic /r/ has a longer duration than singleton consonantal /r/, which is in turn longer than /r/ in a consonant cluster. This raises the possibility that consonantal /r/ and especially cluster /r/ received higher perceptual ratings than vocalic /r/ simply because listeners were less able to notice errors in these contexts. This possibility will be revisited in the following section using acoustic measurements of /r/ in different contexts.

In the discussion to follow, it will be shown that two participants (pseudonyms Jack and Leo) made some degree of progress during the traditional intervention period, as revealed by either an improved score on the midtest probe or a >10% increase in the number of correct productions within treatment. These children thus may not be regarded as true cases of treatment-resistant /r/ misarticulation. To determine whether these two most successful participants had an undue influence on the results reported above, the mixed logit model was re-fitted with their data excluded. The effect of biofeedback treatment remained significant ($\beta = .84$, $z = 6.97$, $p < .0001$), indicating that the group-level gains observed after the application of biofeedback cannot be reduced to the progress made by one or two particularly successful individuals.

Group Results: F3 Measurements

Acoustic data were collected to complement the perceptual accuracy ratings described in the previous section. Group means and standard deviations for F3 in pre-, mid-, and posttest conditions are reported in Table 2. Student's *t* test for paired samples was used to compare F3 measurements across testing times. There was no significant difference between F3 values collected at pretest and midtest, $t(677) = .22$, $p = .82$. However, the paired-samples *t* test did reveal a significant difference between F3 values collected at midtest and posttest, $t(677) = 8.38$, $p < .0001$. The difference remained significant if data from the two participants who showed some degree of response to traditional treatment were excluded from the comparison, $t(551) = 8.87$, $p < .0001$. The mean F3 was lower at posttest than midtest, indicating that /r/ sounds produced at the end of the study were acoustically more consistent with adult /r/ than those produced before biofeedback treatment. This is in agreement with the group-level analysis of perceptual ratings.

TABLE 2. Group mean and standard deviation of F3 measured at pretest, midtest, and posttest.

Time of probe	Average F3	*SD*
Pretest	3040.2 Hz	495.0 Hz
Midtest	3036.3 Hz	514.4 Hz
Posttest	2865.7 Hz	573.0 Hz

Earlier, it was suggested that the higher perceptual ratings assigned to consonantal and cluster /r/ categories may be in part a reflection of the fact that listeners have greater difficulty detecting errors in speech sounds of shorter duration. The acoustic data collected as part of the present study provide evidence in support of this hypothesis. Over the full set of productions, mean F3 was very similar across vocalic /r/, singleton consonantal /r/, and consonantal /r/ in a cluster context; means and standard deviations are reported in Table 3. Students' *t* tests revealed no significant difference in F3 between vocalic /r/ and singleton consonantal /r/, $t(244.5) = 0.94$, $p = .35$; vocalic /r/ and cluster /r/, $t(856.3) = 0.63$, $p = .53$; or singleton and cluster /r/, $t(201.2) = -0.58$, $p = .56$. However, differences emerged when the comparison was conducted across /r/ sounds that received a rating of 1 from at least two out of three certified clinician listeners. As can be seen in Table 3, the mean F3 for cluster /r/ sounds rated 1 was substantially higher than the mean F3 for either vocalic or singleton consonantal /r/ sounds rated 1. Students' *t* tests revealed that these differences were significant: vocalic /r/ versus cluster /r/, $t(130.6) = -4.98$, $p < .001$; singleton consonantal /r/ versus cluster /r/, $t(35.2) = -3.08$, $p = .004$. The difference in F3 values between vocalic /r/ rated 1 and singleton consonantal /r/ rated 1 was not significant, $t(45.4) = -0.85$, $p = .4$. Further investigation of the relationship between acoustic measurements and perceptual ratings of /r/ sounds in different contexts is needed before firm conclusions can be drawn. However, these data do suggest that listener-oriented factors played a role in the finding that /r/ sounds were more likely to be rated correct in consonant clusters than in other contexts.

