# A Treatment for Dysprosody in Childhood Apraxia of Speech

**Article** *in* Journal of Speech Language and Hearing Research · October 2010

DOI: 10.1044/1092-4388(2010/09-0130) · Source: PubMed



**4 authors:**

Kirrie Jane Ballard
The University of Sydney


Donald A Robin
University of New Hampshire


Patricia Joan McCabe
The University of Sydney


Jeannie Mcdonald
Macquarie University

# A Treatment for Dysprosody in Childhood Apraxia of Speech

**Kirrie J. Ballard**
University of Sydney,
New South Wales, Australia

**Donald A. Robin**
The University of Texas Health Sciences
Center at San Antonio and
The University of Texas
at San Antonio

**Patricia McCabe**
**Jeannie McDonald**
University of Sydney

**Purpose:** Dysprosody is considered a core feature of childhood apraxia of speech (CAS), especially impaired production of lexical stress. Few studies have tested the effects of intervention for dysprosody. This Phase II study with 3 children investigated the efficacy of a treatment targeting improved control of relative syllable durations in 3-syllable nonwords representing strong-weak (SW) and weak-strong (WS) stress patterns (e.g., BAtigu or baTIgu). Treatment sessions were structured along the principles of motor learning (PML) approach.
**Method:** Three children, age 7 to 10 years, with mild to moderate CAS and normal language development participated in an intensive 3-week treatment. Within-participant designs with multiple baselines across participants and behaviors were used to examine acquisition, generalization, and maintenance of skill.
**Results:** All children improved in their ability to control relative duration of syllables in SW and WS nonwords. Improvement was also noted in control of loudness and pitch contrasts. Treatment effects generalized to untreated nonword stimuli, but minimal change was seen in production of real words.
**Conclusion:** Findings support the efficacy of this approach for improving production of lexical stress contrasts. Structuring the intervention according to the PML approach likely stimulated strong maintenance and generalization effects.

**KEY WORDS: childhood apraxia of speech, prosody, lexical stress, treatment, motor learning**

There are few studies on the efficacy of treatment for childhood apraxia of speech (CAS). Historically, clinicians have labeled these children as difficult to treat because of limited or slow progress (Forrest, 2003; Hall, Jordan, & Robin, 2007) and a lack of research evidence to guide treatment approaches (Morgan & Vogel, 2008). Nevertheless, the symptoms of CAS typically persist throughout life. Thus, the development of innovative approaches to speech treatment in CAS is critical. To that end, we designed a novel treatment approach that targets the perceived dysprosody that is a predominant feature of CAS (American Speech-Language-Hearing Association [ASHA], 2007).

Though there has been debate over the diagnostic markers of CAS, ASHA (2007) published a position statement to serve as a unifying guide for research on this population. Three diagnostic characteristics were identified in the statement as having wide support, reflecting impaired programming of the spatial and temporal parameters of speech movements. These are "(a) inconsistent errors on consonants and vowels in repeated productions of syllables or words, (b) lengthened and disrupted coarticulatory transitions between sounds and syllables, and (c) *inappropriate prosody, especially in the realization of lexical or phrasal stress* [emphasis added]" (ASHA, 2007, p. 2). The latter two features underlie the common perception that individuals with CAS segment their speech, as if talking syllable by syllable (Robin, Maas, Sanberg, & Schmidt, 2007). Interestingly, these features are also central to the adult form of apraxia of speech (Duffy, 2005; McNeil, Robin, & Schmidt, 1997).

Previous studies that have examined the suprasegmental disruptions in CAS have found abnormal production of lexical, or metrical, stress as measured by perceptual judgments and/or acoustic variables (Munson, Bjorum, & Windsor, 2003; Nijland et al., 2003; Shriberg, Aram, & Kwiatkowski, 1997; Skinder, Strand, & Mignerey, 1999; Velleman & Shriberg, 1999). Lexical stress is phonetically realized through manipulation of three variables: duration of the vowel (ms), vocal intensity (dB), and fundamental frequency ($F_0$ in Hz; Kager, 2007). In English, stressed syllables have longer vowel durations and higher peak intensity and peak $F_0$. Multisyllabic words in English tend to have an alternating stressed-destressed pattern across syllables. Nouns typically follow a trochaic pattern of strong-weak (SW) stress, while verbs more typically have an iambic pattern of weak-strong (WS). Children as young as 12 months of age have the capacity to manipulate these phonetic features in their prelinguistic productions (Davis, MacNeilage, Matyear, & Powell, 2000). Pollock, Brammer, and Hageman (1993) analyzed SW and WS CVCV nonword productions in children age 2 to 4 years. They reported that the 2-year-olds were able to signal lexical stress accurately over CVCV nonwords through relative vowel duration. The duration contrast became more pronounced between 2 and 3 years, primarily through shortening of the unstressed syllable. Although 2-year-olds were not signaling stress through pitch and intensity, by 3 years of age children were able to do so. There was little change observed in relative duration, pitch, and intensity measures from 3 to 4 years of age.

The speech motor impairment of CAS appears to interfere with development of the fine rapid control of articulatory muscles that is required for expression of subtle lexical stress contrasts across syllables (Shriberg et al., 1997; Skinder et al., 1999). In one study, 52% of 53 children with suspected CAS were perceived to use excessive, equal, or misplaced stress in connected speech samples, which was a much higher incidence than the 10% level for a broader population of children with speech delay (Shriberg et al., 1997). While Velleman and Shriberg (1999) found that children with CAS were perceived to produce the same types of lexical stress errors as children with non-CAS speech delay, they noted that the stress errors persisted in children with CAS as late as 14 years. Nijland et al. (2003) performed acoustic measures of word and segment durations in children with CAS. They found significantly longer durations in CAS than normally developing peers, and, unlike peers, the CAS children did not shorten vowel duration in weaker stressed initial syllables. The findings of Nijland et al. are consistent with those of Shriberg and colleagues, but they did not comment about whether their acoustic measures correlated well with perceptual judgments of lexical stress. Munson et al. (2003) performed both acoustic and perceptual measures of CVCV nonword productions from five CAS and five phonologically delayed (PD) children between the ages of 3;9 (years;months) and 8;11. They reported that all children were able to mark WS nonwords with durational contrast and SW nonwords with pitch and intensity contrasts. However, the productions of the CAS children were perceived by listeners to be less accurate in stress placement than the PD children. Munson et al. did not comment on how acoustic measures may have correlated with degree of perceptual accuracy.

In summary, perceptual and acoustic measurement studies have supported the hypothesis that CAS is a speech motor disorder that, in part, affects control of temporal parameters of speech movements that underlie production of prosodic features at the syllable level. However, the relationship between acoustic and perceptual measures has not been fully explored. This is an interesting question, given that perceptual judgments are often considered the gold standard in assessment and treatment planning for speech disorders (Duffy, 2005).

CAS intervention studies have focused primarily on improving segmental errors, with few targeting impairments of coarticulation or prosody. For example, interventions have focused on accuracy of phonemes (Williams & Stephens, 2004), often taking a linguistic approach based on training phonological processes (Powell, 1996) or phonological awareness (Moriarty & Gillon, 2006). Other studies have included children with severe speech impairment and very young children and so have focused on developing intelligible production of core vocabularies (Strand, Stoeckel, & Bass, 2006) or augmentative and alternative forms of communication (Cumley & Swanson, 1999). However, it is important to note that Strand and colleagues (2006) advocate for varying prosody during treatment for articulatory accuracy of words. In their Dynamic Temporal and Tactile Cueing treatment approach, they routinely have children produce words with randomly varying intonational patterns, as modeled by the clinician. This team has not specifically measured or reported changes in prosody in the children treated.

Given the agreement among clinicians and researchers that CAS is an impairment of speech motor control and that dysprosody is a core symptom of CAS (ASHA, 2007), it is logical to develop treatments to improve speech motor skills underlying production of prosodic contrasts. Though treatments for dysprosody exist (Hall et al., 2007; for a review, see Hargrove, Anderson, & Jones, 2010), no studies have experimentally tested intervention for prosody in CAS (Morgan & Vogel, 2008).

Furthermore, few interventions for CAS or production of prosodic contrasts have been cast in a motor learning framework. The principles of motor learning (PML) approach provides a strong starting point (Schmidt & Lee, 2005). The PML approach was developed within the Schema Theory of Motor Control and Learning (Schmidt, 1975). Over the past 30 years, several hundred published studies have led to the development of a set of simple principles that facilitate *maintenance* and *generalization* of trained motor skills in children and adults (Maas et al., 2008; Schmidt & Lee, 2005). These studies have largely considered limb movements in healthy individuals, but an increasing number have shown similar effects in adults with limb and speech motor impairments (see Maas et al., 2008, for a review).

PML guide the structure and frequency of *practice* and the provision of augmented feedback. Existing therapy programs for CAS can be easily adapted to a PML structure without losing their core elements such as the tactile and kinesthetic cues of the Prompts for Restructuring Oral-Muscular Phonetic Targets (PROMPT) method.

Principles of practice structure that facilitate long-term learning of motor skills include (a) high intensity of practice, (b) training multiple and varied skills in parallel with (c) random ordering of stimuli within session, and (d) initiating training at high levels of task or stimulus complexity. In relation to the complexity principle, it has been shown that training a motor skill of given complexity (e.g., *st* consonant blends) results in generalization down the hierarchy (i.e., to *s* and *t*) but not up the hierarchy (i.e., *str* blends; Maas et al., 2008). Principles of feedback frequency and structure that facilitate long-term motor learning include (a) restricting feedback to information on simple accuracy (i.e., knowledge of results, or *KR feedback*) rather than detailed performance characteristics (i.e., knowledge of performance, or *KP feedback*), (b) providing feedback on only 50% of responses, and (c) inserting a 3–5-s silent pause between a response and the provision of feedback. Generally, the principles serve to increase the difficulty of the practice task and encourage self-evaluation of responses. These principles show high overlap with the recently developed principles of experience-dependent neural plasticity (Kleim & Jones, 2008; Ludlow et al., 2008). Several treatment efficacy studies have supported use of PML with acquired apraxia of speech (Austermann Hula, Robin, Maas, Ballard, & Schmidt, 2008; Ballard, Maas, & Robin, 2007; Knock, Ballard, Robin, & Schmidt, 2000; Maas, Barlow, Robin, & Shapiro, 2002) and CAS (Strand et al., 2006).

