AJSLP

Article

Ultrasound Biofeedback Treatment for Persisting Childhood Apraxia of Speech

Jonathan L. Preston,[a,b] Nickole Brick,[b] and Nicole Landi[a,c]

Purpose: The purpose of this study was to evaluate the efficacy of a treatment program that includes ultrasound biofeedback for children with persisting speech sound errors associated with childhood apraxia of speech (CAS).
Method: Six children ages 9–15 years participated in a multiple baseline experiment for 18 treatment sessions during which treatment focused on producing sequences involving lingual sounds. Children were cued to modify their tongue movements using visual feedback from real-time ultrasound images. Probe data were collected before, during, and after treatment to assess word-level accuracy for treated and untreated sound sequences. As participants reached preestablished performance criteria, new sequences were introduced into treatment.
Results: All participants met the performance criterion (80% accuracy for 2 consecutive sessions) on at least 2 treated sound sequences. Across the 6 participants, performance criterion was met for 23 of 31 treated sequences in an average of 5 sessions. Some participants showed no improvement in untreated sequences, whereas others showed generalization to untreated sequences that were phonetically similar to the treated sequences. Most gains were maintained 2 months after the end of treatment. The percentage of phonemes correct increased significantly from pretreatment to the 2-month follow-up.
Conclusion: A treatment program including ultrasound biofeedback is a viable option for improving speech sound accuracy in children with persisting speech sound errors associated with CAS.

Key Words: articulation, efficacy, intervention, children, speech sound disorders, childhood apraxia of speech

Childhood apraxia of speech (CAS) is a subtype of speech sound disorder (SSD) with unique features that include deficits in speech sound accuracy, prosody, coarticulatory transitions, and consistency on repeated attempts at words (American Speech-Language-Hearing Association [ASHA], 2007). These deficits are believed to stem from an impairment in planning or programming movements for speech (ASHA, 2007). The functional impact of CAS can be quite significant, as many children with CAS often have reduced speech intelligibility and speech sound errors that persist well into school age (Lewis, Freebairn, Hansen, Iyengar, & Taylor, 2004). Errors produced by children with persisting SSDs (i.e., past the age of 9 years) are commonly on lingual sounds, although children with CAS may produce inconsistent errors on other sounds and sound sequences as well. Presently, little research is available to guide treatment decisions for these children whose speech errors do not resolve. Therefore, the aim of the present study was to investigate an approach to treatment for children with persisting speech sound errors associated with CAS.

Current treatments for school-age children with CAS involve a variety of approaches, including (a) integral stimulation, which includes an emphasis on sequencing articulatory gestures in increasingly complex words and phrases with manipulation of auditory and visual cues (Strand & Debertine, 2000; Strand & Skinder, 1999; Strand, Stoeckel, & Baas, 2006); (b) phonological awareness training paired with production training (McNeill, Gillon, & Dodd, 2009; Moriarty & Gillon, 2006); and (c) the Prompts for Restructuring Oral Muscular Phonetic Targets technique (Chumpelik, 1984; Hayden & Square, 1994). Because many children with CAS continue to have persisting speech errors in spite of intense treatment, there continues to be a critical need to study the effects of different treatment approaches. A recent review by ASHA (2007) indicated the need for studies "to test the efficacy of alternative treatment programs for children of all ages, types, and severities of expression of CAS" (p. 59).

We took the view that feedback of motor performance is an essential part of learning speech. The speech production system involves feed-forward mechanisms in which a child's representations for speech sounds are paired with prosodic specifications (that govern rate, loudness, and intonation) to plan and execute speech movements (Kent, 2000; Smith, 2006). Feedback is also a critical part of the speech production system, and several current models of production specify

[a]Haskins Laboratories, New Haven, CT
[b]Southern Connecticut State University, New Haven
[c]Yale Child Study Center, New Haven, CT
Correspondence to Jonathan Preston: preston@haskins.yale.edu

Editor: Carol Scheffner Hammer
Associate Editor: Lynn Williams
Received October 16, 2012
Revision received January 21, 2013
Accepted May 5, 2013
DOI: 10.1044/1058-0360(2013/12-0139)

the importance of feedback loops during speech development as well as during online monitoring of phonetic output (Bohland, Bullock, & Guenther, 2010; Hickok, Houde, & Rong, 2011; Terband, Maassen, Guenther, & Brumberg, 2009; Tourville & Guenther, 2011).

The feedback available to a child may include auditory and somatosensory information that compares a speaker's actual productions with the intended plan for the speech sound(s). This feedback allows for adjustments to be made as the movements are being produced—that is, when errors arise during development in feed-forward speech processes (such as planning, programming, or executing movements), feedback mechanisms can be used to detect and correct those sounds (Tourville & Guenther, 2011). When substantial disruption occurs in the feed-forward processes (as may be the case in individuals with CAS), interventions that focus on enhancing feedback may be useful for teaching children to recognize errors and adjust their productions.

Biofeedback

Biofeedback refers to (instrumental) feedback of a physiological function, usually by providing visual information about an individual's performance. The motor learning literature has reported positive results for nonspeech motor learning using various biofeedback approaches (Huang, Wolf, & He, 2006). Biofeedback approaches such as spectrograms and electropalatography have been used in treatment for children with SSDs (Carter & Edwards, 2004; Dent, Gibbon, & Hardcastle, 1995; McAllister Byun & Hitchcock, 2012; Shuster, Ruscello, & Toth, 1995). However, the application of biofeedback training specifically to children with CAS is lacking. The evidence base behind biofeedback approaches is growing, although the expense and availability of appropriate instrumentation and training may limit their use. Because of the intense treatment required for many children with CAS (Campbell, 1999), these alternate approaches might be viewed as clinically and economically viable if they can result in rapid and sustained gains in speech production.

In the context of principles of motor learning, biofeedback is used to provide "knowledge of performance," which is information about the nature of a movement and how it differs from the target movement (Maas et al., 2008; Ruscello, 1995). Knowledge of performance may be useful during the early stages of motor learning to enhance the acquisition of motor skills. This is different from knowledge of results, or feedback on the "correctness" of the motor behavior, which may be more useful for generalization in motor learning (Maas et al., 2008). In speech treatment, knowledge of performance may take the form of clinician feedback on performance (e.g., "The back of your tongue didn't go up when you made that /k/ sound"; e.g., Strand & Skinder, 1999); however, tongue positions and movements, which are not highly visible, can be challenging to verbally cue and describe, and verbal feedback on tongue movements is temporally delayed. Thus, knowledge of performance in the form of real-time visual feedback of tongue positions and movements might overcome these challenges and facilitate motor learning. The hypothesized mechanism of learning, therefore, is that biofeedback may provide knowledge of an individual's performance that can be used to update, modify, and stabilize motor plans for speech.

The current study involved an early-stage investigation of ultrasound as a biofeedback tool designed to provide school-age children with CAS greater knowledge of performance. Ultrasound (the same technology used to image a fetus or a heart) was selected over other forms of biofeedback (e.g., electropalatography) because it does not require custom-fit appliances (e.g., a pseudopalate) and might therefore be more financially feasible to ultimately implement on a larger scale clinical basis. Ultrasound has the advantage of providing an explicit image of tongue configuration in real time and therefore can be used to cue a speaker to modify his or her tongue position when producing lingual sounds. Ultrasound images provide a "moving line" that represents the contour of the tongue. In a sagittal view, the whole tongue cannot be seen in one image, so the transducer is angled slightly forward or back. When angled forward, the tongue blade is usually visible (though the tongue tip is often not). When angled further back, the tongue root can be observed, although a "shadow" often appears in the image due to the hyoid bone. The tongue body is observable in most sagittal images. In a coronal view, the sides of the tongue and the center of the tongue can be observed.

To date, case studies have shown ultrasound biofeedback to be effective at improving phonetic accuracy for speakers with hearing impairment and persisting articulation disorders (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Bacsfalvi & Bernhardt, 2011; Bernhardt et al., 2008; Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005; Modha, Bernhardt, Church, & Bacsfalvi, 2008). For example, Adler-Bock et al. (2007) reported substantial gains in the accuracy of rhotics for two children ages 12 and 14 years whose speech errors had been resistant to previous treatment. Bernhardt et al. (2005) provided evidence that adolescents with hearing impairment and residual articulation problems learned vowels, liquids, and lingual fricatives and affricates using ultrasound biofeedback. The present investigation explored the application of this approach specifically to the study of children with speech sound errors associated with CAS.

Purpose

The primary purpose of the study was to determine if a treatment approach that includes biofeedback of tongue movements would improve the accuracy of target speech sequences in school-age children with persisting speech sound errors associated with CAS. We hypothesized that using ultrasound biofeedback can result in improved accuracy of treated sound sequences. By teaching children about articulatory targets using visual feedback of tongue movements, and by sequencing these movements in various words/phrases/prosodic contexts, the relationship between the speech motor plan and the actual movements may be strengthened.

Method

Participants

Six children, ages 9;10 (years;months) to 15;10, with persisting CAS were recruited through contact with local schools and clinics. All of the participants were Caucasian males and were from homes in which at least one parent had attended college. All had longstanding histories of a diagnosis of CAS based on parent report and clinician referral. Additionally, all had been receiving speech-language services since the age of 2 or 3 years, were reportedly making limited progress in their speech sound accuracy, and were enrolled in speech-language treatment through their schools at the time of the study.

To confirm the diagnosis of CAS, a licensed speech-language pathologist (SLP) who was not involved in the treatment administered the assessment protocol described below; the first author, also a licensed clinician, was present but only observed the evaluation. Both the SLP and the first author had to agree that they observed signs of CAS based on the following criteria. Children with CAS were required to exhibit an SSD as defined by at least 1.5 *SD*s below the mean on the Goldman Fristoe Test of Articulation—Second Edition (GFTA–2; Goldman & Fristoe, 2000). Additionally, because treatment was designed to target the production of lingual phonemes, the children had to exhibit errors that could be addressed using biofeedback of the tongue.