Effect Size and Clinical Significance

The effect size of the change in F3 height between mid- and posttest measurements was calculated using a modification of Cohen's (1988) *d* statistic in which the correlation between measurements at midtest and posttest was used as a correction for paired samples (Morris & DeShon, 2002). The value of *d* was calculated to be .32, which is a small effect size (Cohen, 1988). Thus, even though there was a statistically significant acoustic change in participants' /r/ productions before and after biofeedback treatment, the small magnitude of the change raises questions regarding its clinical significance.

Individual Results: Perceptual Ratings of /r/ Produced Within Treatment Trials

For a more detailed picture of changes in response to traditional and biofeedback treatment for /r/, this section will examine individual participants' trajectories of performance over the course of the study. For each child, Figure 3 reports the number of /ɝ/ and /ɝ/+vowel trials that were assigned a rating of 1 (fully correct) by a blinded graduate

TABLE 3. Mean and standard deviation of F3 in vocalic, singleton, and cluster /r/.

Context	All productions		Productions rated 1 (correct) by at least 2 raters	
	M	*SD*	*M*	*SD*
Vocalic /r/	3093.7 Hz	685.6 Hz	2164.5 Hz	564.4 Hz
Consonantal /r/: Singleton	3031.3 Hz	688.7 Hz	2265.9 Hz	450.6 Hz
Consonantal /r/: Cluster	3067.6 Hz	601.8 Hz	2604.2 Hz	558.1 Hz

student rater. These scores reflect the accuracy of /r/ production in a context where feedback was provided. In Figure 3, the two sessions in each week of treatment are combined, with a single number from 0 to 120 reflecting the total number of perceptually accurate /r/ trials produced in that week. Recall that the transition from traditional treatment to biofeedback was staggered across participants in order to provide information about the specificity of the relationship between gains in /r/ production accuracy and the initiation of biofeedback intervention. A dotted line represents this transition point in Figure 3.

The top section of the graph reports the performance of four participants who received 4 weeks of traditional treatment followed by 6 weeks of biofeedback treatment (pseudonyms Joe, Leo, Michael, and Owen). For one participant in this group, Michael, no change in /r/ production accuracy was observed in either period of intervention. Two participants, Joe and Owen, received near-zero accuracy scores throughout the traditional treatment period but showed improvement after the introduction of biofeedback. Joe's maximum weekly accuracy was 87/120 correct productions (73%). He was generally more accurate in producing vocalic /r/ than consonantal /r/. Owen progressed more rapidly, and by his last week of biofeedback treatment, he produced correct /r/ in 108 out of 120 trials (90%). Like Joe, Owen demonstrated greater accuracy in vocalic /r/ trials. He reached the CVC nonword level of complexity for vocalic /r/ but did not advance to a higher level for consonantal /r/. The last participant, Leo, made progress early in the traditional treatment period. Because his accuracy increased by more than 10% from week 1 to week 2, he was considered to have demonstrated some degree of response to traditional intervention. However, his accuracy during traditional treatment never exceeded the maximum of 25% achieved in the second week of the study. Following the introduction of biofeedback, though, Leo's total accuracy increased to 80% and remained at or above that level for the rest of the study, reaching a maximum of 95% in the fourth week of biofeedback. Leo advanced to the CVC word level for vocalic /r/. He had greater difficulty with consonantal /r/, and his accuracy fell off somewhat in the last 2 weeks as he attempted to blend /ɝ/ with a vowel to produce syllable-initial /r/.