The current study is a Phase II trial testing the efficacy of a new treatment approach in three cases of CAS, using within-subject experimental designs. The treatment targets rapid and fluent production of lexical stress contrasts in multisyllabic strings. It is assumed that impaired lexical stress production in CAS stems from a deficit in rapid and fluent control of temporal and spatial parameters of articulator movement required to produce the fine variations in duration, vocal intensity, and $F_0$ across syllables. The approach uses multisyllabic strings to provide intensive practice in transitioning between segments and syllables with varying stress patterns. Nonwords are selected because they (a) allow random sequencing of syllables to increase the variety of transitions practiced and (b) simulate encountering novel words and rapidly planning movements and movement sequences without the influence of learned motor plans or linguistic representations. In this instance, the level selected for treatment was three-syllable varied nonsense strings; at the commencement of the treatment, the three participants were able to produce these strings with high segmental accuracy. Therefore, the effects of the treatment on production of lexical stress could be examined independently of the influence of segmental errors. However, the children produced the strings with relatively equal duration, intensity, and pitch levels across the first and second syllables in both SW and WS forms, giving rise to the perception of syllable segregation.

Our treatment approach was guided by the PML approach in the choice of complex and varied stimuli, high-intensity practice, random order of stimulus presentation during practice, and low frequency of KR feedback. Stimulus complexity was defined by number of syllables and different phonemes in a string (Schneider & Frens, 2005). High-complexity stimuli were novel four-syllable strings with three different consonants and vowels (e.g., butagitu), midcomplexity stimuli were three-syllable varied strings (e.g., butagi), and low-complexity stimuli were three-syllable strings with varied consonants (e.g., bataga). As a central problem of CAS is planning varied sequences of syllables for fluent production, it was predicted that manipulating prosody in novel and varied three-syllable sequences would be challenging. The specific hypotheses were as follows:

1. Treatment, structured in accordance with PML, will result in improved ability to produce two contrasting stress patterns (i.e., SW and WS) in three-syllable varied strings (e.g., BAtigu and baTIgu). That is, the children would move from a largely unimodal distribution (e.g., equal stress on both SW and WS items) to a bimodal distribution (i.e., SW and WS stress) for relative duration, peak intensity, and peak $F_0$ across adjacent syllables in a string.
2. Treatment effects will generalize to less complex three-syllable strings (e.g., BAtaga) but not to more complex four-syllable strings (e.g., BAtiguta).

3. Treatment effects may generalize to untreated SW and WS three-syllable real words, as they contain the same number of syllables, and existing lexical representations and semantic linkages may facilitate production. However, generalization may not occur because the selected stimuli (see Appendix) contain on average 7.2 phonemes, including consonant clusters, compared with six singleton phonemes in the treated stimuli.
4. Treatment and generalization effects will be retained at 4 weeks posttreatment, with performance better than baseline levels.
5. Acoustic measures of relative duration, intensity, and/or pitch will correlate highly with perceptual judgments of lexical stress, demonstrating that these variables accurately capture the perceived lexical stress errors and treatment-related changes of CAS speakers.

# Method

## Participants

Three children (M1: male, age 10;10; F1: female, age 9;2; and M2: male, age 7;8) from a single family were referred by their parents, in response to a recruitment advertisement for a broader CAS treatment study. M1 and F1 were left-handed, although there was no family history of left-handedness. All had a history of severe speech sound disorder (F1 > M1 > M2) in the context of normal receptive and expressive language skills, with no documented hearing or visual impairment. Currently, clinical diagnosis of CAS is based solely on perceptual judgment; there is no standardized test available. As such, diagnosis of CAS was based on presence of the core perceptual features of CAS (ASHA, 2007) during the speech tasks of the Motor Speech Examination (i.e., diadochokinesis, production of mono- and multisyllabic words, and connected speech; Duffy, 2005); the Goldman Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000); the Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996), for M1 and F1 only; and the Inconsistency Assessment from the Diagnostic Evaluation of Articulation and Phonology (Dodd, Hua, Crosbie, Holm, & Ozanne, 2002), for M2 only.

There was unanimous agreement on CAS diagnosis between the first three authors, all experienced speech-language pathologists. No orofacial structural abnormalities, muscle weakness, or altered muscle tone and reflexes were identified; this assessment ruled out a frank dysarthria (Motor Speech Examination; Duffy, 2005). None of the children had been identified with, or received services for, language reading or intellectual impairment. The children were reading above age level, as reported by their mother, who was a qualified elementary school reading teacher. Assessment of language skills using the Australian version of the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF–4; Semel, Wiig, & Secord, 2006) confirmed language skills within normal limits or higher for all children. No formal cognitive assessments were administered. However, CELF–4 scores correlate moderately to highly with measures of intelligence in children (Pearson Education, 2008). Results of language and speech testing are presented in Tables 1 and 2.

All three children had previously received traditional articulation therapy, working on one sound at a time while moving from sounds in isolation and simple words up through connected speech. Individual therapy was provided one to two times per week, and the family strictly adhered to a daily home practice regimen throughout. M1 attended therapy from age 3;2 to 6;9, working sequentially on velar plosives, fricatives, affricates, and liquids. F1 attended therapy from age 2;10 to 7;0 and worked on plosives, fricatives, affricates, and liquids. M2 attended therapy from 3;9 to 5;6, working sequentially on fricatives, affricates, and liquids. In no case was there direct treatment of prosody. Treatment was terminated when speech sound accuracy was at an age-appropriate level. Of note, all children also received occupational therapy for fine motor skills over the same period as speech therapy, and M2 was still receiving occupational therapy at the time of this study. Physical therapy had not been advised. The children had not received speech therapy in the 2 years prior to the study. Upon enrollment, they produced only a few segmental errors that were all sound distortions, and their speech intelligibility was 100%. However, their parents reported that family and friends described their speech as robotic, with M2 most affected. Following the position statement on CAS (ASHA, 2007), lexical stress was the focus here, as one aspect of speech naturalness.

## Stimuli

Four sets of stimuli were created (see Appendix) and used with all the children. Set 1 included the to-be-treated three-syllable nonsense strings containing three different plosive consonants and three different long vowels. There were 36 possible CVCVCV combinations (e.g., batigu, butiga), and these were duplicated to make a set of 72 with two different stress conditions: 36 with an SW stress pattern over the first two syllables and 36 with a WS pattern. Twenty-five of the syllable strings with SW and 25 with WS were randomly selected for treatment. Set 2 contained the remaining 22 strings that were left untreated to test for generalization of treatment effects to untreated exemplars of the treated behavior. Set 3 included an additional 20 syllable strings

**Table 1.** Results of pretreatment speech and language testing for the three participants (M1, F1, and M2).

| Test | M1 | | | F1 | | | M2 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Standard score | Percentile | Results | Standard score | Percentile | Results | Standard score | Percentile | Results |
| CELF–4 Receptive Language subtests | | | | | | | | | |
| Word Classes | 11 | 63 | WNL | 13 | 84 | WNL | 13 | 84 | WNL |
| Concepts and Following Directions | 11 | 63 | WNL | 12 | 75 | WNL | 7 | 16 | WNL |
| Sentence Structure | — | — | — | — | — | — | 10 | 50 | WNL |
| Receptive Language score | 106 | 66 | WNL | 115 | 84 | WNL | 98 | 45 | WNL |
| CELF–4 Expressive Language subtests | | | | | | | | | |
| Formulated Sentences | 14 | 91 | > NL | 15 | 95 | > NL | 11 | 63 | WNL |
| Word Classes | 15 | 95 | > NL | 9 | 37 | WNL | — | — | — |
| Word Structure | — | — | — | — | — | — | 14 | 91 | > NL |
| Recalling Sentences | 14 | 91 | > NL | 15 | 95 | > NL | 17 | 99 | > NL |
| Expressive Language score | 126 | 96 | > NL | 118 | 88 | > NL | 124 | 95 | > NL |
| | Raw score | | | Raw score | | | Raw score | | |
| Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996) | 31/40 | | < NL | 26/40 | | < NL | — | | — |
| Diagnostic Evaluation of Articulation and Phonology: Inconsistency Assessment (Dodd et al., 2002) | — | | — | — | | — | 10/25 | | Mild |
| | Errors | | | Errors | | | Errors | | |
| Goldman Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000) | Mild distortion of vowels and /r/; voicing of /f/; schwa insertion in some /l/ blends | | | Mild distortion of vowels, /s/, and /r/; devoicing of medial /v/; schwa insertion in some /l/ blends | | | Distortion of vowels, /r/, and /w/; devoicing of [th]; audible nasal air emission on final /m/ | | |

*Note.* CELF–4 = Clinical Evaluation of Language Fundamentals, Fourth Edition (Semel et al., 2006); WNL = within normal limits; NL = normal limits. Dash indicates not administered or not appropriate for child's chronological age.

generated to test for generalization to untreated but related behaviors. These included 10 less complex three-syllable strings with plosives (e.g., bataga, in both SW and WS forms, to total 20 stimuli) and 10 more complex four-syllable strings (e.g., butigabi, in both SW and WS forms, to total 20 stimuli). Set 4 included 10 familiar real words (e.g., motorbike) that shared features of Set 1 (i.e., three syllables and SW/WS stress pattern) but were more complex in number of phonemes ($M = 7.2$) and presence of consonant blends and later developing phonemes.

**Table 2.** Presence of characteristics of childhood apraxia of speech, based on perceptual judgment by three experienced speech-language pathologists.

| Feature of apraxia of speech | M1 | F1 | M2 |
|---|---|---|---|
| Slow speech rate | Mild | Mild | Mild-moderate |
| Equal stress across syllables | Mild | Mild | Moderate |
| Syllable segregation | Mild | Mild | Moderate |
| Inconsistent errors | Mild | Mild | Mild-moderate |
| Repetitions and revisions | Mild | Mild | Mild |
| Sound distortions | Mild | Mild | Mild |

All stimuli were presented orthographically in 24-point Times New Roman font on individual cards. Strong syllables in the nonwords were indicated with bold font, and vowels were written as they sounded (e.g., SW: **baa**teegoo; WS: baa**tee**goo).

## Design and Procedures

Single-subject designs with multiple baselines across behaviors and participants were used. Participants had three (M2) or four (M1 and F1) baseline tests on the speech behaviors of interest. Treatment was then provided in 60-min sessions 4 days per week for 3 weeks during a school vacation. Baselines and experimental and retention probes were administered to measure treatment effects, generalization to untreated items, and experimental control.

### Baseline and Experimental Probes

Each baseline test involved the participants reading aloud a randomly selected set of 10 treated strings, as well as 10 untreated strings of the same complexity, 10 less complex strings, 10 more complex strings, and 10 real words, for a total probe of 50 productions. Experimental probes identical to the baseline were administered after every fourth treatment session, prior to any treatment for the day to avoid the influence of recent clinician feedback and rehearsal on performance. Three experimental probes were completed during the treatment phase, with the third probe being the first posttreatment test. A fourth probe was given 4 weeks posttreatment to test retention of any treatment and generalization effects.