To distinguish children with CAS from those with residual SSDs that are not associated with CAS, participants were also required to score below 85% on the Sequencing subtest of the Verbal Motor Production Assessment for Children (VMPAC; Hayden & Square, 1999), which assesses the ability to sequence consonants (e.g., /pʌtʌkʌ/) and vowels (e.g., /u-i-a/). Additionally, detailed speech samples were collected and audio recorded from four measures: (a) the GFTA–2; (b) the Word Inconsistency subtest of the Diagnostic Evaluation of Articulation and Phonology (Dodd, Hua, Crosbie, Holm, & Ozanne, 2002); (c) a 125-item picture-naming task assessing many consonant clusters, multisyllabic words, and all English consonants and vowels at least twice (Preston, 2008); and (d) 17 imitated sentences that included many samples of later developing sounds, including liquids and sibilants (e.g., *The computer screen flashed*). The percentage of consonants correct (PCC) and PCC Late-8 were computed from the audio recordings (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). It was required that children with CAS produce sequencing errors as defined by omissions or additions of sounds or syllables in phonologically complex words, metathetic errors (i.e., switching sounds in words), and/or migration errors (i.e., sounds moving to other positions in the word). In addition to the six participants who met the criteria for the treatment study, five children were screened but did not meet the eligibility criteria.

Based on these detailed speech samples, all of the children produced errors on rhotics, but all had other errors as well. For example, several participants had distortions or substitutions of vowels (such as /i, ɛ, e/) and consonants (including alveolars /s, z, t, d, n, l/ and affricates /ʤ, ʧ/) or omissions of sounds in consonant clusters such as /sk, kl, dr/. Treated sound sequences were selected individually based on a review of phonetic transcriptions from the speech production measures listed above, while considering perceived impact on intelligibility and imageability with ultrasound. For example, if a participant had consonant singletons and vowels in error (such as /r, n, e, i/), CV and VC sequences were selected that paired these phonemes so that both elements could be addressed for accuracy (such as /re/ and /in/). If a particular cluster was observed to be in error multiple times during the testing (such as /sk/ clusters, in spite of correct /s/ and correct /k/ as singletons), this sequence was selected for treatment.

Prior to treatment, the children participated in additional testing to quantify their language and cognition abilities. Measures included the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999); the Peabody Picture Vocabulary Test, Fourth Edition (PPVT–4; Dunn & Dunn, 2007); the Expressive Vocabulary Test, Second Edition (EVT–2; Williams, 2007); the Recalling Sentences and Formulated Sentences subtests of the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF–4; Semel, Wiig, & Secord, 2003); and the Elision and Blending subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999). The children's scores on these measures are presented in Table 1.

All participants attended speech-language treatment at their schools. However, to enhance the validity of the study, the researchers were in contact with school-based clinicians, who agreed to put their focus on treating aspects of communication other than articulation of the target sounds (e.g., by focusing on written language, syntax, etc.).

Intervention Design, Probes, and Target Sequences

A multiple baseline across behaviors design was used (replicated across six participants). Eight target sequences were selected per child to be probed at every session. Following three baseline pretreatment probes, one target sequence was selected for treatment, and the remaining seven were untreated. Two to four sessions later, a second target sequence was added, and the remainder of the sessions involved treating two target sequences. Once the child achieved 80% accuracy for a treated target sequence on two consecutive probes, treatment was discontinued on this target sequence (although probes were still administered), and a new target sequence replaced the one that had reached criterion. Additionally, to avoid frustration, if 0% improvement was observed after six consecutive sessions, the target was discontinued and a new target was introduced (this occurred only for the /ar/ sequence for participant U007). Treatment was provided for a total of 18 sessions, with no more than two treatment targets being addressed during any one session.

Probe data for sound sequences. Based on the pretreatment data, eight sound sequences that involved errors on lingual sounds were identified for each participant (e.g., /re/, /ir/, /kl/, etc.). Eight words associated with the target sequence were probed at the end of each session (8 target sequences × 8 words per sequence = 64 words). The eight words for any

Table 1. Descriptive data for the six male participants with childhood apraxia of speech (CAS) before treatment.

	Participant					
Variable	**U002**	**U005**	**U007**	**U008**	**U009**	**U012**
Age	9	12	13	15	12	13
GFTA–2 standard score	46	61	<40	69	59	43
VMPAC Sequencing	69.5	80	78	85	80	83
VMPAC Focal Oral Motor	93	98	87	95	90	91
PCC from 125-item picture naming	80	81	65	97	89	93
PCC-Late-8 from 125-item picture naming	53	69	19	81	70	74
WASI Matrix Reasoning *T* score	45	39	53	40	45	65
EVT–2 standard score	103	93	80	89	83	133
PPVT–4 standard score	97	78	92	92	83	126
CTOPP Elision scaled score	12	5	7	3	8	8
CTOPP Blending scaled score	8	9	4	6	7	10
CELF–4 Formulated Sentences scaled score	8	7	6	4	5	7
CELF–4 Recalling Sentences scaled score	13	1	6	5	2	10
Other clinical concerns	PDD-NOS	ADHD, LI, RD	Trisomy 8, LI, limb apraxia, dysarthria, VPI	LI, RD	LI	OME, hypernasal
Judgment of CAS severity	Moderate	Mild-moderate	Severe	Mild	Mild-moderate	Moderate

Note. GFTA = Goldman Fristoe Test of Articulation—Second Edition; VMPAC = Verbal Motor Production Assessment for Children; PCC = percentage of consonants correct; WASI = Wechsler Abbreviated Scales of Intelligence; EVT–2 = Expressive Vocabulary Test, Second Edition; PPVT–4 = Peabody Picture Vocabulary Test, Fourth Edition; CTOPP = Comprehensive Test of Phonological Processing; CELF–4 = Clinical Evaluation of Language Fundamentals—Fourth Edition; PDD-NOS = pervasive developmental disorder, not otherwise specified; ADHD = attention deficit hyperactivity disorder; LI = language impairment; RD = reading disability; VPI = velopharyngeal insufficiency; OME = history of otitis media with effusion.

given target sequence always included 3–4 monosyllabic words and 4–5 multisyllabic words. For example, for target /re/, probe words included *rain, rake, race, range, raisin, railway, radio,* and *racecar*. Only half of the words on the probe list were treated. Therefore, for the child to reach the preestablished criterion of 80% accuracy, generalization to untreated words containing the target sequence was necessary. Scoring of probes focused exclusively on the accuracy of the target sequence; hence, for the target /kl/, if a child produced *closet* with the /kl/ sequence correct, the production was scored as accurate, regardless of the accuracy of the remaining sounds. Accuracy on the probes was scored live by the treating clinician and by a second listener (either live or via audio recording).

Probes were administered via direct imitation without feedback in order to track the progress of treated target sequences and monitor untreated sequences. Each session, all of the 64 words (8 words from 8 target sequences) were elicited once, then half of the target sequences were elicited two more times to obtain a more reliable score (hence, probe scores were based on either 8 or 24 productions of the target sequence). The larger sample was elicited from sound sequences that were treated that day and from two other sequences that were rotated. For example, at the end of a session when /ar/ and /kl/ had been treated, the larger sample (24 productions) was elicited for /ar, kl/ and from two untrained sequences /ks, ru/; the smaller sample (8 productions) was elicited for the remaining sequences /re, gr, sk, rz/. The Appendix provides an example of a probe list and the items that were scored during one session.

The eight sound sequences per child were selected to be phonetically dissimilar if possible (i.e., sampling different consonants, vowels, and syllable positions). In some cases, it was possible to collect probe data on a sound sequence that was unlikely to show generalization effects (e.g., U002 was probed on /si/, and no treatment on alveolar sibilants was provided). However, because of the nature of individual error patterns, phonetically similar targets were sometimes selected and therefore revealed generalization effects.

Treatment Procedure

Treatment was provided either by a certified SLP (the first author, for 26% of sessions) or a graduate student who was usually supervised by the first author (for 74% of sessions). Treatment sessions were scheduled twice per week for 60 min each and usually occurred after school. It took between 10 and 16 weeks to complete the 18 sessions due to holidays, illness, and so forth. Thirty minutes of each session were devoted to ultrasound biofeedback. However, we sought to facilitate motor learning without children relying exclusively on biofeedback. Therefore, 15–20 min of each treatment session were devoted to other tabletop activities in which more traditional approaches were used to target the same sequences; this included drill and drill-play activities (e.g., card games, Jeopardy games) using speech sound training techniques such as modeling and imitation of target words, shaping to elicit sounds in isolation (when necessary), phonetic cues/verbal descriptions related to articulatory positions and movements, and self-monitoring practice (Bernthal, Bankson,

& Flipsen, 2008; Secord, Boyce, Donohue, Fox, & Shine, 2007).

Up to two target sequences were addressed per session, using the ultrasound in two blocks of 15 min (a timer was used to ensure adherence to this). An example session outline would be as follows: 15 min biofeedback on /re/, 8–10 min of a tabletop activity, 15 min biofeedback on /kl/, 8–10 min of a tabletop activity, and 10 min for administration of the probe list for data collection. Data from individual practice attempts during the biofeedback and tabletop activity were not collected because we wished to compare accuracy on the probe lists of treated and untreated sound sequences under the same conditions (i.e., imitated words without feedback).

During biofeedback blocks, real-time ultrasound images were used to teach the participants visual representations of tongue movement sequences. An Interson PI 7.5 MzH ultrasound transducer was connected to a Dell Precision laptop with a 17-in. screen. The ultrasound images were video recorded using screen capture software. Either the ultrasound transducer was placed beneath the child's chin and the child held the transducer in place, or a microphone stand with a clamp was used to keep the transducer in place as the child leaned on it. Depending on the nature of the sound sequence that was targeted, either a sagittal (front-to-back) or coronal (left-to-right) view was used. During the first session, children were oriented to the image, and they all showed understanding of the visual display by the end of the first session.