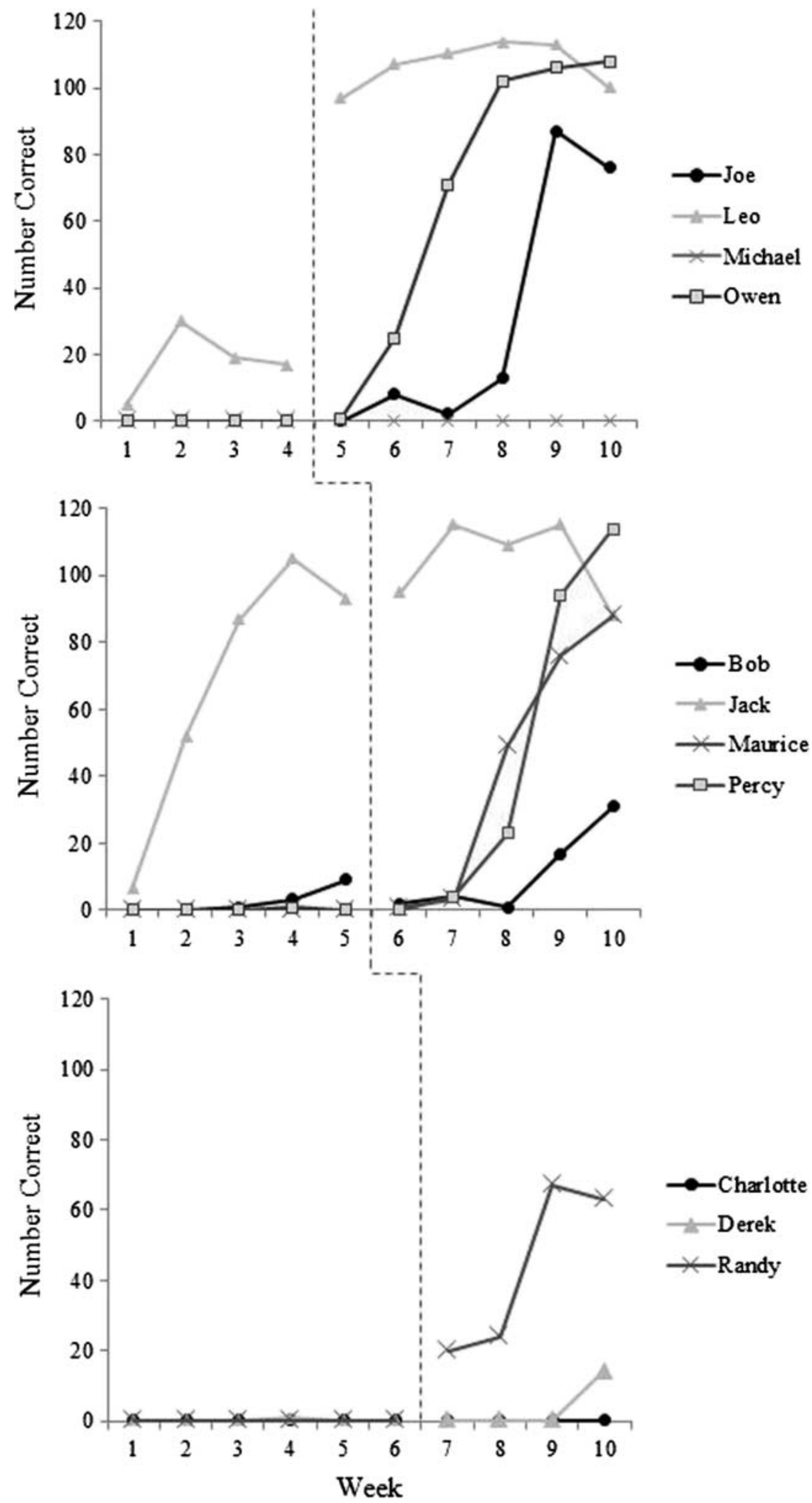
The middle section of Figure 3 depicts changes in /r/ production accuracy for the four participants who received 5 weeks of traditional treatment followed by 5 weeks of biofeedback treatment (pseudonyms Bob, Jack, Percy, and Maurice). Like Joe and Owen in the previous group, Percy and Maurice demonstrated a pattern of near-zero accuracy throughout the traditional intervention period, with rapid gains following the introduction of biofeedback. In the final week of treatment, they reached maximum accuracies of 95% and 73%, respectively. Percy and Maurice displayed no particular advantage for vocalic /r/ or consonantal /r/ in treatment trials. Another participant, Bob, remained below 10% accuracy throughout the traditional treatment phase and the first 3 weeks of biofeedback intervention, but by the final week of biofeedback, he demonstrated a modest increase to a maximum of 26% correct production. Bob's pattern of performance was less typical in that he was judged to produce consonantal /r/ with slightly greater accuracy than vocalic /r/ in treatment trials. The fourth participant, Jack, made pronounced gains beginning in the second week of traditional intervention; he thus cannot be regarded as a true case of treatment-resistant /r/ misarticulation. Jack continued to make progress during biofeedback treatment, advancing through the CVC word level of complexity for vocalic /r/. However, these gains did not immediately transfer to consonantal /r/, and like Leo, Jack showed decreased accuracy in his last week of intervention as he attempted to blend vocalic /r/ with a vowel to produce syllable-initial /r/.