### Treatment

Treatment sessions began once all baseline tests had been completed. One student clinician was assigned to each participant and administered all treatment sessions and experimental probes. All sessions were supervised by the first or fourth authors via a one-way mirror and were video recorded for fidelity and reliability analyses. Each session included a prepractice and a practice component (Schmidt & Lee, 2005).

*Prepractice*. A random set of 10 treatment stimuli were selected and randomly inserted into sentence-final position of each of 10 carrier phrases (e.g., "He bought a____" or "Can you find my ____?"). Participants made attempts at producing these sentences with modeling from the clinician and detailed feedback on performance (i.e., KP feedback) to shape correct responses. This KP feedback informed the participants of the parameters of a correct response and prepared them for being able to adjust productions independently during the practice when no models or KP feedback would be provided. All participants needed verbal KP feedback addressing (a) relative duration of the first to second syllable in the target string, to realize SW or WS stress patterns; (b) maintaining habitual speech rate, rather than slowing speech rate; and (c) avoiding pauses between syllables in the string, between the carrier phrase and the string, or between presentation of the stimulus and onset of the response. Participants M1 and M2 required KP feedback to discourage increasing overall loudness for the syllable string. M2 required KP feedback to discourage exaggerating pitch variation across the syllable string. He also required rhythmic tapping by the clinician and himself, as well as visual aids, to shape appropriate relative durations of syllables without markedly slowed speech rate. The visual aids included short and long blocks arranged on the table. To aid the participants in processing these three types of feedback, a feedback sheet was made with the words *emphasis, fluency,* and *loudness*. The clinicians defined these terms to the participants (i.e., emphasis referred to getting the durational contrast; fluency to speaking at a habitual rate without pauses, hesitations, or repetitions; and loudness to overall intensity level for the sentence). They referred to each term as they provided the KR and KP feedback for each response (e.g., "Emphasis and fluency were spot on, but your voice was too loud"). Prepractice continued until the participant had produced five consecutive target sentences correctly, as perceived by the clinician, without a model.

*Practice*. The practice part of each session followed immediately after the prepractice. Between 100 and 120 practice trials were completed per session, comprising 10–12 trials for each of the 10 treated strings. Syllable strings, embedded in carrier phrases, were presented orthographically and read aloud by the participants. Stress pattern was indicated with bold font on the strong syllable. All stimuli were presented in random order, and KR feedback was provided on 50% of responses fading from 100% on the first 10 trials to 10% on the final 10 trials. Participants were told that they would not receive any modeling from the clinician, would be given feedback on accuracy (KR feedback) only, and should listen to their productions and self-evaluate each attempt. KR feedback initially was provided for stress pattern only, with a correct response being perceived as relative duration of about 3:2 for SW strings (i.e., the first syllable about 50% longer than the second) and 2:3 for WS strings (i.e., the second syllable about 50% longer than the first). Once success rates for stress assignment were above 50%, KR feedback was given on stress assignment, fluency (i.e., speech rate and presence of pauses), and, for M1 and M2 only, overall loudness level. For responses receiving KR feedback, the clinician would first rate stress assignment (e.g., "Good emphasis") and, when this was correct, then give feedback on fluency (e.g., "Nice and fluent") and loudness (e.g., "Not good on loudness"). To aid understanding of this more complex KR feedback, the feedback sheet with the words "emphasis," "fluency," and "loudness" (described above) was placed on the table. Rest breaks (i.e., a 3–5-min board game) were interspersed.

As is typical in clinical practice, all KP and KR feedback was based on the clinician's perceptual judgments of stress assignment, syllable duration and speech rate, vocal intensity, pitch variation, and speech naturalness. Perceptual judgments of stress assignment correlate highly with acoustic measures of stress assignment (Davis et al., 2000; also see Results section below).

## Equipment

All baselines and experimental probes were recorded in a quiet room using a Marantz PMD670 PC Card Recorder at 48 kHz. An Audio-Technica ATM75 cardioid headset microphone was placed 5 cm from the mouth.

## Dependent Measures

### Acoustic Measures

Acoustic measurements of syllable and/or vowel duration (ms), peak vocal intensity (dB), and peak fundamental frequency ($F_0$ in Hz) were made on all nonword and word productions in all tests. Acoustic analyses were done using the Praat signal-processing software (Boersma & Weenink, 2001). In nonsense syllable strings, all consonants were plosives. Syllable duration for the first and second syllables was measured as the time from onset of one plosive burst to onset of the next plosive burst. For real word stimuli, vowel duration was measured because the words were not controlled for number or type of phonemes within syllables. Measurement of onset and offset of vowels in the first two syllables followed the guidelines of Peterson and Lehiste (1960), utilizing pitch and intensity contours and formant trajectories generated within Praat software. Peak intensity (dB) and peak $F_0$ (Hz) measures for all syllables were made by selecting the vowel portion and generating the measures automatically using the Praat software algorithms.

*Pairwise variability indices of lexical stress*. For each stimulus item and each acoustic measure, the pairwise variability index (PVI; Low, Grabe, & Nolan, 2000; see Equation 1 below) was calculated to determine degree of asymmetry across the first two syllables of a string. PVI provides a measure normalized for speech rate (or loudness level, or $F_0$ level), for more accurate comparison across participants and within participants over time. A positive PVI is consistent with an SW pattern (i.e., greater duration, intensity, or $F_0$ on the first syllable), and a negative PVI is consistent with a WS pattern (i.e., greater duration, intensity, or $F_0$ on the second syllable), with increasing values indicating more pronounced contrast. A zero PVI value indicates equal stress over both syllables. The formula for duration is given by Equation 1:

$$\mathrm{PVI(dur)} = 100 \times \{(d_k - d_{k+1})/[(d_k + d_{k+1})/2]\}, \quad (1)$$

where $d$ is the duration of the $k^{\mathrm{th}}$ syllable.

Similar formulas were used for peak intensity in dB [PVI(int)] and peak $F_0$ in Hz [PVI($F_0$)].

*Total duration of first two syllables*. The average duration (DUR) of the first two syllables of the nonsense strings and real words was calculated (see Equation 2) to document changes in fluency (i.e., speech rate). This addressed whether control of lexical stress (i.e., relative duration, intensity, and pitch) was compromising speech rate, which can also affect speech naturalness:

$$\mathrm{DUR} = d_k + d_{k+1}, \quad (2)$$

where $d$ is duration of the $k^{\mathrm{th}}$ syllable.

The DUR measures were divided by number of phonemes to normalize the values across the nonword and real word data sets because real words were not controlled for number of phonemes in each syllable, as shown in Equation 3:

$$\mathrm{NORMDUR_{nor}} = (d_k + d_{k+1})/(m_k + m_{k+1}), \quad (3)$$

where $d$ is duration of the $k^{\mathrm{th}}$ syllable, and $m$ is the number of phonemes in the $k^{\mathrm{th}}$ syllable.

The average value for real word NORMDUR at baseline was used as a within-subject benchmark for the desired NORMDUR for the nonsense strings at posttreatment. All nonsense strings contained four phonemes (CVCV), while four real words had four phonemes, and six contained five phonemes. Durations of the four-phoneme and five-phoneme real words did not differ significantly within-subject ($p > .05$ for all comparisons; $t$ tests with Welch's correction for unequal variances). Similarly, no comparisons of the durations for SW versus WS tokens for real words at baseline or for nonsense strings posttreatment differed significantly ($p > .05$ for all comparisons). Therefore, values for SW and WS stimuli were pooled within each real and nonsense word stimulus list.

### Perceptual Measures

First, the percentage of responses perceived by each clinician to be correctly produced during each practice session was tallied to document change in ability to produce SW and WS stress contrasts on three-syllable nonsense strings in sentences. Criteria for a correct response were the same as those described for production of target stress pattern ("emphasis") and habitual speech rate ("fluency") during practice (see above). Accuracy for stress pattern alone and for stress pattern plus speech rate was recorded to demonstrate an accuracy/speed trade-off.

Second, three adults blinded to the study hypotheses and timing of samples perceptually judged (a) the 20 productions of the treated and untreated $C_1V_1C_2V_2C_3V_3$ strings and (b) the 10 real words from each child's final baseline test and immediate posttreatment test. The purpose of the latter judgments was to identify perceptible changes in production of the targeted SW and WS nonwords and words. For each stimulus type, the samples from each child were randomly ordered within participant. Samples were presented in free field from a laptop computer in a quiet room. Raters were instructed to adjust the volume to a comfortable level, listen to each sample only once, and rate it on a 5-point equal-appearing interval scale where 1 = *a clear SW pattern,* 2 = *tending to SW,* 3 = *equal stress over the first two syllables,* 4 = *tending to WS,* and 5 = *a clear WS pattern*. Order of nonwords and words for each participant and order of participants were counterbalanced across raters.

No measures of segmental accuracy or intelligibility were included in the study. The children were 100% intelligible and made few segmental errors pretreatment. The consonants in the experimental stimuli were earlier developing sounds that had been mastered by all children.

## Reliability

Interrater agreement on the independent variable (i.e., a checklist including administration of each step in the treatment protocol, judgment of response accuracy, and provision of KR feedback during practice) was calculated on a randomly selected 10% of trials across the three participants. Interrater measures were completed live by the first and fourth authors, who viewed all sessions via video camera. While these measures were not blinded, this procedure allowed any discrepancies in protocol administration to be discussed and corrected between sessions, without compromising the overall fidelity of the treatment. Point-to-point agreement was 85.1% for M1, 85.0% for F1, and 87.0% for M2. Intrarater agreement on the independent variable was measured by having each clinician view the recording of randomly selected trials at least 2 weeks later. For 18% of trials across the three participants, point-to-point agreement was 96.2% for M1, 98.0% for F1, and 95.6% for M2.

Interrater reliability on the primary dependent variable of syllable duration in the nonword strings was calculated for 12% of responses on probes across the three participants, and intrarater reliability was calculated on 8% of responses. For interrater measures on each child, the clinicians assigned to the other children were given randomly selected samples. For intrarater measures, a clinician remeasured a set of randomly selected samples at least 2 weeks after original measurement. In both cases, date and original syllable duration values were concealed. Wilcoxon matched-pairs signed-ranks tests showed no significant differences for inter- or intrarater measures ($p > .05$, Pearson $r = .98$ and $.99$, respectively). The average point-to-point interrater difference for syllable duration was 1.22 ms ($SE = 2.90$), and the average intrarater difference was 0.17 ms ($SE = 0.38$). Reliability was not calculated on the measures of peak vocal intensity and $F_0$ for vowels, as these measures were automatically generated using Praat software (Boersma & Weenink, 2001).