Treatment focused specifically on the target sequence by teaching the tongue configurations needed to produce the speech sounds clearly. In general, a slower speaking rate was used during most productions with the ultrasound, as this facilitated use of the visual feedback display. The visual display from the ultrasound provided feedback to the child as he spoke, and the clinician used this feedback to cue tongue gestures. For example, to cue /k/, the tongue dorsum was identified, and elevation of the dorsum was demonstrated and described. A visual target was provided using a transparency over the laptop screen with marks for the child to "hit" with the tongue dorsum. To cue /r/, the clinician cued multiple aspects of articulation depending on the child's error pattern. For example, a sagittal view could be used to cue elevation of the tongue tip/body. (No a priori assumption was made about whether a retroflex or bunched tongue configuration was best; this was decided on a case-by-case basis.) A coronal view could be used to cue elevation of the lateral margins of the tongue while creating a "dip" in the center of the tongue.

To elicit correct productions, both verbal and visual cues were provided, including shaping techniques (Secord et al., 2007; Shriberg, 1975), until a clear production was achieved. The biofeedback along with verbal description of the target sequence were provided on all practice attempts until the sequence reached at least five correct productions in isolation or syllables in a 15-min block; then, treatment commenced at the syllable and word levels. Some participants achieved five correct productions in isolation or in syllables in as few as 15 trials, whereas others were never able to produce five correctly in a 15-min block. In these instances, to avoid frustration, participants advanced to the word level after ~10 min of training in isolation/syllables. Although the frequency of verbal feedback was not systematically dictated, a mix of verbal feedback was provided along with the ultrasound: feedback about movement (e.g., "I saw/didn't see the tongue tip go up"), about general accuracy based on the visual display ("Yes, that looked right" or "No, that didn't look quite right"), and on how the production sounded (e.g., "That sounded clear" or "That didn't sound quite right"). As children learned the primary aspects of tongue movement that were desired for the sounds, they were encouraged to cue themselves by pointing to the screen to identify specific aspects of tongue movement that were being addressed.

Eight to nine words (both monosyllabic and multisyllabic) containing the target sequence were addressed during each session, with only four of these words coming from the probe list. Once the participants achieved accuracy on isolated sounds, syllables, and words, backward chaining was used to facilitate multisyllabic productions (Chappell, 1973; Young, 1987). For example, when training /re/ in *race,* the monosyllabic word could be used to elicit a more complex production such as *erase*. Additionally, short cloze-type phrases were also introduced. These phrases were designed to either begin or end with the target sequence. For example, to target /re/, the participants were encouraged to complete the phrase "race ___" (e.g., *race to the store, race home*); to target /ar/, a phrase such as "___ car" was used (e.g., *blue car, dirty car*). This was done to add simple linguistic formulation demands while also requiring the planning of multisyllabic utterances. The biofeedback was then used to focus the participant on the movement at either the beginning or end of the short phrases.

Although the primary focus of our study was on articulatory accuracy, children with CAS are often observed to produce errors in speech sounds as well as errors in loudness, rate, and intonation (ASHA, 2007; Ballard, Robin, McCabe, & McDonald, 2010; Shriberg, Aram, & Kwiatkowski, 1997). Feedback on tongue movement is not intended to improve prosody, but incorporating prosody into treatment may facilitate coordination of tongue movement with respiratory/phonatory mechanisms involved in planning speech. Thus, our treatment approach incorporated prosodic manipulation by encouraging varied practice through cuing alterations in rate, intonation, and loudness of the word or phrase (e.g., "Say it slower/faster," "Make it a question," or "Say it quiet"; cf. Strand & Skinder, 1999). However, feedback was not given on prosody, only on tongue movement.

A research assistant reviewed the video recordings of the ultrasound blocks from four sessions per participant (24 sessions). In 52% of the production trials (practice attempts) with the biofeedback, participants relied only on the visual feedback (i.e., they were engaged in self-evaluation based on the visual feedback). In 32% of trials with biofeedback, verbal feedback on accuracy was also provided by the clinician (e.g., "You got it" or "Not quite"). On 16% of trials, a qualitative description of the articulation accompanied the visual display (e.g., "I saw your tongue tip go up"). Biofeedback was provided for 30 min per session, not a specified number of trials. However, a high rate of practice trials was sought (cf. Edeal & Gildersleeve-Neumann, 2011). Of the 24 sessions reviewed, an average of 228 trials were elicited with biofeedback per session (SD = 80), and 22 of the 24 sessions yielded at least 150 trials.

2-Month Follow-Up Assessment

At a 2-month follow-up, the GFTA–2 and the 17-sentence imitation task were re-administered to the participants. A research assistant who was blind to the intervention status of the participants transcribed recordings of the GFTA–2 and 17 sentences (total of ~650 phonemes) from both before treatment and 2 months after the treatment had ended. PCC was used as a global measure of speech sound accuracy to compare the pre- and posttreatment results. Probe lists were also re-administered to assess whether gains on the sound sequences that had been observed during treatment were maintained 2 months later. As with the probes collected before and during treatment, the posttreatment probe data reflect the average of two listener judgments.

Interrater Reliability

Two listeners (the first author and the treating clinician or a research assistant) scored the responses on the probes either live or via an audio recording (although in a few circumstances, audio recording errors prevented this double scoring). Interrater agreement was as follows: 79.3% for U002, 84.4% for U005, 91.5% for U007, 88.1% for U008, 84.3% for U009, and 82.6% for U012. Cohen's kappa was 0.67, representing good agreement. The final data are presented in Figures 1–6 and represent the average of the two listeners' judgments.

Results

Data from the probes were used to calculate percent accuracy for each of the eight sound sequences per child. Across the six participants, 23 of 31 treated sound sequences reached the performance criterion of 80% accuracy for two consecutive sessions (the number of targets treated varied across participants because some met criterion more quickly than others). An average of five sessions were needed for these 23 sequences to reach performance criterion. As shown in the figures, a high degree of accuracy was observed during posttreatment probes for most of the treated sound sequences that met criterion, showing retention after the performance criterion was met. Eight of the 31 treated sequences did not meet performance criterion (see Table 2); half of these eight were introduced toward the end of the study period (i.e., just prior to the 18th session) and received as few as three treatment sessions.

Summary statistics were computed for each treated sound sequence (see Table 2). Of the treated sequences, an average increase of 53% was observed from the pretreatment mean to the posttreatment mean. The standardized mean difference (SMD; Gierut & Morrisette, 2011; Olive & Smith, 2005) was computed by calculating the mean difference between pre- and posttreatment divided by the baseline standard deviation. Hence, an SMD of 1.0 represents an increase in 1 SD from baseline. Each participant showed an average increase of at least 2 SDs on his treated sound sequences. The percentage of nonoverlapping data points (the percentage of posttreatment data points that were above the highest pretreatment data point) indicated consistently higher accuracy posttreatment than pretreatment for most of the treated sound sequences.

Table 3 presents PCC based on a listener's blind rating of the target words on the GFTA–2 and the 17-sentence imitation task before and 2 months after treatment. All participants showed higher scores after treatment as judged by the blind listener, although the magnitude of change was modest. A Wilcoxon signed-ranks test for related samples confirmed statistically significant differences from pretreatment to the 2-month follow-up (p = .028).

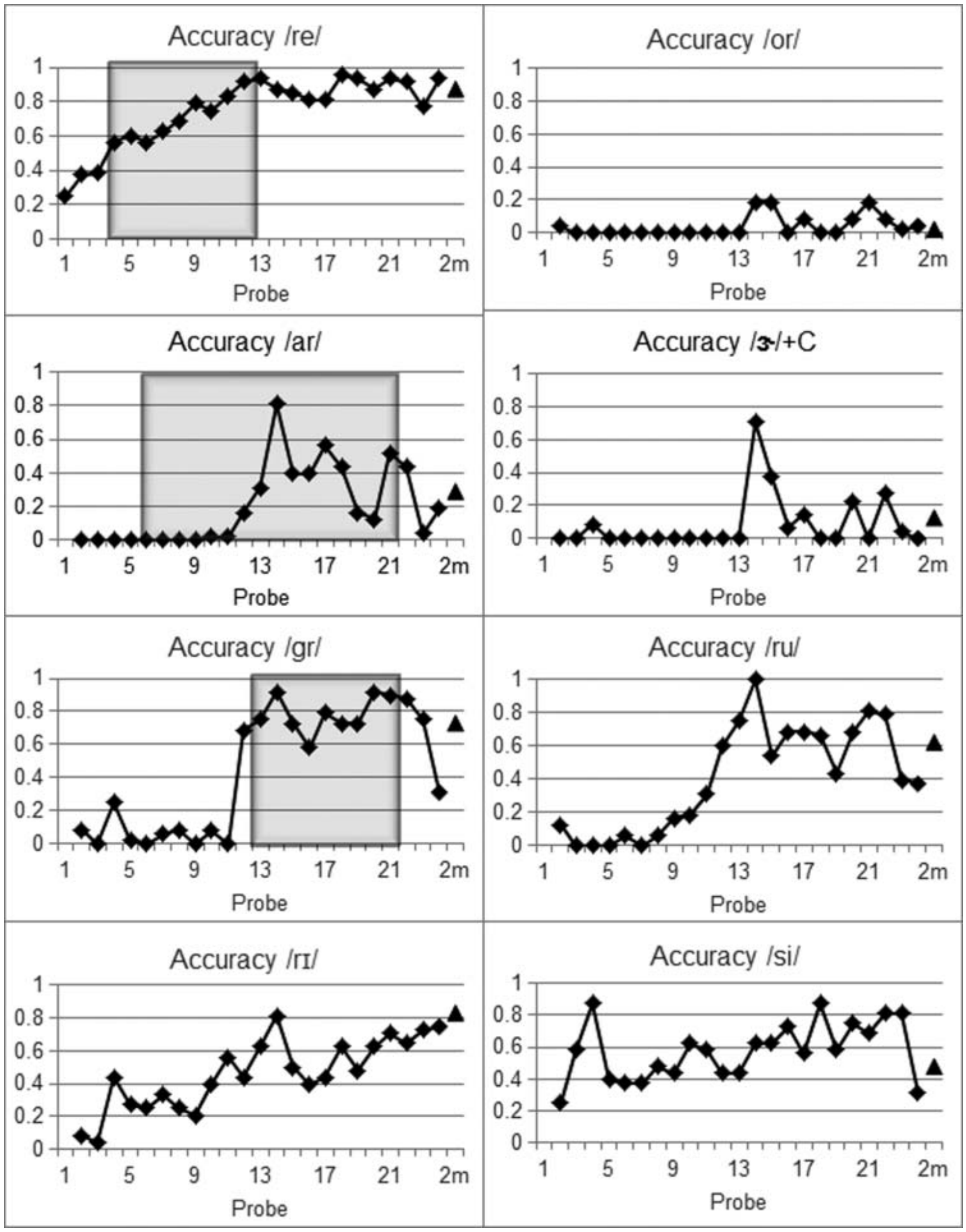
Graphical displays of probe data for the six participants are presented in Figures 1–6. The x-axis represents the probe number, and the y-axis represents the proportion correct of each target sound pattern on the probe (the average of two listeners). Shaded boxes represent sessions in which the sound pattern was treated. Triangles (the final data point on each graph) represent performance from the 2-month follow-up probe. Because some of the target sequences were phonetically similar (e.g., U002 was probed for both /re/ and /rɪ/), there was evidence of generalization of treatment to untreated target sequences. However, little improvement was seen in sequences that were phonetically unrelated to the treatment target sequences (e.g., for U007, improvement in /sk/ and /kl/ did not facilitate improvement in /rz/). Additionally, some sequences were in error during initial testing, but when individualized probes were developed and administered, a high performance was observed; thus, new sequences were probed partway through the study (e.g., U005's /skr/ probe) to replace sequences that were consistently high and would not be selected for treatment. Individual performances of each participant are summarized below.