The bottom section of Figure 3 shows the trajectories of /r/ production accuracy for the three participants who received 6 weeks of traditional treatment followed by 4 weeks of biofeedback treatment (pseudonyms Charlotte, Derek, and Randy). Two participants, Charlotte and Derek, made little or no change during either phase of the study. However, Derek did produce his first perceptually accurate /r/ sounds in the final week of biofeedback treatment, achieving a maximum accuracy of 12%. It is possible that further gains might have been observed if biofeedback intervention had continued. The final participant in this group, Randy, produced no correct /r/ trials during traditional treatment but made gains during the biofeedback phase. He reached a maximum accuracy of 53% across consonantal and vocalic /r/. Like Bob, Randy demonstrated a slight advantage for consonantal /r/ over vocalic /r/ in treatment trials.

Individual results: Perceptual ratings of /r/ probe measures

Although gains made in the treatment setting are an important indicator of progress, a child's progress becomes functionally meaningful only if those improvements can be transferred to other targets and settings. The question of generalization is particularly important in the context of biofeedback. If a participant can produce /r/ while receiving visual biofeedback but not when the feedback is withdrawn, these gains are likely to be of limited clinical significance.

FIGURE 3. Number of correct productions of treated stimuli within treatment sessions.



This section will examine generalization by evaluating changes in the accuracy with which individual participants produced /r/ without biofeedback in pre-, mid-, and posttest probes. The /r/ probe measure represented an untreated level of linguistic complexity for participants who did not advance to the word level within treatment (i.e., all participants other than Jack, Leo, and Owen).[4]

Individual mixed logit models were created to evaluate factors influencing the perceptually rated accuracy of /r/ sounds that were elicited in pre-, mid-, and posttest probes. In an individual child's model, a significant effect of treatment with a positive coefficient suggests that the child achieved generalization to untreated words that were elicited without feedback. Again, separate models were created for data that were collected before and after traditional intervention (pretest vs. midtest) and before and after biofeedback intervention (midtest vs. posttest). The individual models were the same as the previously described group models with the random effect of participant eliminated.

All significant effects of treatment that emerged from the individual mixed logit models are summarized in Table 4. In the analysis of data that were collected before and after traditional intervention, treatment emerged as a significant predictor of /r/ production accuracy for the most successful participant, Jack. The positive coefficient indicates that Jack's /r/ sounds in untreated words were more likely to be rated accurate after the period of traditional intervention. Treatment was also a significant predictor of accuracy for Michael. However, in this case, the significant effect of treatment carries a negative coefficient, indicating that Michael's /r/ productions were *less* likely to be rated perceptually correct after traditional intervention than they were at the start of the study.

In the model that examined the impact of biofeedback intervention by analyzing midtest and posttest data, treatment emerged as a significant predictor of /r/ accuracy for four participants: Bob, Jack, Leo, and Owen. The factor of treatment also approached significance for Percy ($p = .05$). For the other three participants who demonstrated progress within the treatment setting, the mixed logit model indicated that generalization to untreated items in a no-feedback context had not reached a significant level. In four out of five cases, the participants who demonstrated a significant or near-significant degree of generalization on the posttest probe measure were the same individuals who showed the greatest accuracy for /r/ trials during intervention. The one exception is Bob, whose progress within the treatment context had appeared much weaker than that of other participants whose mixed logit models did not yield a significant effect of biofeedback treatment, like Joe and Maurice. The present results thus support the notion that progress on treated targets and generalization to untreated contexts are independent aspects of learning that must be probed separately for a complete picture of a child's progress in the course of intervention.

TABLE 4. Significant and near-significant effects of treatment from individual mixed logit models.