## Data Analysis

The perceptual measures from the practice sessions and the PVI measures from the baseline and experimental probes were graphed for visual inspection of treatment, generalization, and maintenance effects. Changes in PVI(dur), PVI(int), and PVI($F_0$) for SW versus WS syllable sequences from pre- to posttreatment were analyzed using the Kruskal–Wallis nonparametric analysis of variance (ANOVA) with Dunn's multiple comparisons post hoc test. Data for each participant and for each of the four stimulus sets (i.e., one treated set and three untreated sets) were analyzed separately. A conservative alpha level of .01 was used to adjust for multiple within-subject comparisons. Data for each participant were pooled across the final two baselines and the two posttreatment probes (i.e., Probe 3 and the 4-week retention probe) to increase statistical power for the untreated stimulus sets, which contained a small number of tokens.

The DUR measures for treated and untreated strings were compared from pre- to posttreatment using the nonparametric Mann–Whitney $U$ test. Again, data from the final two baselines and from the two retention probes were pooled for each participant. This analysis tested whether rate of production changed over the course of the study. Finally, the NORMDUR measures for real words were compared with the NORMDUR measures of the three nonsense string sets at pretreatment and at posttreatment using the Kruskal–Wallis nonparametric ANOVA with Dunn's multiple comparisons test to determine whether speech rate on the syllable strings differed from habitual rate.

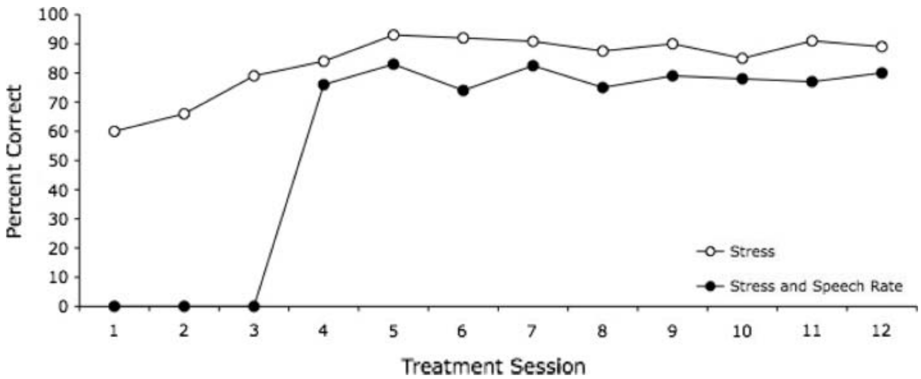
Perceptual ratings of stress production for each nonword or word in the baseline and immediate posttreatment probes were averaged over the three listeners. Average within-participant ratings from pre- to posttreatment were compared using the Mann–Whitney $U$ statistic to determine whether changes in stress production were perceptually robust. In addition, the correlation between average ratings and PVI values for each participant was calculated to determine the strength of relationship between these perceptual and instrumental measures.

# Results

## Perceptual Measures During Treatment

As shown in Figures 1–3, all three participants were perceived to improve in their ability to produce the two stress contrasts of SW and WS in the three-syllable treatment stimuli over the course of the 12 sessions. From the first to the final practice session, M1 improved from 0% to 80% correct for production of stress at habitual rate of speech, F1 improved from 0% to 75.8% correct, and M2 improved from 0% to 38% correct. M2, the youngest and most dysprosodic of the cases, did not reach the criterion of 80% correct over three consecutive sessions, but he did perform at or above 80% correct for production of the stress pattern alone in four of the last five practice sessions.

**Figure 1.** Percentage of M1's productions perceived to be correct for either stress pattern alone (open circles) or for stress pattern and habitual speech rate (closed circles) during the 12 treatment sessions.




## *PVIs of Lexical Stress*

### Treatment Effect

A positive treatment effect is represented by PVI values in experimental probes for SW and WS stimuli separating, with SW stimuli having a positive PVI posttreatment and WS having a negative PVI. Data for the treated $C_1V_1C_2V_2C_3V_3$ stimuli and the untreated $C_1V_1C_2V_2C_3V_3$ stimuli were pooled, as performance on both sets was not significantly different for all participants (Mann–Whitney *U* tests).

All children showed a significant treatment effect by developing appropriate durational contrasts for SW versus WS stimuli. For the treated stimuli at baseline, no differences were found in PVI(dur), PVI(int), or PVI($F_0$) across SW and WS stimuli for any participants (see Table 3 and Figures 4–6). All participants showed significant differences in PVI(dur) for SW compared to WS at posttreatment. M1 achieved this primarily through increasing the durational contrast in SW strings (i.e., more positive PVI), while F1 and M2 increased the durational contrast in WS strings (i.e., more negative PVI). M1 also showed a significant difference in PVI(int) and PVI($F_0$) for SW and WS nonwords at posttreatment, with some shift in PVI values for both stimulus types. F1 and M2 showed a significant posttreatment difference between SW and WS in PVI(int), although, for M2, the distinction was reduced at 4 weeks posttreatment. F1 and M2 did not use pitch to contrast SW and WS posttreatment.

### Generalization of Treatment Effects

As noted above, effects of treatment on the complex three-syllable strings (i.e., $C_1V_1C_2V_2C_3V_3$) generalized to the untreated exemplars of the same-level stimuli. Generalization to less complex three-syllable strings was evident to some degree in all participants (see the second row of charts in Figures 4–6 and Table 3). For M1, the

**Figure 2.** Percentage of F1's productions perceived to be correct for either stress pattern alone (open circles) or for stress pattern and habitual speech rate (closed circles) during the 12 treatment sessions.

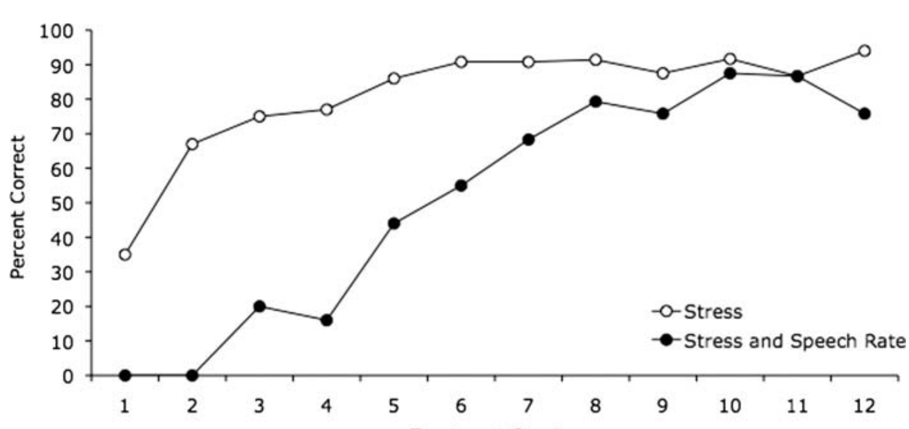
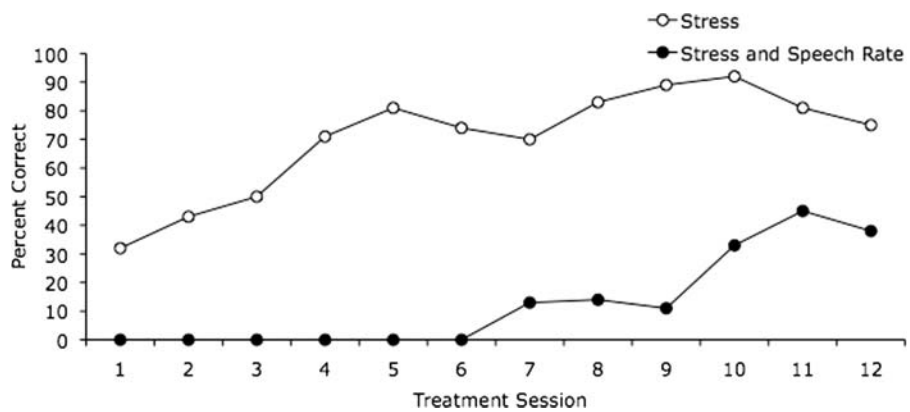


**Figure 3.** Percentage of M2's productions perceived to be correct for either stress pattern alone (open circles) or for stress pattern and habitual speech rate (closed circles) during the 12 treatment sessions.




separation of the SW and WS stimuli on all PVI measures can been seen in the final probes in Figure 4 and with the significant difference between SW and WS stimuli on PVI(dur) and PVI(int) measures at posttreatment (see Table 3). A similar change from pre- to posttreatment was seen in F1, although the focus of change was on the PVI(dur) and PVI($F_0$) measures (see Figure 5 and Table 3). For M2, PVI(dur) and PVI(int) for less complex WS stimuli improved significantly with treatment (see Figure 6 and Table 3). However, duration and intensity contrasts between SW and WS at posttreatment were not statistically reliable, likely due to loss of skill in the 4 weeks following treatment.

Generalization to more complex four-syllable strings was evident but more limited than the changes seen in less complex stimuli (see the third row of charts in Figures 4–6 and Table 3). For M1, the only finding of note is the significant improvement in differentiation of SW and WS productions for the PVI($F_0$) measure at posttreatment. Significant change in control of relative duration of syllables did not reach significance due to the variable performance on the PVI(dur) measure during baseline (see Figure 4). For F1, relative durations of syllables in SW and WS stimuli changed from being not significantly different in baseline to being a significant difference posttreatment (see Figure 5 and Table 3). This likely reflects limited generalization of treatment effects, with PVI(dur) on more complex productions changing but not PVI(int) or PVI($F_0$), as shown in Table 3. For M2, there was no reliable improvement differentiation of more complex SW and WS stimuli at posttreatment, compared to baseline (see Table 3). There was an undesirable shift of PVI(dur) for SW stimuli from near zero at baseline to negative values posttreatment (see Figure 6 and Table 3).

Production of real words showed a somewhat different and unexpected pattern of results in all participants. M1 at baseline marked the SW and WS stress contrasts primarily by intensity and pitch variations, with significant differences between SW and WS stimuli for PVI(int) and PVI($F_0$) values but not for PVI(dur), as indicated in Figure 4 and Table 3. With treatment focusing on relative durations of syllables, M1 produced a different pattern posttreatment. That is, PVI(dur) values were significantly different for SW and WS real word stimuli, but PVI(int) and PVI($F_0$) were no longer significantly different. F1 and M2, on the other hand, showed a significant difference between SW and WS words at baseline and posttreatment for PVI(dur), as shown in Figures 5 and 6, respectively, and Table 3. For M2, PVI(int) differentiated SW and WS words in baseline but not posttreatment.