U002

Based on U002's pretreatment assessment performance, rhotics and /s/ were in error in monosyllabic and multisyllabic words; thus, these were probed (Figure 1). The sequence /re/ was at 25%–38% accuracy during baseline, then after nine treatment sessions, U002 reached the criterion of 80% in two consecutive sessions. Target /ar/ was at 0% during baseline probes and was initially unresponsive to treatment. Some improvement in /ar/ was observed after several sessions. Performance remained above baseline levels during the last few treatment sessions, but the performance criterion was not met for /ar/. For /gr/, U002's baseline data were consistently below 25%, with one exception (69%, which may have been the result of generalization from treating /re/). Treatment for /gr/ was introduced once /re/ met criterion, and /gr/ met criterion after nine sessions. Posttreatment probes indicated that U002 retained high accuracy for both /gr/ and /re/.

Overall, U002 demonstrated improved accuracy in the treated prevocalic contexts /re/ and /gr/ but only limited

Figure 1. U002's performance on the probes for eight target sequences over the duration of treatment. Shaded boxes represent the sessions in which the target sequence was treated. Triangles represent the 2-month follow-up.



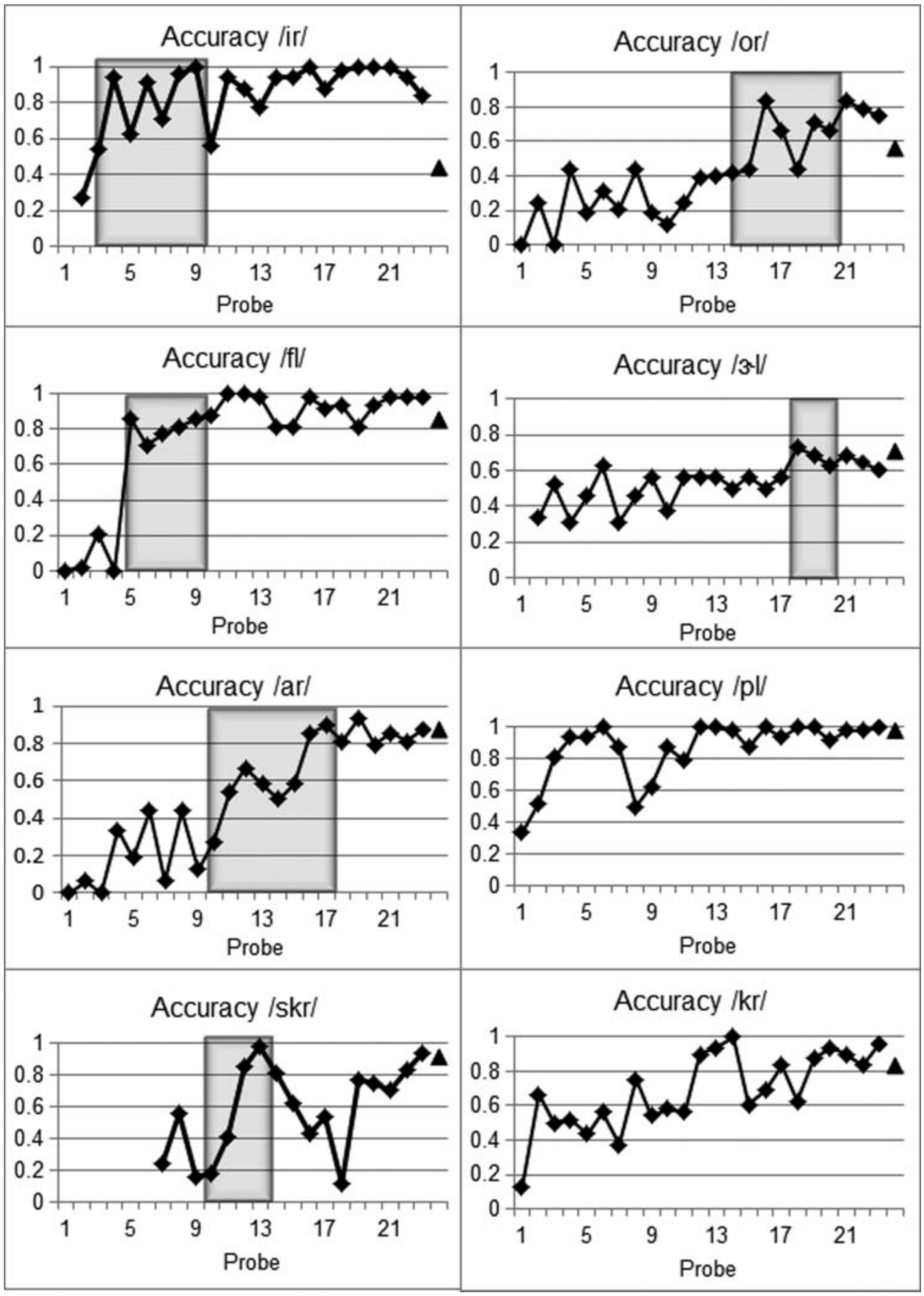
improvement in /ar/. Gains in untreated prevocalic rhotics /rɪ/ and /ru/ were also observed and may be the result of generalization from the treatment of other prevocalic rhotics (e.g., /re/, /gr/). No improvement was observed for untreated contexts /or/, /ɝ/+C (coronal consonant), and /si/.

U005

Based on U005's pretreatment assessment performance, liquid clusters and postvocalic rhotics were in error and so were probed (Figure 2). The first sound sequence to be treated was /ir/.[1] U005 reached criterion for /ir/ in seven treatment sessions, and a high level of accuracy was observed for nearly all of the posttreatment probes. Baseline data for /fl/ was below 21%, and immediate improvement was observed when treatment began. After five sessions, U005 met criterion for