	β	*z*	*p*
Pre versus mid			
Jack	1.14	5.14	<.0001
Michael	–1.19	–2.36	.02
Mid versus post			
Bob	1.13	2.97	.003
Jack	.74	3.17	.002
Leo	.97	4.18	<.001
Owen	3.60	8.74	<.001
Percy	.72	1.96	.05

Discussion

At the group level, no significant change was noted in the accuracy of /r/ production before and after a period of traditional intervention. This was true whether accuracy was evaluated using trained listener ratings or an acoustic measure, the height of F3. By contrast, a mixed logit model showed that /r/ sounds that were produced after the period of biofeedback treatment were significantly more likely to be given a rating of 1 (perceptually correct) by certified clinician raters. In addition, F3 frequencies measured at posttest were significantly lower than F3 frequencies measured at midtest, indicating that /r/ sounds produced after the period of biofeedback intervention were closer to the adult acoustic target for /r/ than those produced before biofeedback.

Generally compatible results were obtained when perceptual ratings of the accuracy of /r/ trials elicited within treatment sessions were used to examine individual participants' response to traditional and biofeedback intervention. Nine out of 11 participants in the present study showed a truly treatment-resistant /r/ error: their /r/ misarticulation did not change in the course of 4 to 6 weeks of individual intervention for /r/, in addition to previous treatment received elsewhere. During the phase of biofeedback intervention, six of these participants did exhibit gains in /r/ production accuracy. This supports the hypothesis that some individuals who are unable to benefit from traditional treatment for /r/ misarticulation can make gains in response to biofeedback intervention. However, only two participants with true treatment-resistant misarticulation showed a significant degree of generalization to /r/ in untreated words produced without feedback. In a third participant, generalization of correct /r/ appeared as a trend approaching significance.

It is important that the results reported here be considered alongside a clear statement of what conclusions *cannot* be drawn from the present study. All participants received a structured period of traditional articulatory intervention preceding the biofeedback treatment phase, and traditional cues continued to be incorporated throughout the course of biofeedback intervention. The decision to include traditional intervention techniques was intended to correct for differences in previous treatment history and also to maximize

[4]However, isolated /r/ was also elicited at pre-, mid-, and posttest time points as part of the stimulability probe. The only participants who were judged by an independent rater to have produced perceptually accurate /r/ in the posttest stimulability probe were Jack, Leo, and Owen. This is in agreement with the results of the word-level assessment.

the likelihood of a successful response to intervention, because some participants may need articulator placement cues to benefit from biofeedback treatment (Shuster et al., 1992). The incorporation of traditional cues into biofeedback treatment was also intended to enhance the clinical validity of this study because clinicians who choose to adopt biofeedback are likely to use it in combination with the cues and techniques they have found most helpful in their previous experience administering traditional treatment. The drawback of this design is that the present results cannot provide information about the efficacy of biofeedback intervention when it is not preceded by and integrated with a program of traditional articulatory treatment. First, it is possible that the gains in /r/ production accuracy that were observed during the biofeedback period were actually a late-emerging response to traditional intervention techniques, not a specific response to biofeedback. Evidence from the staggered transition to biofeedback treatment makes this interpretation less likely. Except for the two participants who responded early in the course of traditional intervention, participants who made gains in treatment maintained a stable, near-zero level of accuracy up until the initiation of biofeedback, followed by a sharp increase in accuracy that typically occurred within 2 weeks of the switch. Second, there is no way to know whether participants would have made similar gains if biofeedback treatment were provided with no preceding period of traditional treatment, or if traditional cues for articulator placement were not provided during biofeedback intervention. These are crucial questions from both a theoretical and a clinical standpoint, and follow-up research to address these issues is currently being planned.

Investigation of Predictors of Individual Progress in Treatment

For both theoretical and clinical reasons, it is important to look for individual participant characteristics that could be predictive of a good or poor prognosis in biofeedback intervention. The following factors were investigated as possible predictors of response to biofeedback treatment: participant age, duration of previous intervention for /r/, and history of speech errors other than /r/. We examined the correlation of these three factors with the number of responses rated correct by at least two out of three clinicians on the posttest /r/ probe. No correlation reached significance at the $p < .05$ level.[5] However, it is possible that the small number of participants in the present study prevented these correlations from reaching significance. Brief qualitative comments are thus provided to inform future studies that may investigate these predictors in greater depth.