### Retention of Treatment Effects

For the statistical analysis, data from immediate posttreatment probes and the 4-week retention probes were pooled, and so long-term retention was not analyzed separately. However, the top row of charts in Figures 4–6 shows the final retention data point for the treated stimuli. There was no or minimal deterioration in performance on the treated items over the 4-week posttreatment period for M1 and F1. M2's ability to differentiate the SW and WS stimuli in production was not maintained; while he produced both SW and WS stimuli with equal stress (i.e., PVI[dur] close to zero) in baseline, he tended to produce all treated strings as WS in the final retention probe (i.e., negative PVI values for both SW and WS stimuli).

## *Total Duration of First Two Syllables*

Two measures were used to track changes in duration of the initial two syllables of the treated and untreated stimulus items: DUR and NORMDUR. Consistent with an increase in familiarity and fluency with the treated nonword stimuli from intensive practice, values

**Table 3.** Results of statistical comparisons (Kruskal–Wallis test [KW] and Dunn's multiple comparisons post hoc test) of the pairwise variability index (PVI) for strong-weak (SW) versus weak-strong (WS) stimuli at pretreatment (Pre) and posttreatment (Post) for M1, F1, and M2.

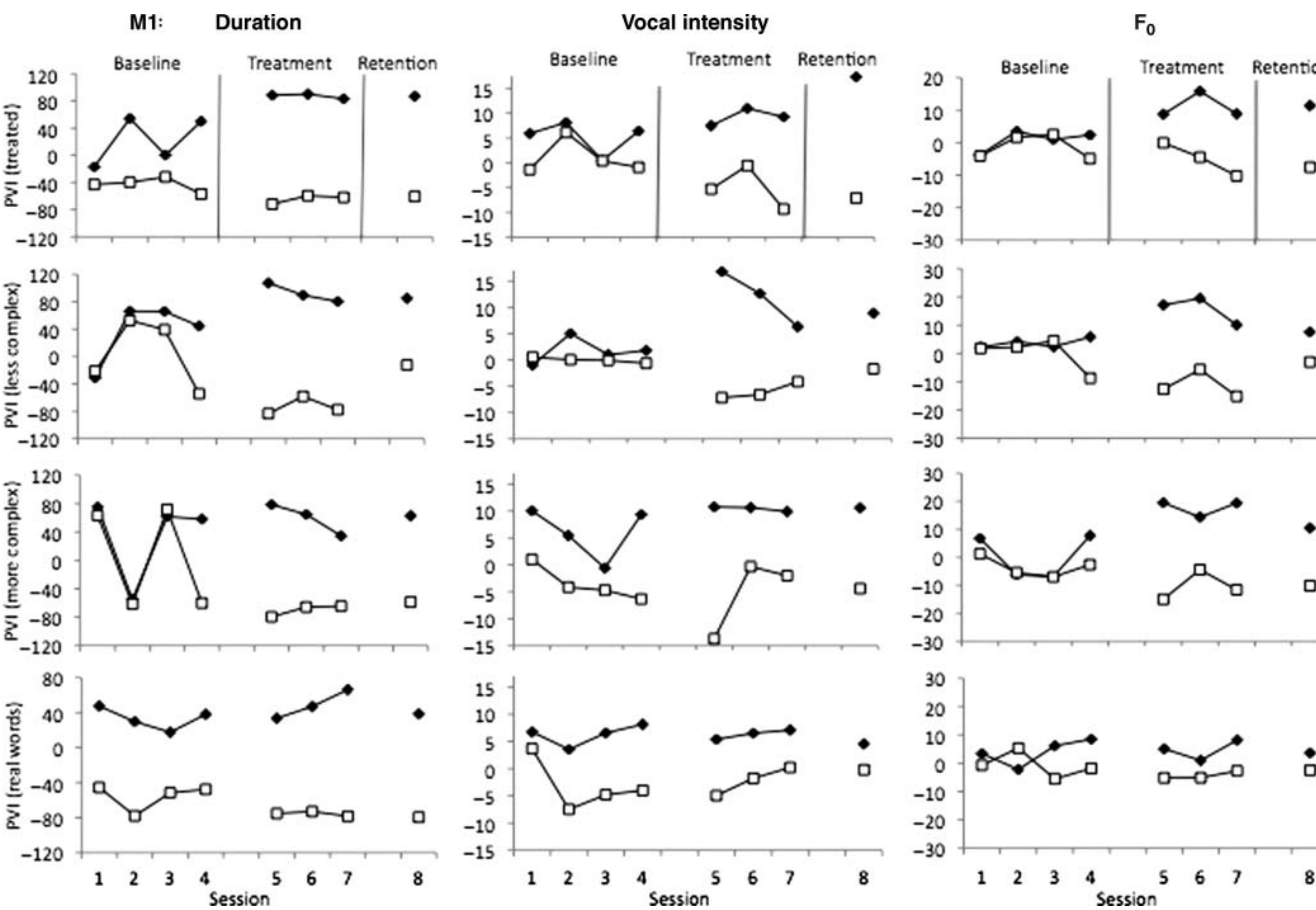
| Participant | PVI(dur) | PVI(int) | PVI($F_0$) |
|---|---|---|---|
| M1 | | | |
| Treated | KW = 45.34, *p* < .0001 | KW = 26.22, *p* < .0001 | KW = 29.55, *p* < .0001 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .001[a] | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *p* < .001 |
| | Pre vs. Post for SW: *p* < .01; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| Less complex | KW = 23.42, *p* < .0001 | KW = 17.29, *p* < .01 | KW = 16.77, *p* < .001[b] |
| | SW vs. WS at Pre: *p* < .05; Post: *p* < .01 | SW vs. WS at Pre: *ns*; Post: *p* < .01 | SW vs. WS at Pre: *p* < .05; Post: *p* < .05 |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| More complex | KW = 13.91, *p* < .01 | KW = 10.59, *p* < .05 | KW = 14.86, *p* < .01 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *p* < .01 |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| Real words | KW = 26.93, *p* < .0001 | KW = 19.64, *p* < .001 | KW = 18.86, *p* < .001 |
| | SW vs. WS at Pre: *p* < .05; Post: *p* < .001 | SW vs. WS at Pre: *p* < .01; Post: *ns* | SW vs. WS at Pre: *p* < .01; Post: *p* < .05 |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| F1 | | | |
| Treated | KW = 34.52, *p* < .0001 | KW = 31.58, *p* < .0001 | KW = 8.08, *p* < .05 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *ns* |
| Less complex | KW = 16.86, *p* < .001 | KW = 10.65, *p* < .05 | KW = 18.77, *p* < .001 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *ns* | SW vs. WS at Pre: *ns*; Post: *p* < .01 |
| | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *p* < .05 | Pre vs. Post for SW: *ns*; WS: *p* < .05 |
| More complex | KW = 22.35, *p* < .001 | KW = 10.14, *p* < .05 | KW = 7.44, *ns* |
| | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *p* < .05 |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| Real words | KW = 23.71, *p* < .0001 | KW = 15.42, *p* < .01 | KW = 4.99, *ns* |
| | SW vs. WS at Pre: *p* < .01; Post: *p* < .01 | SW vs. WS at Pre: *ns*; Post: *p* < .01 | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |
| M2 | | | |
| Treated | KW = 16.41, *p* < .001 | KW = 25.25, *p* < .0001 | KW = 10.76, *p* < .05 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .01 | SW vs. WS at Pre: *ns*; Post: *p* < .001 | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *ns* |
| Less complex | KW = 14.35, *p* < .01 | KW = 13.01, *p* < .01 | KW = 11.14, *p* < .05 |
| | SW vs. WS at Pre: *ns*; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *p* < .01 | Pre vs. Post for SW: *ns*; WS: *p* < .05 |
| More complex | KW = 20.79, *p* < .0001 | KW = 12.40, *p* < .01 | KW = 4.84, *ns* |
| | SW vs. WS at Pre: *ns*; Post: *ns* | SW vs. WS at Pre: *ns*; Post: *ns* | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *p* < .01; WS: *p* < .05 | Pre vs. Post for SW: *ns*; WS: *p* < .05 | Pre vs. Post for SW: *ns*; WS: *ns* |
| Real words | KW = 23.78, *p* < .0001 | KW = 27.21, *p* < .0001 | KW = 4.35, *ns* |
| | SW vs. WS at Pre: *p* < .01; Post: *p* < .01 | SW vs. WS at Pre: *p* < .001; Post: *p* < .05 | SW vs. WS at Pre: *ns*; Post: *ns* |
| | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* | Pre vs. Post for SW: *ns*; WS: *ns* |

*Note.* PVI was calculated over the first two syllables in each string, and comparisons are presented for duration (dur), intensity (int), and fundamental frequency ($F_0$). Pretreatment data are pooled across the final two baselines and posttreatment data across the two retention probes for each participant. Post hoc results show significance of comparisons between SW and WS at each time point and between time points for each lexical stress condition.

[a]Effect sizes for all significant post hoc tests were large (i.e., $d > 1.0$). [b]One of the remaining two post hoc comparisons was significant; these comparisons are not presented here because they were not considered informative (i.e., PVI for SW at pretreatment vs. WS at posttreatment or PVI for WS at pretreatment vs. SW at posttreatment).

of DUR decreased significantly for all participants from pre- to posttreatment (see Table 4). For untreated less complex nonwords, DUR decreased for M1 and M2 from pre- to posttreatment but remained unchanged for F1 (see Table 4). However, M1's and M2's posttreatment duration values were still longer than normal, with average durations of 571.1 ms and 678.6 ms, respectively. This equates to a speaking rate of about 3.5 syllables/s for M1 and 2.9 syllables/s for M2. More complex words decreased in duration for M1 only, reaching an average

**Figure 4.** M1's performance on baseline and experimental probes for treated and untreated stimulus sets. The dependent variables are the median pairwise variability indices (PVIs) for duration (left panel), intensity (middle panel), and $F_0$ (right panel). For each PVI variable, the treated three-syllable strings are in the top chart, the untreated less complex strings are in the second chart from the top, the untreated more complex strings are in the third chart, and the untreated real words are in the bottom chart. Filled symbols represent the strong-weak (SW) stress pattern, and open symbols represent the weak-strong (WS) pattern. A positive PVI indicates an SW stress pattern (e.g., PVI[dur] of 40 is equivalent to a 3:2 ratio), a negative PVI indicates a WS stress pattern, and a zero PVI value indicates equal stress.