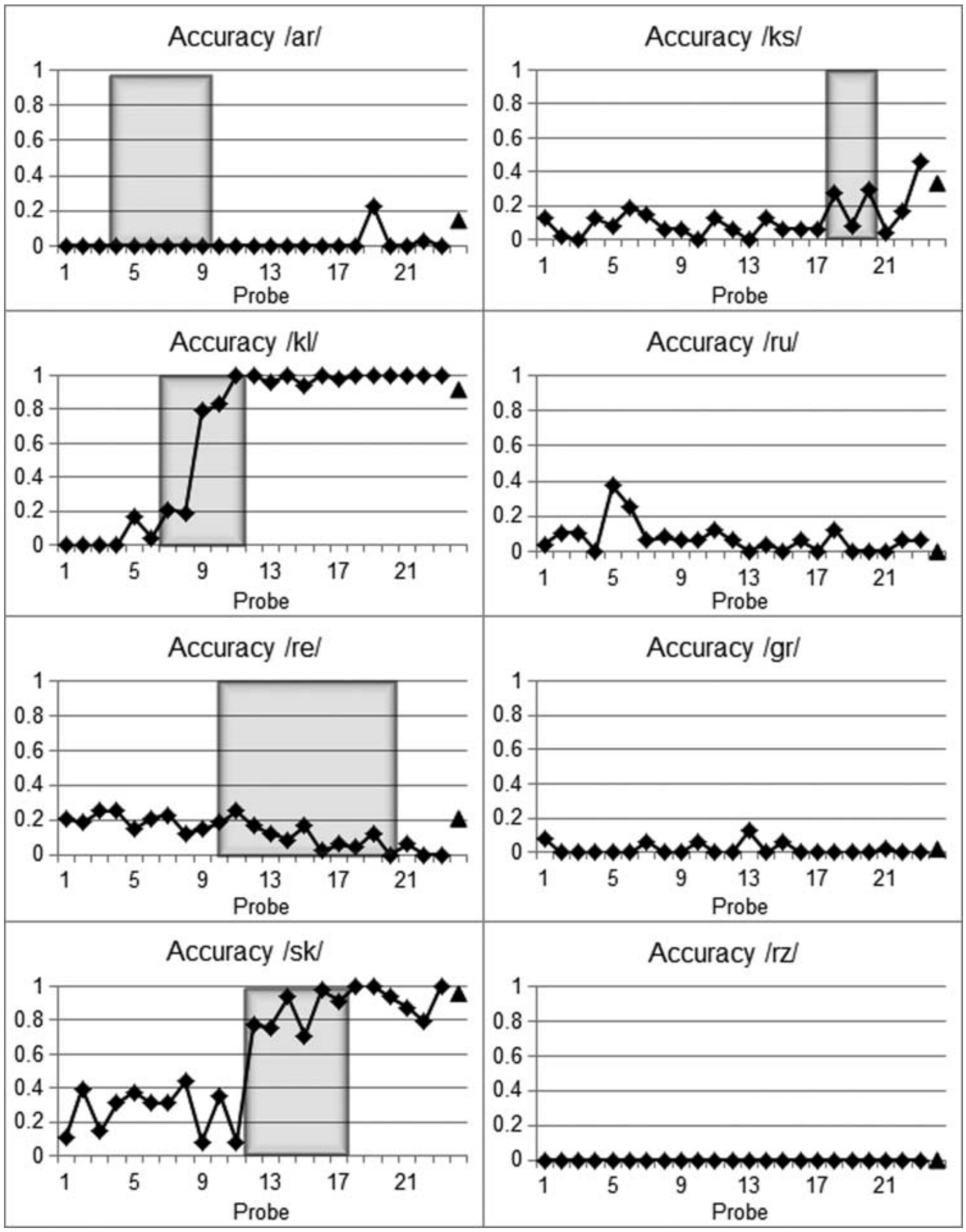
[1]Due to an error in the execution of the protocol, /ir/ was treated first instead of /ar/, and insufficient baseline data were collected. However, even without the data for /ir/, the within-subject design revealed some positive treatment effects for U005.

Figure 2. U005's performance on the probes for eight target sequences over the duration of treatment.



/fl/, and he continued to achieved 80% or higher for the posttreatment probes. Accuracy for /ar/ varied during U005's baseline probes, ranging from 0% to 43%. A gradual increase in accuracy was observed once treatment began, and criterion was reached after eight sessions; a high degree of accuracy was maintained during all posttreatment probes. For /skr/, baseline data fluctuated between 16% and 56%. During treatment of /skr/, U005 reached criterion after four sessions, and 70% accuracy or higher was observed during the final five posttreatment probes. Baseline data for /ɝl/ ranged from 31% to 62%, and U005's accuracy increased at the start of treatment on this sequence, but criterion was not met before the final treatment session. U005's posttreatment accuracy for /ɝl/ remained at ~60%. Finally, U005's baseline data for

Figure 3. U007's performance on the probes for eight target sequences over the duration of treatment.



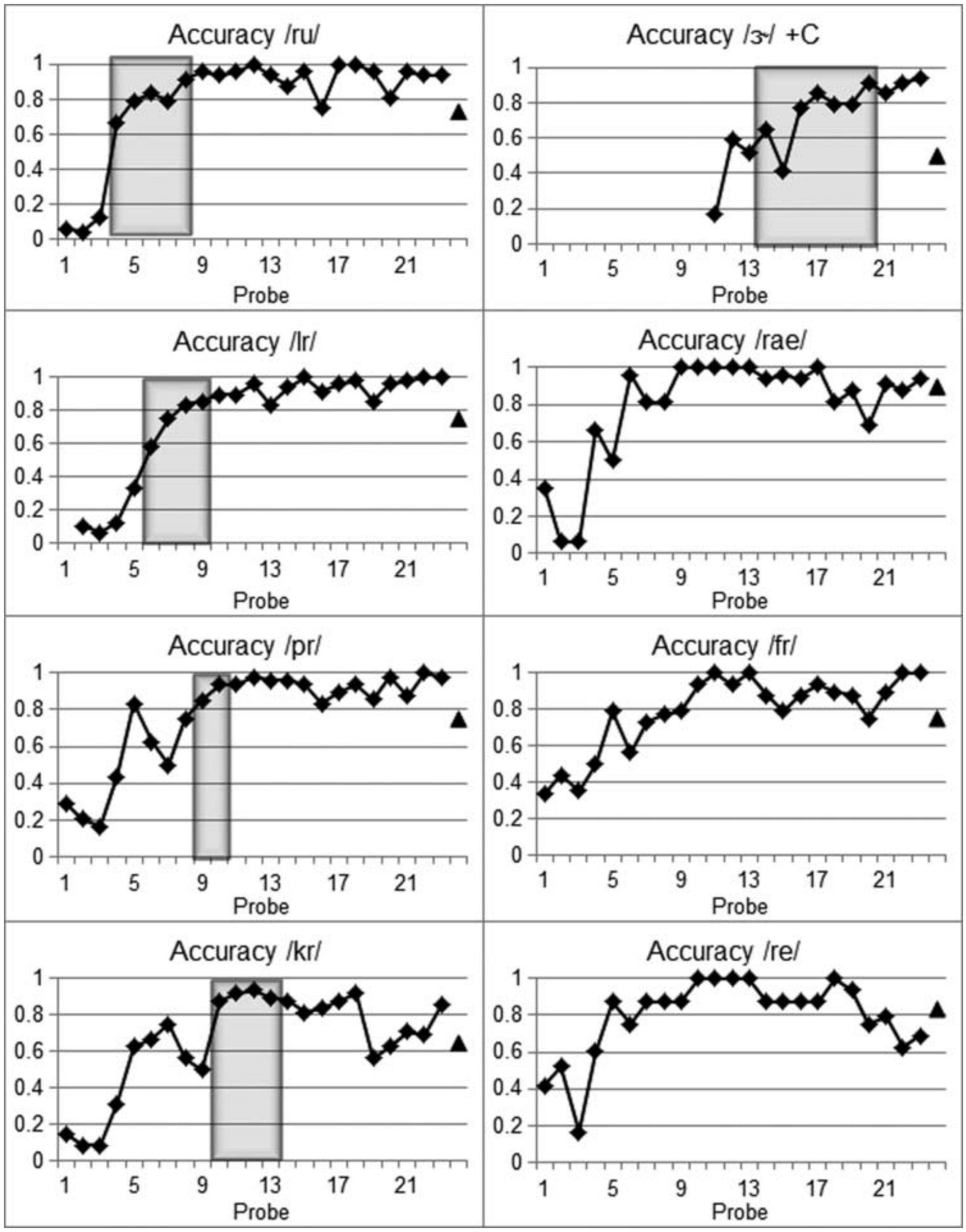
/or/ ranged from 0% to 44%. Treatment lasted for seven sessions, until the study ended. U005 achieved 80% on /or/ only once during treatment and therefore did not meet criterion. However, posttreatment probes for /or/ were greater than all pretreatment probes.

Overall, U005 achieved and maintained increased accuracy for the treated sequences /ir/, /fl/, and /ar/. Improvement was also seen during treatment of /skr/, but accuracy fluctuated during posttreatment probes. Improved accuracy was observed for treated sequences /ɝl/ and /or/, but U005 did not meet performance criterion despite an upward trend over the course of treatment. An increase in accuracy was seen for untreated sequences /pl/ and /kr/, which may be due to generalization effects from treatment of similar contexts. Two-month follow-up data revealed retention of posttreatment accuracy levels.

U007

Based on U007's pretreatment assessment performance, /s/ clusters, velar clusters, and rhotics were probed (Figure 3). He did not respond to treatment addressing rhotic sequences /ar/ or /re/. However, positive treatment effects were observed for two velar cluster targets. U007's baseline

Figure 4. U008's performance on the probes for eight target sequences over the duration of treatment.