Although there was no significant correlation between age in months and accuracy on the posttreatment /r/ probe ($r = -.09$, $p = .79$), qualitative examination of our results provides preliminary evidence that participant age may serve as a predictor of progress in biofeedback intervention. The four oldest participants (Maurice, Percy, Leo, and Jack) had generally favorable outcomes in treatment, with three demonstrating significant or near-significant generalization of correct /r/ production to untreated words. Meanwhile, two of the three youngest participants (Charlotte and Michael) remained at 0% correct /r/ production throughout the biofeedback treatment phase. This finding suggests that some young children may not be ready to benefit from spectral biofeedback; further research would be needed to establish whether cognitive, motor-control, or other factors drive this lack of readiness. On the other hand, the youngest child in the study (Owen, 6;0) had one of the most favorable outcomes in treatment, including a significant degree of generalization in the posttest probe measure. There is a clear need for further research to elucidate the relationship between age and response to biofeedback intervention.

There was no significant correlation between the duration in months of previous /r/ treatment and accuracy on the posttreatment probe measure ($r = -.062$, $p = .85$). However, the duration of previous treatment was an estimate provided by the parent, not a confirmed record of intervention history. It also includes no indication of the number of hours of treatment provided, or whether it was provided individually or in a group setting. Therefore, the lack of a significant correlation in this case might reflect the limited nature of the data available. Similarly, although there was no significant correlation between history of errors affecting phonemes other than /r/ and accuracy on the posttreatment probe ($r = .44$, $p = .17$), it is again possible that a finer grained view of the data might yield a different outcome. It has been argued that children who produce distortions of commonly misarticulated sounds such as /r/ and /s/ but have no history of errors affecting other sounds belong to a different diagnostic subtype than children with a history of substitutions or omissions affecting other sounds (Shriberg, 1994). As distinct groups with different epidemiological properties (Shriberg, Flipsen, Karlsson, & McSweeny, 2001), these two subtypes might also be expected to respond differently to biofeedback treatment. However, the data available to us through parent report were not sufficient for the separation of children into these two subtypes. A goal for future work is to collaborate with SLPs who can provide detailed longitudinal data about participating children. This will permit a more meaningful examination of the relationships among error history, treatment history, and response to biofeedback.

Comparison With Previous Findings

Keeping the aforementioned limitations in mind, the present results can be compared to the results of other studies examining intervention for persistent /r/ misarticulation. To date, most treatment research addressing persistent /r/ errors has been limited to the case study level. One exception is a group study that tested the efficacy of a removable appliance that is designed to position the client's tongue in an appropriate configuration for /r/ (Clark, Schwarz, & Blakeley, 1993). In that study, 13/18 children with treatment-resistant /r/ errors (72%) achieved consistent correct production

[5]To accommodate participants who improved during biofeedback treatment but did not achieve significant generalization, we also considered the correlation of these factors with an index of progress within treatment, the greatest number of /r/ trials rated 1 in a single session of biofeedback intervention. Again, none of the correlations examined was significant. Only the results of the correlation with accuracy on the posttreatment /r/ probe are reported above.

of /r/ in isolation after 6 weeks of intervention with the /r/ appliance. The gains made by the /r/ appliance group significantly exceeded the gains made by a group of 18 children who received traditional articulatory treatment with no appliance. Children trained with the /r/ appliance exhibited a significant degree of generalization to untreated words and even to conversation. A large effect size was calculated for the change in /r/ production accuracy reported by Clark et al. (1993; Meline & Schmitt, 1997). These results suggest that the /r/ appliance could represent a more efficient treatment approach than spectral biofeedback because it produced more generalization to other levels of complexity. A drawback of the /r/ appliance is that it requires a dentist's office visit to take an impression of the child's palate, which is an expense and inconvenience that may act as a deterrent to clinicians and parents. Although spectral biofeedback software may require an initial monetary investment, there is no recurring per-child cost associated with this treatment approach.