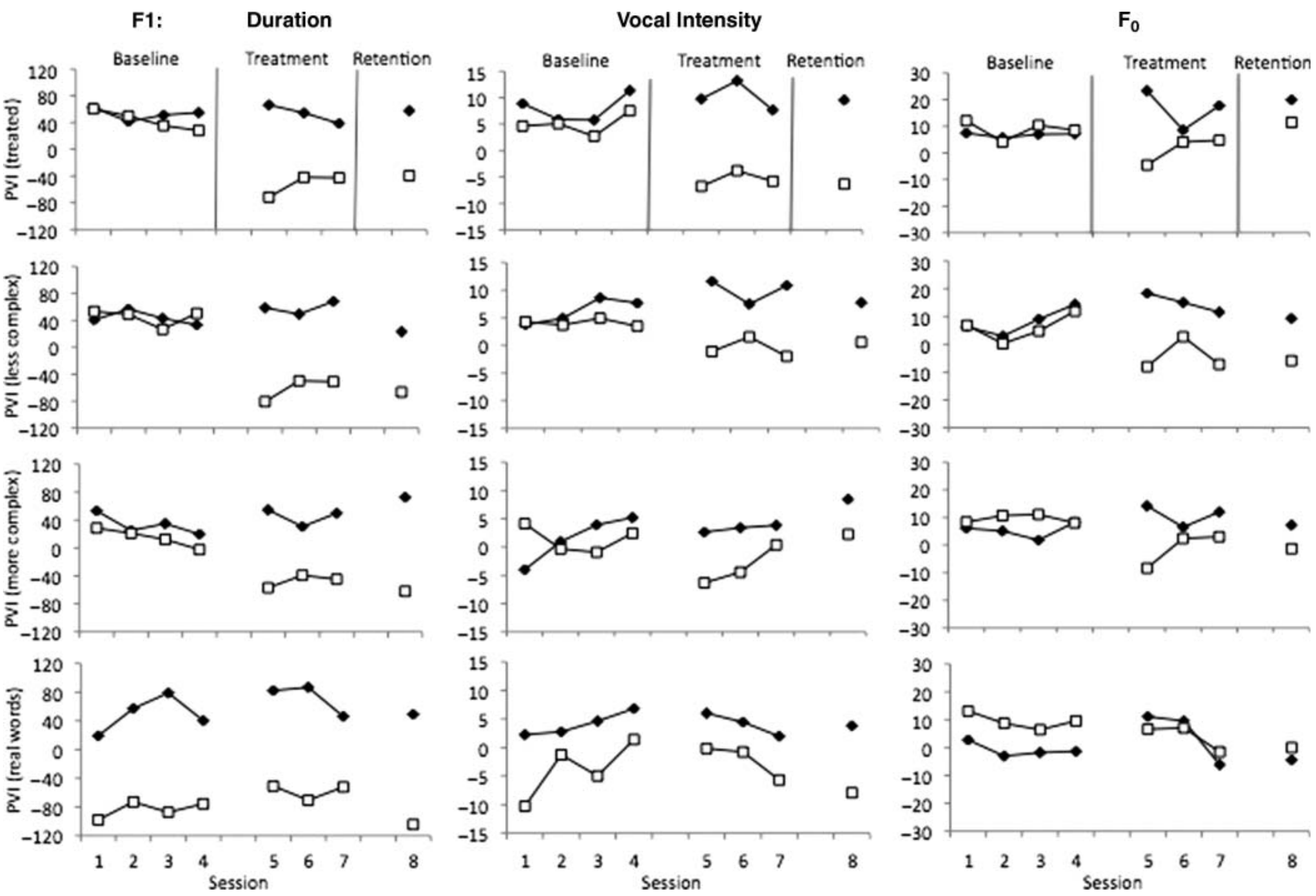



of 585.2 ms. For M2, the average durations of the more complex stimuli decreased from 913.2 ms to 757.1 ms but did not reach significance due to high variability on the baseline tests. For real words, as expected, the summed duration of the first two vowels, excluding consonants, showed no changes in duration for any participant.

The normalized average duration of the first two syllables of the real words (NORMDUR) at baseline was used as a within-subject benchmark for the desired duration of the first two syllables of the nonsense strings at posttreatment. For all children at pretreatment, NORMDUR was significantly different across stimulus sets (M1: Kruskal–Wallis [KW] = 53.93, $p < .0001$; F1: KW = 43.94, $p < .0001$; M2: KW = 43.72, $p < .0001$), with Dunn's multiple comparisons post hoc tests showing that the real words were significantly shorter than the treated and untreated nonsense strings ($p < .001$ for all comparisons). For M1, the durations of the treated strings were significantly shorter than the more complex strings at pretreatment ($p < .01$), but no other post hoc comparisons were significant. For F1 and M2, the durations of the treated and untreated strings did not differ from each other.

At posttreatment, the difference in NORMDUR across the four stimulus sets continued to be significantly different (M1: KW = 36.08, $p < .0001$; F1: KW = 42.92, $p < .0001$; M2: KW = 41.48, $p < .0001$), with post hoc tests showing that the real words continued to be significantly shorter ($p < .01$ for all comparisons) than the treated and untreated nonsense strings. For M1 and F1, the average NORMDUR for the treated strings was significantly shorter than the more complex untreated strings (M1: $p < .01$; F1: $p < .05$), but the difference between treated and less complex untreated strings did not reach

**Figure 5.** F1's performance on baseline and experimental probes for treated and untreated stimulus sets. See Figure 4 for the key.



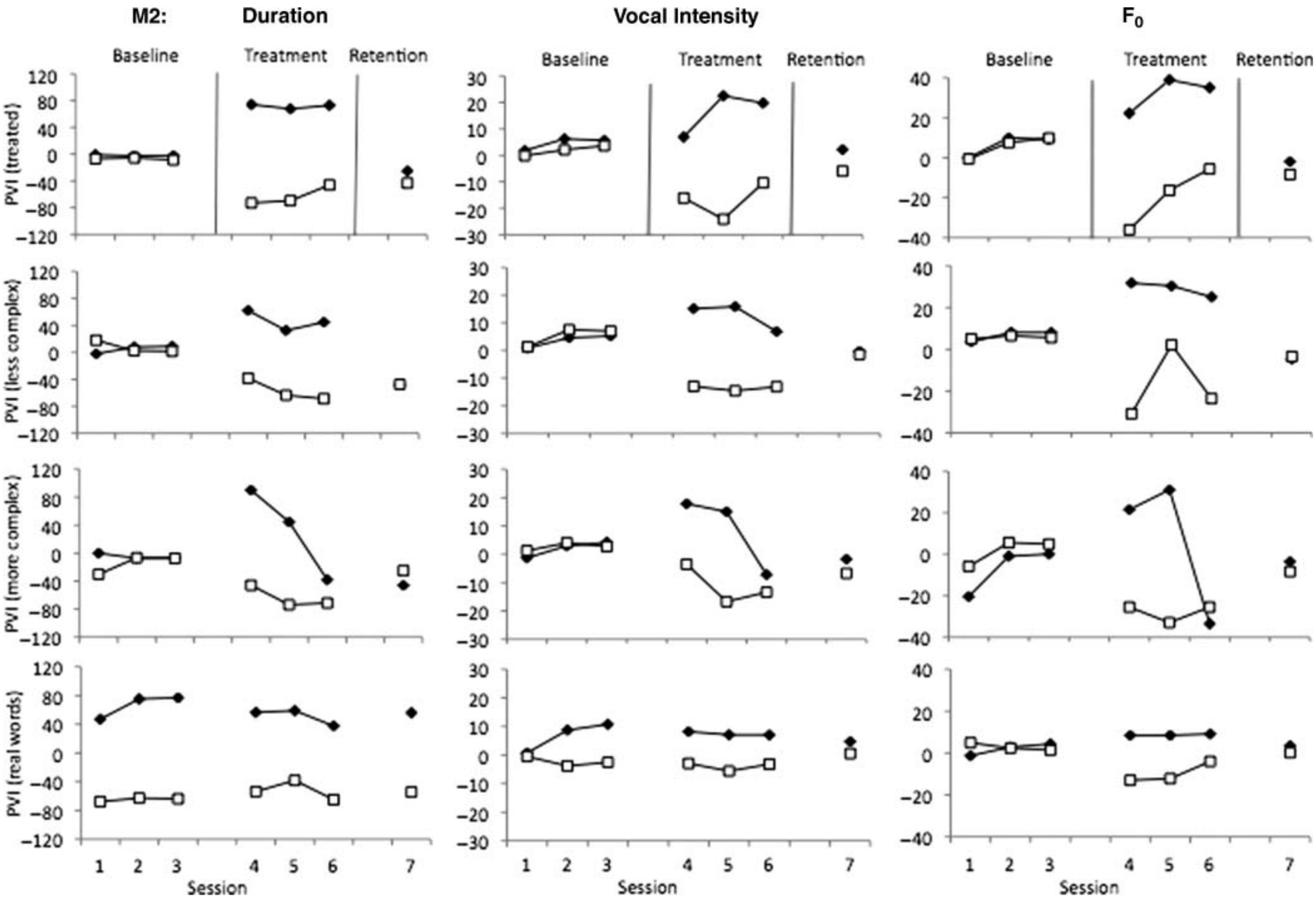

significance. For M2, NORMDUR was not significantly different across treated and untreated nonsense stimulus sets.

## Relationship Between Perceptual and Instrumental Measures

The average rating (out of 5) across three listeners for each nonword and real word response was entered into a nonparametric pre- to posttreatment paired comparison for each subject (see Table 5). An average rating near 1 indicated a clear SW pattern, a rating near 3 indicated equal stress, and a rating near 5 indicated a clear WS pattern. For M1, ratings of SW nonwords improved significantly, from an average of 2.39 to 1.24; ratings of WS strings did not differ significantly from pretreatment to posttreatment, being 4.11 and 4.26, respectively. For F1, only ratings of WS nonwords improved significantly, from an average of 2.45 to 4.39; ratings for SW strings were 2.00 at pretreatment and 1.64 at posttreatment. For M2, average ratings for both SW and WS nonwords improved significantly with treatment (SW: 2.67 to 2.09; WS: 2.67 to 4.04). No pre–post statistical comparisons were significant for average perceptual ratings of real words. This is likely due to the small sample size. However, all three children showed a tendency for stress on SW real words to become more distinct from pre- to posttreatment (i.e., average rating moving closer to 1; see Table 5). There was minimal change in WS stress production on the real words. These findings are remarkably similar to the results of the pre- to posttreatment statistical comparisons of PVI(dur) measures (see Table 3).