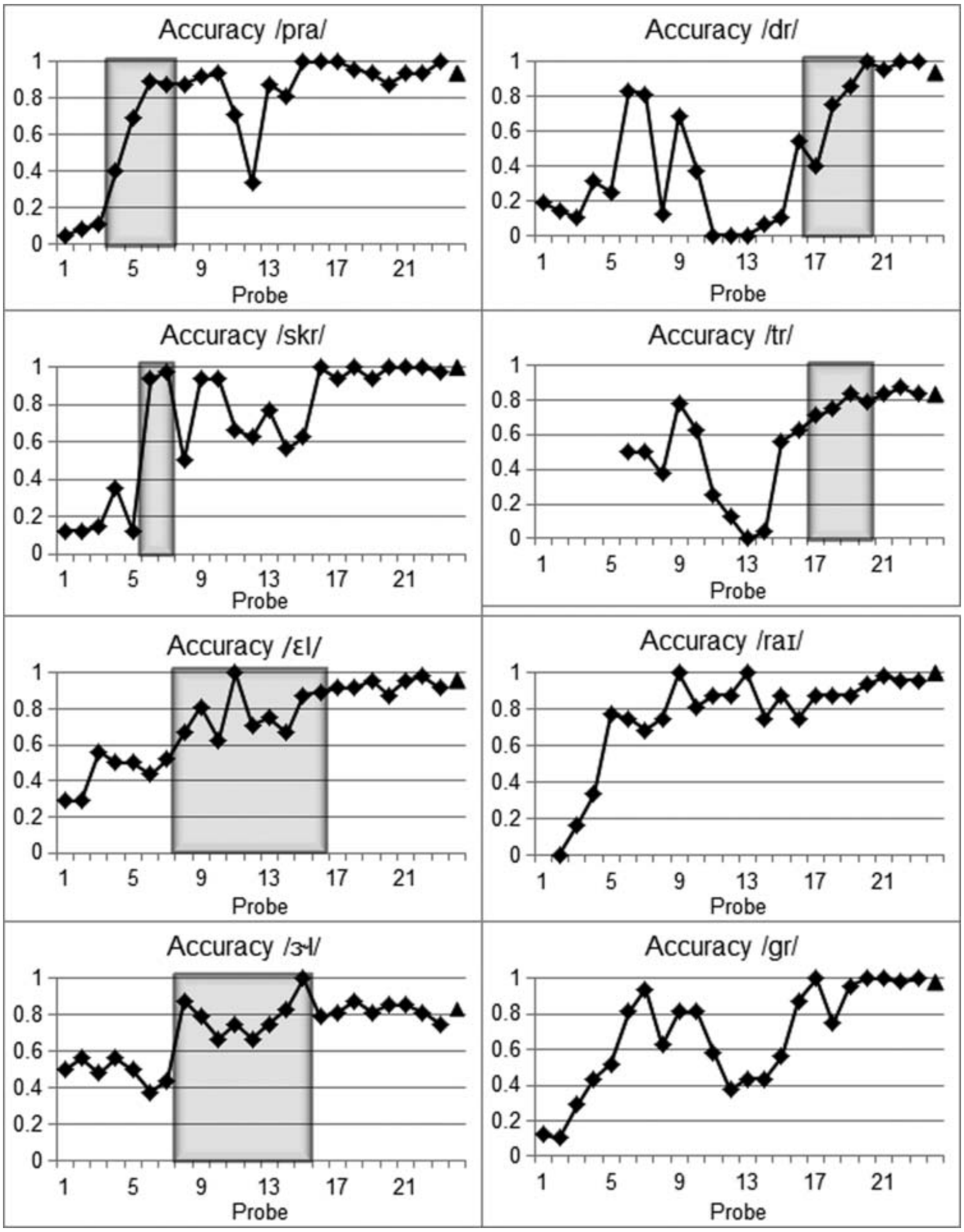
accuracy for /kl/ ranged from 0% to 21%, he reached criterion after five sessions, and he continued to demonstrate nearly 100% accuracy for all posttreatment probes. For /sk/, U007's baseline accuracy ranged from 10% to 44%. Upon introducing treatment for /sk/, U007's accuracy increased and he met the criterion after six sessions. Maintenance of treatment effects was demonstrated by posttreatment data at or above 80%. U007's baseline data for /ks/ fluctuated slightly from 0% to 19%. Treatment of /ks/ was conducted over the final three treatment sessions, with a slight increase in accuracy during and after treatment.

Of the five treated sequences, U007 improved his accuracy for two (i.e., /kl/ and /sk/), and the 2-month follow-up revealed retention of these. Minimal changes were observed for /ks/. No improvement was seen for the treated contexts /ar/ or /re/, nor for the untreated contexts /ru/, /gr/, and /rz/.

U008

Based on U008's pretreatment assessment performance, he made errors on rhotics. Because U008's parents

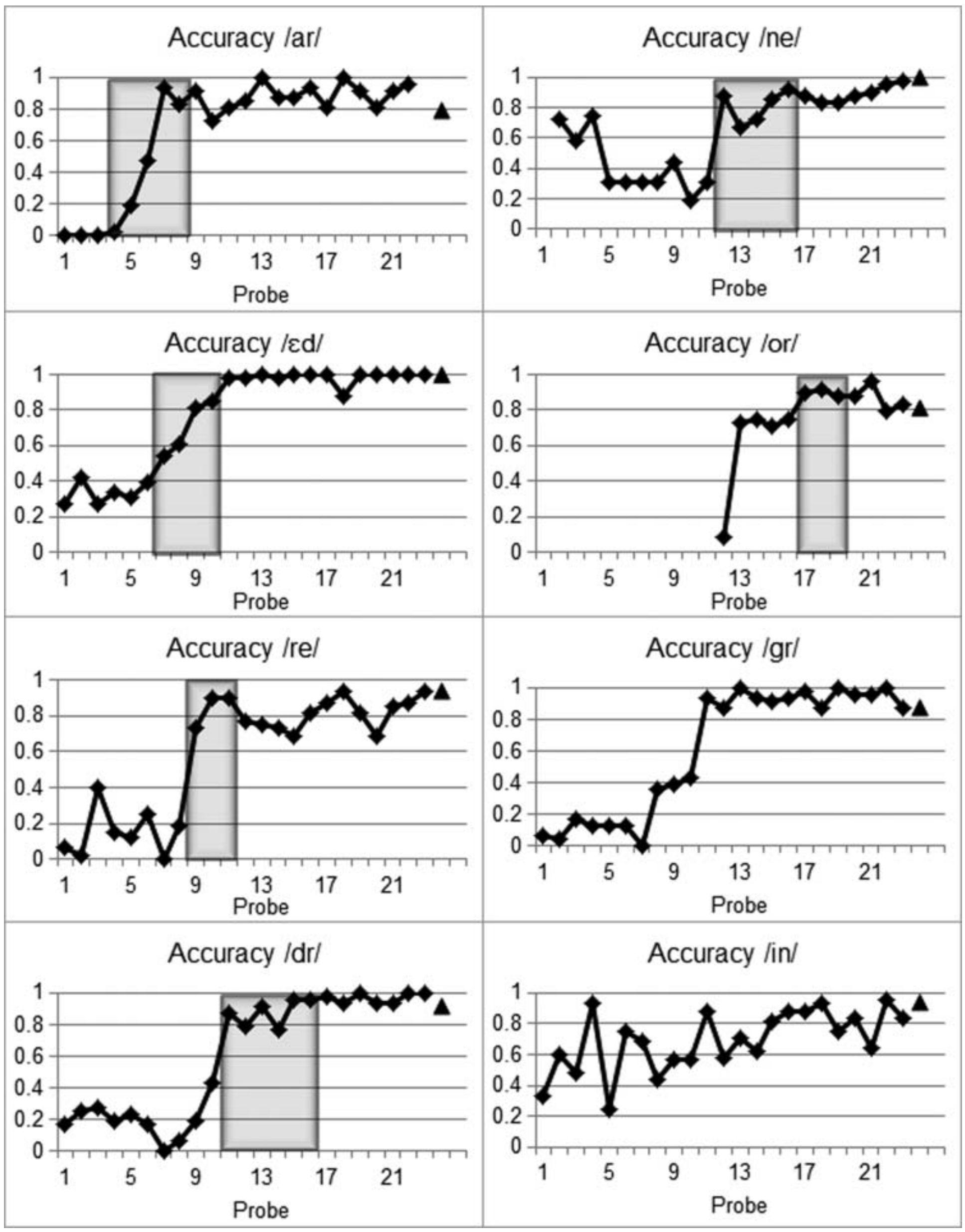
Figure 5. U009's performance on the probes for eight target sequences over the duration of treatment.



were from England, the vocalic /ɝ/ and postvocalic /r/ targets were not selected for treatment. As seen in Figure 4, treatment began with /ru/, and U008's baseline data for /ru/ varied between 4% and 13%. Once treatment began, an immediate increase in accuracy was seen, and criterion was met in five sessions. Posttreatment probes demonstrated accuracy ranging from 81% to 100%. For /lr/, baseline measurements ranged from 6% to 33%. Once treatment was introduced, criterion was met in four sessions, and U008 continued to achieve above 80% for /lr/ following the completion of direct treatment. U008's accuracy for /pr/ varied greatly across baseline probes, ranging from 17% to 83% (an increase in accuracy was observed likely due to generalization effects). Target /pr/ was treated for two sessions in which accuracy quickly reached criterion levels. Posttreatment probes revealed performance between 83% and 100%. Similarly, baseline data for /kr/ showed accuracy from 8% to 75% before the onset of treatment. Treatment on /kr/ resulted in immediate improvement, with a slight drop in accuracy after treatment was discontinued.

Because improvement was observed in most other target sequences, baseline data were collected for /ɝ/+C

Figure 6. U012's performance on the probes for eight target sequences over the duration of treatment.



partway through the study at the request of U008's mother. Although the family was from England and the parents spoke a nonrhotic dialect of English, U008's mother indicated that U008's /ɝ/ productions sounded neither like General American English nor like British English productions. U008's baseline probe accuracy ranged from 17% to 59% ("accurate" was defined as an American English rhotic production, as the clinicians were unable to reliably train the British English nonrhotic vowel). A trend toward higher accuracy was seen for /ɝ/+C, but the performance criterion was not met by the end of treatment. Posttreatment measurements indicated a continued increase in accuracy.

Overall, U008 improved accuracy on all five treated sequences (i.e. /ru/, /lr/, /pr/, /kr/, and /ɝ/+C). Additionally, accuracy for untreated sequences /re/, /ræ/, and /fr/ were at or below 52% for the first three probes, but all of these sequences were at or above 63% during the final three probes, suggesting generalization effects to untrained rhotics. The 2-month follow-up revealed a slight decrease in accuracy on several sound sequences, although performance was higher than most pretreatment data points.

Table 2. Comparisons of pre- and posttreatment accuracy and effect sizes for the treated sound sequences.