The present results can also be compared to a study of 13 children ages 7 to 15 who received a combination of traditional and ultrasound biofeedback intervention for persistent /r/ misarticulation (Bernhardt et al., 2008). Participants in that study completed three phases: (a) six to seven sessions of traditional articulatory treatment, (b) between 1 and 3 hr of ultrasound biofeedback treatment, and (c) seven to eight additional sessions of traditional treatment. Production of /r/ in single words was probed before treatment, after the traditional intervention phase, and again after all treatments were completed. The rated accuracy of /r/ production after traditional treatment did not differ from the pretreatment levels. However, a significant increase in /r/ production accuracy was reported at the end of the study, with scores on the posttest probe ranging from 0 to 45 correct productions out of 49 /r/ word targets. Particularly promising results were observed among participants who had received a higher dosage of ultrasound treatment (2–3 hr), with four out of six participants in this group producing perceptually appropriate /r/ in 42–45 words in the posttest probe. The fact that these gains were observed after such a short period of biofeedback suggests that ultrasound may constitute a particularly efficient form of intervention for /r/ errors. However, differences in study design prohibit any conclusions regarding the relative efficacy of spectral and ultrasound biofeedback treatment: Participants in the present study received a longer duration of biofeedback, but they did not receive a follow-up phase of traditional intervention.

Facilitating Retention and Transfer in Biofeedback Treatment

Although a majority of participants in the present study made rapid gains in /r/ production accuracy during acoustic biofeedback treatment, only half of these also made a significant change in the accuracy of /r/ produced at the word level without feedback. The finding that gains made during acoustic biofeedback intervention did not automatically generalize to other contexts is consistent with the results of previous research investigating another form of biofeedback treatment. Studies using electropalatography (EPG) to treat speech sound errors have reported that EPG biofeedback is most effective in the early stages of intervention, when a target sound is first being established (Fletcher, Dagenais, & Critz-Crosby, 1991; Gibbon & Paterson, 2006). Gibbon and Paterson (2006) reported the results of a survey in which SLPs were asked to describe the outcomes achieved by 60 children who had received EPG biofeedback treatment for misarticulation during the period from 1993 to 2003. Eighty-seven percent of the participants were characterized as having achieved improved articulation through biofeedback intervention; however, 88% were described as having at least some difficulty generalizing the articulatory patterns they learned into spontaneous speech.

The notion that biofeedback is effective for establishing but not generalizing motor patterns receives further support from a broader body of research on general principles of motor skill learning. Certain conditions of practice and feedback have proved more facilitative of the initial acquisition of a motor plan, whereas different conditions may maximize retention and transfer (e.g., Maas et al., 2008). Most notably, KP feedback—of which biofeedback is one subtype—has been found to be effective when the motor task is novel or the nature of the target is unclear to the learner (Newell, Carlton, & Antoniou, 1990). However, KP feedback may lose its advantage in later stages of learning; under certain conditions, it may even become detrimental to learning (Maas et al., 2008). Therefore, participants in the present study might have exhibited greater generalization if biofeedback treatment were followed by a phase of nonbiofeedback intervention with only KR feedback. In case studies describing the use of biofeedback for /r/ errors, participants who established correct production through biofeedback were transferred back to their school SLPs for conventional treatment. This follow-up phase could explain why the case study participants described by Shuster et al. (1995) were more successful than participants in the present study in generalizing correct /r/ production to higher levels of linguistic complexity after the end of biofeedback treatment.

Conclusion

The present results support the hypothesis that biofeedback can facilitate perceptually and acoustically correct production of /r/ in children whose errors do not respond to traditional methods of intervention. Out of nine participants judged to present with true treatment-resistant /r/ misarticulation, six demonstrated marked improvement over an interval of biofeedback treatment. However, only half of the participants who made progress within treatment also generalized correct production to contexts in which biofeedback was not provided. Drawing on principles of motor skill learning, it was proposed that biofeedback can be used to establish the basic motor plan for /r/ in clients with persistent errors, but successful generalization may depend on prompt fading of biofeedback and reintroduction of traditional intervention. This proposal requires further investigation, and there is still a need for controlled comparison of outcomes between children receiving exclusively traditional treatment versus exclusively biofeedback treatment. In spite of these limitations, the present results do suggest that visual biofeedback can represent a valuable tool for clinicians who are working with children with persistent speech sound errors, and more widespread adoption of this technology could prove advantageous to clients and clinicians alike.