The relationship between perceived stress pattern of nonwords and real words and acoustic PVI measures was examined to determine whether the acoustic measures were capturing the perceptually based diagnostic feature of dysprosody in CAS. The analyses included the items that had been perceptually rated by independent listeners (see the Perceptual Measures subsection above)—that is, 20 treated and untreated $C_1V_1C_2V_2C_3V_3$ nonwords and the 10 real words from each child's final baseline test and immediate posttreatment test (i.e., 40 nonwords and 20 real words per child). For each production, the average perceptual rating for the three

**Figure 6.** M2's performance on baseline and experimental probes for treated and untreated stimulus sets. See Figure 4 for the key.




listeners was used. For M1, the perceptual ratings were highly correlated with all PVI measures, PVI(dur): $r = -.91$, $p < .0001$, 95% CI [−.95, −.82]; PVI(int): $r = -.51$, $p < .001$, 95% CI [−.71, −.23]; PVI($F_0$): $r = -.70$, $p < .0001$, 95% CI [−.83, −.50]. For F1, the perceptual ratings were significantly correlated with PVI(dur) and PVI(int) but not PVI($F_0$), PVI(dur): $r = -.80$, $p < .0001$, 95% CI [−.89, −.65]; PVI(int): $r = -.42$, $p < .01$, 95% CI [−.65, −.12]; PVI($F_0$): $r = -.26$, $p > .05$, 95% CI [−.53, −.06]. Similar to M1, all of M2's PVI measures were highly correlated with

**Table 4.** Average summed duration (and standard deviation) of the first two syllables in the treated and untreated nonsense syllable strings and the first two vowels, excluding consonants, in the real words compared from pretreatment to posttreatment for M1, F1, and M2.

| | M1 | | F1 | | M2 | |
|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post |
| Treated | 595.5 (81.7) | 516.1 (94.6)** | 636.6 (143.9) | 544.2 (69.9)** | 773.9 (97.7) | 674.2 (97.8)*** |
| Less complex | 669.7 (72.3) | 571.1 (58.8)*** | 613.4 (143.1) | 597.7 (63.9)$^{ns}$ | 819.6 (125.9) | 678.6 (178.9)* |
| More complex | 708.7 (107.9) | 585.2 (65.1)*** | 677.4 (83.4) | 674.1 (97.8)$^{ns}$ | 913.2 (310.9) | 757.1 (65.4)$^{ns}$ |
| Real words | 209.7 (76.7) | 205.5 (80.7)$^{ns}$ | 171.3 (49.8) | 150.7 (45.6)$^{ns}$ | 247.0 (93.1) | 250.4 (90.2)$^{ns}$ |

*Note.* Pretreatment (Pre) data are pooled across the final two baselines and posttreatment (Post) data across the two retention probes for each participant. Asterisks indicate statistically significant differences between pre- and posttreatment measures for each stimulus type. Nonparametric two-sided Mann–Whitney *U* test was used. A conservative alpha level of $p < .01$ was used to adjust for multiple comparisons within participants. Effect sizes for all significant comparisons were large (i.e., $d > 0.82$).

*$p < .01$. **$p < .001$. ***$p < .0001$.

**Table 5.** Average perceptual rating (and standard deviation) of lexical stress pattern for the treated and untreated nonsense syllable strings and the real words at pretreatment and posttreatment for M1, F1, and M2.

| | M1 | | F1 | | M2 | |
|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post |
| Treated strings | | | | | | |
| SW | 2.39 (1.03) | 1.24 (0.30)** | 2.00 (0.65) | 1.64 (0.71)$^{ns}$ | 2.67 (0.36) | 2.09 (1.31)* |
| WS | 4.11 (0.44) | 4.26 (0.57)$^{ns}$ | 2.46 (1.05) | 4.39 (0.75)** | 2.67 (0.33) | 4.04 (1.10)* |
| Real words | | | | | | |
| SW | 1.94 (0.61) | 1.50 (0.59)$^{ns}$ | 1.94 (0.57) | 1.67 (0.47)$^{ns}$ | 2.06 (0.68) | 1.50 (0.18)$^{ns}$ |
| WS | 4.25 (0.50) | 4.25 (0.88)$^{ns}$ | 4.12 (0.19) | 4.25 (0.74)$^{ns}$ | 3.83 (0.33) | 3.92 (0.92)$^{ns}$ |

*Note.* Ratings (where 1 = *clear fluent SW pattern,* 3 = *equal stress,* and 5 = *clear fluent WS pattern*) were provided by three listeners. Asterisks indicate statistically significant differences between pre- and posttreatment ratings for each behavior. Nonparametric two-sided Mann–Whitney $U$ test was used. A conservative alpha level of $p < .01$ was used to adjust for multiple comparisons within participants. Effect sizes for all significant comparisons were large (i.e., $d > 1.5$), with the exception of the SW treated strings for M2, which showed a medium effect size ($d = 0.60$).

*$p < .01$. **$p < .001$.

perceptual ratings, PVI(dur): $r = -.84$, $p < .0001$, 95% CI [−.91, −.71]; PVI(int): $r = -.80$, $p < .0001$, 95% CI [−.89, −.64]; PVI($F_0$): $r = -.55$, $p < .001$, 95% CI [−.74, −.28].

## Discussion

This study tested the efficacy of a treatment for improving production of lexical stress contrasts in three children with CAS. It was predicted that (a) the children would improve in their ability to produce duration contrasts across syllables in treated three-syllable strings, (b) this improvement would generalize to less complex three-syllable strings, (c) effects of treatment would not generalize to more complex four-syllable strings and possibly not to real words, and (d) perceptual and acoustic measures of lexical stress would be highly correlated. The first two hypotheses were strongly supported. There was some evidence of generalization to more complex stimuli, which is encouraging, albeit counter to the third hypothesis. The strong relationship found between acoustic and perceptual measures of treatment effects supports both use of these objective measures to quantify change in experimental studies and reliance on the faster and more economical perceptual measures in clinical practice.

### *Contributions of the Study*

The current study makes three contributions to our understanding of the nature and treatment of CAS and prosodic disturbances. First, it is the first study to apply within-subject experimental designs to test the efficacy of a treatment for dysprosody in CAS (ASHA, 2007; Morgan & Vogel, 2008). The three children reported in this study provided a unique opportunity to study prosodic deficits in CAS. They are siblings with a history of normal language and intellectual development and remarkably similar motor speech impairment. Previous studies of treatment for dysprosody have included children with diverse coexisting impairments such as language delay (Shea & Tyler, 2001) or autism (Bellon-Harn, Harn, & Watson, 2007). Lexical stress errors have been shown to differentiate children with CAS from children with other speech sound production disorders (Nijland et al., 2003; Shriberg et al., 1997). Furthermore, the dysprosody in CAS has been reported to persist into adolescence (Velleman & Shriberg, 1999), despite these children often receiving considerable intervention for segmental errors. On the other hand, the relative homogeneity of this participant sample may limit generalization of the findings to the broader population of children with CAS and to those who still demonstrate reduced intelligibility due to segmental errors.

Second, the study provides support for the hypothesis that impaired ability for learning to produce prosodic variations is a primary feature of CAS rather than a consequence of disrupted speech sound production. Furthermore, the study highlights the difficulty of children with CAS in controlling temporal aspects of speech production. These children represented a continuum of severity from M2, the youngest with the most prosodically disrupted speech, to M1, whose prosody was only mildly affected. Nonetheless, the treatment task was clearly a challenge for all three participants. All required numerous trials to achieve improvement (see Figures 1–3), and all showed some vocal and nonverbal signs of task-related stress. For example, F1 and M2 tended to move their head upward on the beat of a stressed syllable, M1 made swallowing and oral preparatory movements prior to beginning the sentence, and M1 and M2 increased their overall vocal intensity inappropriately during production of the stimuli on practice trials. These behaviors were discussed with the participants within one session of their appearance and, in all cases, dissipated within three sessions.

Additional evidence that the children experienced difficulty with the treatment task was found in the analyses of syllable durations. The normalized duration of the first two syllables of real words was much shorter than the duration of the first two syllables in the nonsense strings. While durations for the treated strings decreased from pre- to posttreatment, they continued to be significantly longer than the real words. There are likely two explanations for this finding. One is that the strings were not practiced sufficiently to reach the fluency level of real speech. Second, while the strings were practiced in carrier sentences, no meaning was attached to them. If novel but real words had been selected for treatment, it is possible that they would have been integrated into the mental lexicon for use in daily communication, which might have facilitated independent rehearsal or use and thus more fluent or faster production.

It is important to note that M2 showed poorer performance than M1 and F1 during treatment (see Figure 3) and did not retain the treatment effects at 4 weeks posttreatment (see Figure 6). This may be related to the overall greater severity of his dysprosody, suggesting that he required more treatment sessions to maximize and stabilize performance. It is also possible that there was a clinician effect, as each child had a different clinician. This was minimized by having the supervising clinicians involved with all children and ensuring that interrater reliability on the independent variable was high (i.e., perceptual judgment of responses and provision of KP and KR feedback during treatment session). Future studies should control for this effect.

The third contribution of the current study is in understanding the interplay of duration, intensity, and $F_0$ in signaling stress contrasts and their relationship to the perception of lexical stress patterns. During treatment, practice was directed toward controlling only the durational contrast. Despite this, important changes were observed in variation of intensity and $F_0$ contrasts, consistent with the well-established finding that these three features are functionally linked. Davis et al. (2000) reported that adults most often varied just intensity plus frequency with first-syllable stress but varied vowel duration and/or intensity and frequency in almost all strings with second-syllable stress. Their adult data set included both real word and nonword disyllables. Here, we chose to have the children focus on syllable duration to mark first- and second-syllable stress, and, for the nonsense strings, syllable rather than vowel duration was measured. As such, the dominance of duration variation to mark stress is likely to be a bias introduced by our methodology. However, Davis and colleagues observed that, when only one parameter is varied to mark stress, it tended to be duration in infants but intensity in adults. All three children here showed strong durational contrasts, as well as intensity contrasts, to mark stress in real words.

The corroboration of the quantitatively determined treatment effects with data from perceptual ratings is comforting for two reasons. The measures of relative duration, intensity, and $F_0$ within utterance captured the perception of excessive/equal/misplaced lexical stress as documented by Shriberg and colleagues (Shriberg et al., 1997; Velleman & Shriberg, 1999) and identified in the ASHA (2007) position statement on diagnosis of CAS. Our findings support the use of these measures to objectively quantify lexical stress patterns in CAS and to measure change with intervention. In addition, the high correlations between perceptual ratings and the PVI measures support the continued reliance on perceptual judgments in daily clinical practice as valid indicators of impairment and treatment-related change. However, interrater reliability measures will continue to prove useful in safeguarding against problems such as perceptual drift (Kent, 1996).

One problem with the current study lies in the children's ability to differentiate SW and WS stress contrasts, through manipulation of duration and intensity, in highly familiar real words during baseline. The children were referred to the study with parental report of dysprosody, and clinicians in the study corroborated this. It is possible that the acoustic measures used here were not sensitive to more subtle residual disruptions in lexical stress on familiar and well-learned words or that some other measure or measures might have better captured the dysprosody. Nonetheless, the children demonstrated considerable difficulty mastering the experimental stimuli. Future studies might include measures of intersyllabic pauses or coarticulation. PVI measures of real word production in typically developing children would aid interpretation, along with information on speed of response to training on novel words.