Participant	Treated sound pattern	Pretreatment % accuracy *M* (*SD*)	Posttreatment % accuracy *M* (*SD*)	Standardized mean difference	% nonoverlapping data points
U002	/re/	34 (7)	89 (6)	7.5	100
	/ar/[a]	0 (0)	25 (13)		100
	/gr/	12 (20)	75 (7)	3.1	83
	Total of all targets			3.2	
U005	/ir/	27 (0)	86 (18)		100
	/fl/	6 (10)	92 (7)	8.6	100
	/ar/	18 (31)	85 (5)	2.1	100
	/skr/	33 (21)	68 (24)	1.7	73
	/or/[a]	25 (15)	73 (12)	3.3	100
	/ɝl/[a]	49 (10)	66 (5)	1.7	75
	Total of all targets			2.6	
U007	/ar/[a]	0 (0)	2 (6)		20
	/kl/	7 (9)	99 (2)	10.6	100
	/re/[a]	20 (5)	2 (4)	–3.9	0
	/sk/	27 (13)	93 (9)	5.0	100
	/ks/[a]	8 (6)	22 (21)	2.6	50
	Total of all targets			4.0	
U008	/ru/	8 (4)	93 (7)	19.7	100
	/lr/	16 (12)	94 (5)	6.5	100
	/pr/	48 (25)	93 (5)	1.8	100
	/kr/	41 (26)	79 (11)	1.4	71
	/ɝ/+C[a]	43 (23)	89 (8)	2.1	100
	Total of all targets			2.1	
U009	/pra/	8 (3)	88 (16)	25.3	100
	/skr/	14 (2)	84 (18)	37.8	100
	/ɛl/	44 (11)	93 (4)	4.4	100
	/ɝl/	50 (8)	82 (4)	4.0	100
	/dr/	25 (29)	99 (2)	2.5	100
	/tr/[a]	40 (26)	85 (2)	1.7	100
	Total of all targets			2.2	
U012	/ar/	0 (0)	89 (8)		100
	/ɛd/	33 (6)	99 (3)	10.5	100
	/re/	15 (13)	81 (9)	5.0	100
	/dr/	20 (12)	97 (3)	6.7	100
	/ne/	42 (20)	89 (6)	2.2	100
	/or/	60 (29)	88 (6)	0.9	100
	Total of all targets			2.7	

Note. The standardized mean difference (SMD) of all targets represents the difference between the pretreatment data points and all of the posttreatment data points, divided by the standard deviation of all of the pretreatment data points. The SMD cannot be computed if there is no variance in baseline data; those cells are blank. The percentage of nonoverlapping data points is the percentage of posttreatment data points that are above the highest pretreatment data point.

[a]The target never met the performance criterion of 80% accuracy on two consecutive probes.

Table 3. Pre- and posttreatment percentage of phonemes correct based on the GFTA–2 and a 17-sentence imitation task, as determined by a listener who was blind to the intervention status.

	Percentage of phonemes correct	
Participant	Pretreatment	2 months posttreatment
U002	84.7	85.8
U005	87.3	91.8
U007	72.0	73.2
U008	93.7	95.2
U009	91.7	94.2
U012	92.3	95.6

U009

Based on U009's pretreatment assessment performance, he produced distortions of liquids /r, l/ and vowels such as /ɛ/, with many errors in multisyllabic productions. As shown in Figure 5, the first treated sequence was /pra/, and baseline data revealed accuracy at or below 10%. Treatment of /pra/ continued for five sessions until the performance criterion was met; after treatment, U009 demonstrated above 80% on nearly all probes. Accuracy for /skr/ was stable at ~15% on baseline probes. Once treatment of /skr/ began, U009's accuracy immediately increased, and U009 met criterion in only two sessions. Following treatment, he demonstrated some instability, but the final probes indicated a high degree of accuracy. Baseline data for /ɛl/ ranged from 29% to 56%. U009 required nine treatment sessions before performance criterion was reached, and posttreatment accuracy remained above 88%. Another treatment target, /ɝl/, varied between 38% and 56% during pretreatment probes. U009 reached criterion on /ɝl/ after eight sessions, and posttreatment probes showed a high degree of accuracy. For /dr/, baseline data varied greatly from 0% to 83%. Once treatment was initiated for /dr/, criterion was met in four sessions, and posttreatment probes indicated continued high levels of accuracy. Similarly, probes on /tr/ revealed instability, ranging from 0% to 78%. Treatment lasted for four sessions until the final treatment session. Although U009 exceeded 80% for two out of four treatment probes, our performance criteria required him to exceed 80% for two consecutive probes. Because he exceeded 80% on nonconsecutive probes, he did not meet criterion for /tr/. Following treatment, however, U009 demonstrated probe scores above 80% for /tr/.

In summary, U009 increased his accuracy for all six treated sequences (i.e., /pra/, /skr/, /ɝl/, /ɛl/, /dr/, and /tr/). Accuracy for /gr/ and /raɪ/ increased without direct treatment, presumably due to generalization from training rhotics in other contexts. U009 maintained above 75% accuracy over the final three probes for all treated and untreated sequences, as well as during the 2-month follow-up.

U012

Based on U012's pretreatment assessment performance, rhotics and tense front vowels /i, e/ were noticeably distorted, and alveolar sounds were inconsistently produced with elevation of the tongue dorsum and the tongue blade/tip (often resulting in a percept of a velar consonant). As shown in Figure 6, U012 demonstrated a stable baseline of 0% for /ar/ before the start of treatment. Once treatment was initiated, U012's accuracy steadily increased, and he reached the performance criterion after five sessions. He maintained a high degree of accuracy on /ar/ once treatment on this target was discontinued. Target /ɛd/ was introduced next, and U012's baseline data varied from 27% to 42%. U012 reached criterion for /ɛd/ after four treatment sessions, and posttreatment probes revealed accuracy from 87% to 100%. For the sequence /re/, baseline data ranged from 0% to 40%; an immediate increase in accuracy was observed once treatment began, and criterion was met after only three treatment sessions. Posttreatment probes revealed accuracy above pretreatment levels for /re/. Accuracy for /dr/ fluctuated from 0% to 44% during the baseline period. U012 reached the performance criterion for /dr/ in six sessions, and his accuracy continued to be above 93% during the final probes. Similarly, his baseline data for /ne/ were unstable, ranging from 19% to 75%. Treatment lasted five sessions before U012 met the performance criterion. Posttreatment probes showed sustained accuracy for /ne/. Additionally, baseline probes for /or/ were introduced partway through the study, with accuracy ranging from 8% to 75%. U012 met the criterion quickly once treatment was introduced, and his posttreatment probes indicate that he maintained a high degree of accuracy.

Overall, U012 increased and maintained accuracy for all six treated sequences (/ar/, /ɛd/, /re/, /dr/, /ne/, /or/). Untreated sequences /gr/, /tr/, and /in/ showed improvement without direct intervention. The 2-month follow-up revealed that U012 maintained high performance on all target sequences.

Discussion

Real-time ultrasound imaging was used to provide a visual display of tongue configuration and movement for six children with persisting speech sound errors associated with CAS. This is the first study to apply this biofeedback approach specifically to children with CAS. All participants reached the preestablished criterion (80% at the word level for two consecutive sessions) for at least two target sequences. Performance criterion was met for 23 of 31 target sequences included in treatment. The average time to reach this criterion was five sessions, and those treated sequences that reached criterion remained at a high degree of accuracy once treatment on that target was discontinued. The percentage of nonoverlapping data points suggests that the vast majority of posttreatment probes' data points were higher than the maximum pretreatment probe score for the treated sequences. Most of the gains observed during treatment were maintained 2 months later.

This study provides evidence that this treatment program, as implemented here, can facilitate improvement in speech sound accuracy. The results provide support for the notion that a treatment program for children with CAS that includes feedback about lingual movements may facilitate more accurate and stable productions of sound sequences, and that generalization to phonetically similar untreated words occurs (i.e., only half of the probe list included treated words). The hypothesized mechanism for improvement is enhanced motor control by linking the motor plan with explicit visual feedback of the movement. It is well established that auditory, motor, and somatosensory feedback loops are used during phonological development and in online adaptation of speech production (Tourville & Guenther, 2011); the visual feedback provided here may have helped to facilitate participants' speech learning by providing another source of feedback.

Although the present research design cannot be used to determine which specific aspects of the intervention were responsible for the improvement, or which mechanisms were truly responsible for the change, all of the participants had already received many years of traditional treatment, so it is likely that the biofeedback procedures facilitated the improvements. Future studies could explore systematic manipulation of factors such as the relative duration and frequency of visual and verbal feedback, level of prosodic manipulation, feedback response rate, stimulus set size, practice distribution, and so on (e.g., Maas et al., 2008). Direct comparisons to other biofeedback and nonbiofeedback treatment approaches would also be of clinical value.

Caveats and Limitations

The data from this study indicated that improvement was generally rapid (except for a few treated sequences);

however, there was evidence that some of the untreated targets did not show improvement (e.g., U002's /or/, U007's /ru, gr, rz/), which helps to validate the single-subject design. In some instances, though, the untreated target sequences did not necessarily show stable, low accuracy (which is preferred in a multiple baseline design); some rising baselines were observed, potentially due to generalization effects. Although this may be viewed as compromising the single-subject experiment design (because untreated sequences were not necessarily stable), these changes in untreated sequences represent positive clinical outcomes. That is, not all generalization effects could be predicted a priori. Moreover, these children were reportedly making little progress in their traditional treatment, and it is unlikely that spontaneous improvement could account for the growth observed in all cases (particularly due to their age and persistent problems in speech sound accuracy). Treatment targets were selected based on phonetic transcriptions, and priority was given to sequences based on perceived impact on intelligibility and clinical judgment; thus, no systematic algorithm was used to select the targets. Future studies could apply different approaches to target selection in order to evaluate how to maximally achieve generalization effects. For example, it appears as though treating prevocalic rhotics generalized to other prevocalic rhotics, but this was not necessarily the case for postvocalic rhotics.

Because speech movements are slowed down to take advantage of the visual biofeedback, the approach is most useful for targeting short sequences (such as CV, VC, CC contexts). Thus, the approach is primarily effective for eliciting and establishing correct productions in syllables and words. The probe data reported here, therefore, reflected accuracy at the word level. Although the probes sampled both treated and untreated words, including motorically complex words, the word-level probes are not necessarily indicative of sentence- and conversation-level skills. Future studies could explore the generalization of treated sequences to connected speech contexts.