Acknowledgments

This research was supported by an American Speech-Language Hearing Foundation Research Grant for New Investigators to Tara McAllister Byun. The authors also gratefully acknowledge the contributions of the following individuals: for statistical consultation, Haiyan Su; for clinical support, Sarah Granquist; for stimulus rating, Meghan Hemmer, Lacy MacDonald, Amanda Moorehead, Rose Papera, and Beth Rhein; for data collection and analysis, Elizabeth Varall and Elina Izbinsky; and for insightful comments on the study design and/or manuscript, Joseph Attanasio and Dennis Ruscello. We also thank our participants and their families for their cooperation throughout the study.

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,* 128–139.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Bernhardt, B., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., . . . Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics and Phonetics, 22,* 149–162.

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

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.2.9) [Computer software]. Retrieved from http://www.praat.org/.

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

Clark, C. E., Schwarz, I. E., & Blakeley, R. W. (1993). The removable r-appliance as a practice device to facilitate correct production of /r/. *American Journal of Speech-Language Pathology, 2,* 84–92.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crowe Hall, B. J. (1991). Attitudes of fourth and sixth graders toward peers with mild articulation disorders. *Language, Speech, and Hearing Services in Schools, 22,* 334–340.

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

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

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

Freedman, S. E., Maas, E., Caligiuri, M. P., Wulf, G., & Robin, D. A. (2007). Internal versus external: Oral-motor performance as a function of attentional focus. *Journal of Speech, Language, and Hearing Research, 50,* 131–136.

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

Gibbon, F. E., Stewart, F., Hardcastle, W. J., & Crampin, L. (1999). Widening access to electropalatography for children with persistent sound system disorders. *American Journal of Speech-Language Pathology, 8,* 319–334.

Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2008). A motor differentiation model for liquid substitutions: English /r/ variants in normal and disordered acquisition. *Proceedings of Meetings on Acoustics, 1,* 1–9.

Hoffman, P. R. (1983). Interallophonic generalization of /r/ training. *Journal of Speech and Hearing Disorders, 48,* 215–221.

Katz, W., McNeil, M., & Garst, D. (2010). Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology, 24,* 826–837.

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

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

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

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

Meline, T., & Schmitt, J. F. (1997). Case studies for evaluating statistical significance in group designs. *American Journal of Speech-Language Pathology, 6,* 33–41.

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

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

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

Newell, K. M., Carlton, M. J., & Antoniou, A. (1990). The interaction of criterion and feedback information in learning a drawing task. *Journal of Motor Behavior, 22,* 536–552.

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

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

Schmidt, R. A. (1975). Schema theory of discrete motor skill learning. *Psychological Review, 82,* 225–260.

Schmidt, R. A. (2003). Motor schema theory after 27 years: Reflections and implications for a new theory. *Research Quarterly for Exercise & Sport, 74*(4), 366–375.

Shriberg, L. D. (1975). A response evocation program for /ɝ/. *Journal of Speech and Hearing Disorders, 40,* 92–105.

Shriberg, L. D. (1980). An intervention procedure for children with persistent /r/ errors. *Language, Speech, and Hearing Services in Schools, 11,* 102–110.

Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4,* 38–53.

Shriberg, L. D. (2009). Childhood speech sound disorders: From post-behaviorism to the postgenomic era. In R. Paul & P. Flipsen Jr. (Eds.), *Speech sound disorders in children: In honor of Lawrence D. Shriberg* (pp. 1–34). San Diego, CA: Plural.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40,* 708–722.

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

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research, 37,* 1151–1177.

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research, 41,* 941–950.

Shuster, L. I., Ruscello, D. M., & Smith, K. D. (1992). Evoking [r] using visual feedback. *American Journal of Speech-Language Pathology, 1,* 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,* 37–44.

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,* 81–90.

Wulf, G. (2007). Attentional focus and motor learning: A review of 10 years of research. *E-Journal Bewegung und Training, 1,* 4–11.