A related point that begs attention is the use of nonword stimuli in treatment. While the three children were able to differentiate SW and WS real words during baseline, it is reasonable to propose that acquisition of this contrast would have taken longer than normal given the protracted rate of acquisition seen during the treatment. Evidence from studies of typically developing children supports this assumption (Davis et al., 2000; Pollock et al., 1993). Davis et al. (2000) reported that infants as young as 12 months can manipulate vowel duration, intensity, and $F_0$ to produce alternating stress patterns across syllable strings in prelinguistic babble. Pollock et al. (1993) showed that 3-year-old children successfully control vowel duration, intensity, and pitch to mark SW and WS contrasts on imitated $C_1V_1C_2V_2$ nonwords within a single testing session. Therefore, the choice of nonwords here simulated learning of unfamiliar words. It also allowed for careful control of stimulus characteristics in the experimental setting and rapid, highly reliable measurement of syllable durations. In working with younger children, however, novel real words might be a more functional type of stimulus. This also raises the issue that no formal and reliable tests of processing and production of lexical stress are currently available to guide diagnosis and intervention for dysprosody in Australian English-speaking children.

## Future Directions

The current study motivates further development of the treatment approach. The nonword stimuli were highly constrained and were not assigned meaning, although they were presented as nouns in the carrier sentences during practice. More research is required to determine whether best outcomes are obtained with real or nonsense words (see Gierut & Morrisette, 2010). In addition, measures of connected speech could be included to explore generalization of treatment effects to more functional speech tasks.

At times the clinicians reported difficulty with providing rapid, confident judgments of response accuracy during practice. An automated instrumental procedure would assist this, removing potential error and bias associated with perceptual judgments. However, intra- and interrater reliability and the correlation between perceptual and acoustic measures were high, suggesting instrumental measures are not necessary for clinical practice.

In conclusion, this approach shows promise for treating production of lexical stress contrasts in CAS. Although lexical stress is produced through variations in syllable duration and peak intensity and fundamental frequency, participants may only need to focus on one variable to achieve change across all three. This simplifies the therapy task for children. Structuring the intervention according to the PML approach probably stimulated strong retention and generalization effects; however, this was not specifically tested here. Further research is required to test this intervention with a larger sample of children with CAS and measure effects when combined with intervention for segmental accuracy in younger children.

## Acknowledgments

The study was supported by University of Sydney grants to Kirrie J. Ballard and Patricia McCabe. We wish to thank Samantha Warhurst, Lauren van Rees, and Monique Signorelli for assistance with treatment delivery and data analysis. We are indebted to the family for their support and willingness to be involved in the project.

## References

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

**Austermann Hula, S., Robin, D. A., Maas, E., Ballard, K. J., & Schmidt, R. A.** (2008). Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. *Journal of Speech, Language, and Hearing Research, 51,* 1088–1113.

**Ballard, K. J., Maas, E., & Robin, D. A.** (2007). Treating voicing control in apraxia of speech with variable practice. *Aphasiology, 21,* 1195–1217.

**Bellon-Harn, M. L., Harn, W. E., & Watson, G. D.** (2007). Targeting prosody in an eight-year-old child with high-functioning autism during an interactive approach to therapy. *Child Language Teaching and Therapy, 23,* 157–179.

**Boersma, P., & Weenink, D.** (2001). Praat 3.9.15 [Computer software]. Amsterdam, the Netherlands: Institute of Phonetic Sciences.

**Cumley, G. D., & Swanson, S.** (1999). Augmentative and alternative communication options for children with developmental apraxia of speech: Three case studies. *Augmentative and Alternative Communication, 15,* 110–125.

**Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K.** (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development, 71,* 1258–1270.

**Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A.** (2002). *Diagnostic Evaluation of Articulation and Phonology.* Oxford, England: Pearson Assessment.

**Duffy, J. R.** (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Mosby-Yearbook.

**Forrest, K.** (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology, 12,* 376–380.

**Gathercole, S., & Baddeley, A.** (1996). *The Children's Test of Nonword Repetition.* London, England: Harcourt Assessment.

**Gierut, J. A., & Morrisette, M. L.** (2010). Phonological learning and lexicality of treatment stimuli. *Clinical Linguistics & Phonetics, 24,* 122–140.

**Goldman, R., & Fristoe, M.** (2000). *Goldman Fristoe Test of Articulation—Second Edition*. Bloomington, MN: Pearson Assessment.

**Hall, P. K., Jordan, L., & Robin, D. A.** (2007). *Developmental apraxia of speech: Theory and clinical practice* (2nd ed.). Austin, TX: Pro-Ed.

**Hargrove, P., Anderson, A., & Jones, J.** (2010). A critical review of interventions targeting prosody. *International Journal of Speech-Language Pathology, 11,* 298–304.

**Kager, R.** (2007). Feet and metrical stress. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 195–227). Cambridge, England: Cambridge University Press.

**Kent, R. D.** (1996). Hearing and believing: Some limits to auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7–23.

**Kleim, J. A., & Jones, T. A.** (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research, 51,* S225–S239.

**Knock, T. R., Ballard, K. J., Robin, D. A., & Schmidt, R.** (2000). Influence of order of stimulus presentation on speech motor learning: A principled approach to treatment for apraxia of speech. *Aphasiology, 14,* 653–668.

**Low, E. L., Grabe, E., & Nolan, F.** (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43,* 377–401.

**Ludlow, C. L., Hoit, J., Kent, R., Ramig, L. O., Shrivastav, R., Strand, E., … Sapienza, C. M.** (2008). Translating principles of neural plasticity into research on speech motor control recovery and rehabilitation. *Journal of Speech, Language, and Hearing Research, 51,* S240–S258.

**Maas, E., Barlow, J., Robin, D. A., & Shapiro, L. P.** (2002). Treatment of sound errors in aphasia and apraxia of speech: Effects of phonological complexity. *Aphasiology, 16,* 609–622.

**Maas, E., Robin, D. A., Austermann-Hula, S. N., Freedman, S. E., Wulf, W., 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.

**McNeil, M. R., Robin, D. A., & Schmidt, R. A.** (1997). Apraxia of speech: Definition, differentiation, and treatment. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 311–344). New York, NY: Thieme.

**Morgan, A. T., & Vogel, A. P.** (2008). Intervention for childhood apraxia of speech. *Cochrane Database of Systematic Reviews,* Issue 3, Article No. CD006278.

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

**Munson, B., Bjorum, E. M., & Windsor, J.** (2003). Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language, and Hearing Research, 46,* 189–202.

**Nijland, L., Maassen, B., Van Der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R.** (2003). Planning syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 17,* 1–24.

**Pearson Education.** (2008). *Clinical Evaluation of Language Fundamentals, Fourth Edition: Technical report: Correlations between the CELF-4 and WISC-4 Integrated*. Retrieved from http://pearsonassess.com/NR/rdonlyres/D9546406-0147-4B51-A283-ECD252694FC1/0/CELF4_WISC4_TechReport.pdf.

**Peterson, G. E., & Lehiste, I.** (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32,* 693–703.

**Pollock, K. E., Brammer, D. M., & Hageman, C. F.** (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics, 21,* 183–203.

**Powell, T. W.** (1996). Stimulability considerations in the phonological treatment of a child with a persistent disorder of speech-sound production. *Journal of Communication Disorders, 29,* 315–333.

**Robin, D. A., Maas, E., Sandberg, Y., & Schmidt, R. A.** (2007). Motor control and learning and childhood apraxia of speech. In P. K. Hall, L. Jordan, & D. A. Robin (Eds.), *Developmental apraxia of speech: Theory and clinical practice* (2nd ed., pp. 67–86). Austin, TX: Pro-Ed.

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

**Schmidt, R., & Lee, T.** (2005). *Motor control and learning: A behavioral emphasis* (4th ed.). Champaign, IL: Human Kinetics.

**Schneider, S. L., & Frens, R. A.** (2005). Training four-syllable CV patterns in individuals with acquired apraxia of speech: Theoretical implications. *Aphasiology, 19,* 451–471.

**Semel, E., Wiig, E. H., & Secord, W. A.** (2006). *Clinical Evaluation of Language Fundamentals, Fourth Edition: Australian*. Marrickville, New South Wales, Australia: Harcourt Assessment.

**Shea, R. L., & Tyler, A. A.** (2001). The effectiveness of a prosodic intervention on children's metrical patterns. *Child Language Teaching and Therapy, 17,* 55–76.

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

**Skinder, A., Strand, E. A., & Mignerey, M.** (1999). Perceptual and acoustic analysis of lexical and sentential stress in children with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 7,* 133–144.

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

**Velleman, S. L., & Shriberg, L. D.** (1999). Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research, 42,* 1444–1460.

**Williams, P., & Stephens, H.** (2004). *Nuffield Centre Dyspraxia Programme*. Windsor, United Kingdom: The Miracle Factory.

Received July 1, 2009

Accepted February 18, 2010

DOI: 10.1044/1092-4388(2010/09-0130)

Contact author: Kirrie J. Ballard, Speech Pathology, Faculty of Health Sciences, University of Sydney, P.O. Box 170, Lidcombe, New South Wales 1825, Australia. E-mail: kirrie.ballard@sydney.edu.au.

**Appendix.** Stimulus list.

| Treated three-syllable strings | Untreated three-syllable strings | Untreated less complex strings | Untreated more complex strings | Untreated real words | Carrier sentences used in treatment |
|---|---|---|---|---|---|
| baguti | gatibu | bigiti | gubitaba | crocodile (SW) | I saw a |
| bitagu | tabigu | bitigi | tagibutu | cucumber (SW) | Can you find my |
| bugita | bugati | bataga | tibagatu | hamburger (SW) | He bought a |
| tabugi | bigatu | bugutu | batugibi | kangaroo (SW) | Here's the new |
| tigabu | gubati | tigibi | tigabubu | motorbike (SW) | She has a big |
| tubiga | butiga | tabaga | gatibugu | pattycake (SW) | There's my |
| gatubi | gabitu | tugubu | gubatigi | computer (WS) | It's going to |
| gibatu | tibagu | gibiti | gabituga | echidna (WS) | It's a purple |
| gutiba | batugi | gataba | gutibati | spaghetti (WS) | I want a |
| gubita | butagi | gutubu | bigutagi | toboggan (WS) | I went to the |

*Note.* Examples are presented rather than the complete stimulus set. SW = strong-weak stress pattern; WS = weak-strong stress pattern. SW real words all have the SWS stress pattern, although, for some, primary stress is on the final syllable.