Another potential limitation of this study is that blind scoring was not used on the probes. The decision was made to score accuracy live whenever possible in order to mimic true clinical data collection and because treatment decisions about continuing or discontinuing the intervention for a target sequence were dependent on the most recent performance. However, acceptable levels of interrater reliability were achieved, and data presented reflect the combined score of two listeners. An additional limitation is that the majority of the intervention (74% of sessions) was implemented by graduate students rather than a certified SLP. It is possible that greater treatment effects could have been observed with more experienced clinicians.

The present study included children with normal nonverbal cognitive skills and mostly mild or moderate cases of CAS. One participant with severe CAS and dysarthria (U007) showed improvement in two treated sequences (/kl, sk/) but overall was the poorest responder to the intervention (there was essentially no improvement in /ar/ or /re/). The biofeedback treatment approach, therefore, appears to facilitate improvements across a range of severity but might be more appropriate for mild-to-moderate cases of persisting CAS or for children who do not have co-occurring dysarthria. Future studies could explore systematic means of determining candidacy based on individual characteristics.

Finally, it is important to consider the relative drawbacks of the approach. For example, it is drill oriented and requires good sustained attention by the participants to focus on the feedback (which may have played a role in U005's unstable performance due to concomitant attention deficit hyperactivity disorder). Because images are only of the tongue, only lingual phonemes can be targeted using the feedback. Additionally, cost, access, and training with this technology may limit the clinical implementation of ultrasound biofeedback.

Summary

The six children with persisting speech sound errors associated with CAS participated in ultrasound biofeedback training, resulting in improved accuracy of sound sequences at the word level, and some participants showed generalization to untreated targets. Two-month follow-up data indicated that the participants generally maintained the gains they had achieved with this treatment. These results provide an extension of previous studies using ultrasound biofeedback to treat residual speech sound errors (Adler-Bock et al., 2007; Bernhardt et al., 2008; Modha et al., 2008), suggesting that ultrasound biofeedback may be a viable treatment option for improving the accuracy of sound sequences for children with persisting speech errors associated with CAS.

Acknowledgments

Support was provided by the Childhood Apraxia of Speech Association of North America, a Connecticut State University grant, and a donation to the Language and Early Assessment Research Network at Haskins Laboratories. Thanks to Virginia Porto, Shawna Oneil, and Nicole Augustine for providing treatment to participants; Emily Phillips, Cayla Dominello, Amanda Rizzo, and Gabrielle Cella for data entry; Dana Arthur for assisting with assessments; and Suzanne Boyce for providing initial training with ultrasound imaging.

References

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

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. Available from www.asha.org/policy

Bacsfalvi, P., & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics, 25,* 1034–1043.

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 53,* 1227–1245.

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 & Phonetics, 22*(2), 149–162.

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

Bernthal, J. E., Bankson, N. W., & Flipsen, P. (2008). *Articulation and phonological disorders* (6th ed.). San Antonio, TX: Pearson.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*(7), 1504–1529.

Campbell, T. F. (1999). Functional treatment outcomes in young children with motor speech disorders. In A. J. Caruso & E. A. Strand (Eds.), *Clinical management of motor speech disorders in children* (pp. 385–396). New York, NY: Thieme.

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

Chappell, G. E. (1973). Childhood verbal apraxia and its treatment. *Journal of Speech and Hearing Disorders, 38,* 362–368.

Chumpelik, D. (1984). The PROMPT system of therapy: Theoretical framework and applications for developmental apraxia of speech. *Seminars in Speech and Language, 5*(2), 139–156.

Dent, H., Gibbon, F., & Hardcastle, W. J. (1995). The application of electropalatography (EPG) to the remediation of speech disorders in school-aged children and young adults. *International Journal of Language and Communication Disorders, 30,* 264–277.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP).* San Antonio, TX: Pearson.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test*, *Fourth Edition*. San Antonio, TX: Pearson.

Edeal, D. M., & Gildersleeve-Neumann, C. E. (2011). The importance of production frequency in therapy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 20,* 95–110.

Gierut, J. A., & Morrisette, M. L. (2011). Effect size in clinical phonology. *Clinical Linguistics & Phonetics, 25*(11–12), 975–980.

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation*, *Second Edition*. San Antonio, TX: Pearson.

Hayden, D., & Square, P. (1994). Motor speech treatment hierarchy: A systems approach. *Clinics in Communication Disorders, 4*(3), 162–174.

Hayden, D., & Square, P. (1999). *Verbal Motor Production Assessment for Children*. San Antonio, TX: Harcourt Assessment.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407–422.

Huang, H., Wolf, S. L., & He, J. (2006). Recent developments in biofeedback for neuromotor rehabilitation. *Journal of Neuroengineering and Rehabilitation, 3,* 11.

Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders, 33,* 391–428.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Iyengar, S. K., & Taylor, H. G. (2004). School-age follow-up of children with childhood apraxia of speech. *Language, Speech, and Hearing Services in Schools, 35,* 122–140.

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.

McAllister Byun, T. M., & 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,* 207–221.

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009). Effectiveness of an integrated phonological awareness approach for children with childhood apraxia of speech (CAS). *Child Language Teaching and Therapy, 25*(3), 341–366.

Modha, G., Bernhardt, B., Church, R., & Bacsfalvi, P. (2008). Ultrasound in treatment of /r/: A case study. *International Journal of Language & Communication Disorders, 43*(3), 323–329.

Moriarty, B., & Gillon, G. (2006). Phonological awareness intervention for children with childhood apraxia of speech. *International Journal of Language & Communication Disorders, 41*(6), 713–734.

Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology, 25*(2/3), 313–324.

Preston, J. L. (2008). *Phonological processing and speech production in preschoolers with speech sound disorders* (Unpublished doctoral disseration). Syracuse University, Syracuse, NY.

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

Secord, W. A., Boyce, S. E., Donohue, J. S., Fox, R. A., & Shine, R. E. (2007). *Eliciting sounds: Techniques and strategies for clinicians* (2nd ed.). Clifton Park, NY: Thomson Delmar Learning.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals*, *Fourth Edition*. San Antonio, TX: Harcourt Assessment.

Shriberg, L. D. (1975). A response evocation program for /er/. *Journal of Speech and Hearing Disorders, 40*(1), 92–105.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research, 40,* 313–337.

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.

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.

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders, 39*(5), 331–349.

Strand, E., & Debertine, P. (2000). The efficacy of integral stimulation intervention with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 8*(4), 295–300.

Strand, E., & Skinder, A. (1999). *Treatment of developmental apraxia of speech: Integral stimulation methods. Clinical management of motor speech disorders in children* (pp. 109–148). New York, NY: Thieme.

Strand, E., Stoeckel, R., & Baas, B. (2006). Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology, 14*(4), 297–307.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 52,* 1595–1609.

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

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Harcourt Assessment.

Williams, K. T. (2007). *Expressive Vocabulary Test*, *Second Edition*. San Antonio, TX: Pearson.

Young, E. C. (1987). The effects of treatment on consonant cluster and weak syllable reduction processes in misarticulating children. *Language, Speech, and Hearing Services in Schools, 18,* 23–33.

Appendix

Probe Example

Participant #: U007 Probe #9

For participant U007 in this session, /ar/ and /kl/ were treated (treatment addressed only four of the words from the /ar/ list and four from the /kl/ list). The /ar/ and /kl/ probe lists were therefore elicited three times at the end of the session. The probe lists for two untreated sequences, /ks/ and /ru/, were also selected to be elicited three times to obtain a more reliable estimate on this day (during the next session a different pair of untreated targets would be elicit three times). Probe lists for the remaining targets /re, gr, sk, rz/ were elicited once. Note that scoring represents the ratings of a single listener, but a second listener also scored the responses and the final data plotted are the average of the two listeners.

Target: /ar/	1st	2nd	3rd	Target: /ks/	1st	2nd	3rd
1. Car	–	–	–	1. Box	–	–	–
2. Jar	–	–	–	2. Socks	–	–	–
3. Star	–	–	–	3. Rocks	+	–	–
4. Scar	–	–	–	4. Picks	–	–	–
5. Guitar	–	–	–	5. Fox	–	–	–
6. Sonar	–	–	–	6. Beeswax	–	–	–
7. Boxcar	–	–	–	7. Complex	–	–	–
8. Candy bar	–	–	–	8. Chicken pox	–	–	–
Target: /kl/	**1st**	**2nd**	**3rd**	**Target: /ru/**	**1st**	**2nd**	**3rd**
1. Clean	+	+	+	1. Rule	–	+	–
2. Clock	+	+	–	2. Rude	–	–	–
3. Claw	+	+	+	3. Room	–	–	–
4. Clam	+	+	+	4. Roof	–	–	–
5. Closing	+	–	–	5. Ruby	–	–	–
6. Closet	+	+	+	6. Rootbeer	–	–	–
7. Clicking	+	+	+	7. Rudolph	–	–	–
8. Clamoring	+	–	–	8. Ruin	–	–	–
Target: /re/	**1st**	**2nd**	**3rd**	**Target: /gr/**	**1st**	**2nd**	**3rd**
1. Rain	–			1. Grab	–		
2. Rake	+			2. Grape	–		
3. Race	–			3. Green	–		
4. Range	–			4. Grow	–		
5. Raisin	–			5. Gravy	–		
6. Railway	–			6. Grouchy	–		
7. Radio	–			7. Grasshopper	–		
8. Racecar	–			8. Grizzly bear	–		
Target: /sk/	**1st**	**2nd**	**3rd**	**Target: /rz/**	**1st**	**2nd**	**3rd**
1. Skate	–			1. Doors	–		
2. Scoop	–			2. Chores	–		
3. Scam	+			3. Floors	–		
4. Skin	–			4. Cores	–		
5. Skunk	–			5. Fours	–		
6. Skimming	–			6. Alters	–		
7. Scary	–			7. Explores	–		
8. Skeleton	–			8. Achievers	–		