Developmental Neurorehabilitation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ipdr20

Improvements in Speech of Children with Apraxia: The Efficacy of Treatment for Establishing Motor Program Organization (TEMPO^{SM})

Hilary E. Miller, Kirrie J. Ballard, Jenna Campbell, Madison Smith, Amy S. Plante, Semra A. Aytur & Donald A. Robin

To cite this article: Hilary E. Miller, Kirrie J. Ballard, Jenna Campbell, Madison Smith, Amy S. Plante, Semra A. Aytur & Donald A. Robin (2021) Improvements in Speech of Children with Apraxia: The Efficacy of Treatment for Establishing Motor Program Organization (TEMPO^{SM}), Developmental Neurorehabilitation, 24:7, 494-509, DOI: 10.1080/17518423.2021.1916113

To link to this article: https://doi.org/10.1080/17518423.2021.1916113

Published online: 09 Jul 2021.

DEVELOPMENTAL NEUROREHABILITATION
2021, VOL. 24, NO. 7, 494–509
https://doi.org/10.1080/17518423.2021.1916113

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Improvements in Speech of Children with Apraxia: The Efficacy of Treatment for Establishing Motor Program Organization (TEMPO^{SM})

Hilary E. Miller[a*], Kirrie J. Ballard[b], Jenna Campbell[a], Madison Smith[a], Amy S. Plante[a], Semra A. Aytur[a], and Donald A. Robin[a]

[a]University of New Hampshire, Durham, NH, USA; [b]University of Sydney, Sydney, Australia

ABSTRACT
Purpose: This study investigated the efficacy of Treatment for Establishing Motor Program Organization (TEMPO^{SM}) in childhood apraxia of speech (CAS).
Method: A mixed between- and within-participant design with multiple baselines across participants and behaviors was used to examine acquisition, generalization, and maintenance of skills. TEMPO^{SM} was administered in four one-hour sessions a week over a four-week period for eleven participants (ages 5 to 8), allocated to either an immediate treatment group or a wait-list control group. Acoustic and perceptual variables were measured at baseline, immediate post-treatment, and one-month post-treatment.
Results: Children demonstrated significant improvements in specific acoustic measures of segmentation and lexical stress, as well as perceptual measures of fluency, lexical stress, and speech-sound accuracy. Treatment and generalization effects were maintained one-month post-treatment with generalization to untreated stimuli.
Conclusion: TEMPO^{SM} was efficacious in improving segmental and suprasegmental impairments in the speech of children with CAS.

ARTICLE HISTORY
Received August 13, 2020
Revised April 07, 2021
Accepted April 08, 2021

KEYWORDS
Childhood apraxia of speech; prosody; lexical stress; intervention; speech motor programming

Introduction

Childhood apraxia of speech (CAS) is a motor speech disorder characterized by an impairment in the programming of spatial and temporal parameters for accurate speech movement patterns.[1,2] Specifically, apraxia of speech is a breakdown in the translation of intact phonological plans into the exact movement parameters required for segmental and prosodic accuracy of speech production.[2] Although the perceptual features that define the disorder have long been the subject of debate, there is emerging consensus that childhood and acquired forms of apraxia are unified by a common set of differentially diagnostic features including segmentation (increased segment and inter-segment duration), reduced stress contrast across syllables in polysyllabic words, and distortion of speech sounds.[2,3]

Apraxia of speech has been conceptualized within Klapp's motor programming model in which there is an internal working memory buffer (INT) that holds motor units prior to execution, and a sequencer (SEQ) that places those units in the correct serial order after movement onset.[4,5] Adult speakers with apraxia demonstrate impairments in INT, and not SEQ, indicating a deficit in the efficient programming of motor units.[6] As motor learning occurs, individual speech units (e.g. syllables) are concatenated into larger motor programs such as polysyllabic words;[7] however, this process is impaired in speakers with apraxia as more complex programs place an increased load on INT.[6] Specifically, inefficient concatenation explains the perception that individuals with apraxia "speak one syllable at a time," or segment their speech.[8] Concatenation also allows for accurate programming of coarticulatory effects and prosodic patterns across syllables, both of which are notably impaired in speakers with apraxia of speech, since the application of the suprasegmental features that underlie lexical stress (e.g., changes in frequency, intensity and duration across syllables within a word) occurs in INT.[6]

Children with CAS are thought to require years of intensive therapy,[1] reportedly up to 81% more therapy than severe phonological disorders to achieve similar functional outcomes,[9] perhaps due in part to a poor understanding of the disorder mechanism up until recently. Indeed, symptoms can persist into adulthood and result in a substantial disability affecting intelligibility, social communication, academic performance, and overall quality of life.[10,11] Treatment approaches for CAS have primarily targeted improved accuracy of segmental features to expand phonemic inventory or develop a core vocabulary. Although there are a variety of treatment approaches, the efficacy of each has not been established (for a review, see ref. [12]) Consequently, there is a critical need for development and implementation of innovative treatment approaches that effectively target the underlying impairment in CAS and the resulting disruption in temporal control of syllable-level prosody – specifically lexical stress contrasts and

CONTACT Hilary E. Miller hilary.miller48@gmail.com Department of Speech, Language, and Hearing Sciences, Boston University, 677 Beacon St., Boston, MA 02215

*Current Address: Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA, USA.
Supplemental data for this article can be accessed on the publisher's website.

syllable segmentation – that characterizes the disorder. Mechanistic-based treatment approaches may result in increased treatment efficacy, with faster and more effective outcomes to intervention.

The current study investigates the efficacy of Treatment to Establish Motor Program Organization (TEMPO^{SM}) as an intervention for CAS. TEMPO^{SM} targets the hypothesized underlying impairment in CAS (i.e., segmentation of speech into individual sound/syllable units) by training concatenation of syllables into longer motor units. Additionally, TEMPO^{SM} explicitly and simultaneously targets each of the three diagnostic features of CAS through repeated practice of fluent transitions between syllables, syllable stress contrasts, and accurate speech sounds in multisyllabic pseudowords for improved intelligibility and naturalness of speech. TEMPO^{SM} is structured within a motor learning framework and fully incorporates principles of motor learning (PML) in order to maximize treatment efficacy. Broadly, PML delineate important factors in stimulus selection, practice structure, and feedback that increase the difficulty of training tasks and encourage self-evaluation of productions in order to promote retention and generalization of trained motor skills (see Methods for a complete list).[13,14] Robin and colleagues first demonstrated a significant improvement in acoustic and perceptual measures of lexical stress in three children with CAS following treatment using this novel approach, now named TEMPO^{SM}.[15] Subsequent studies of this approach, published using the name Rapid Syllable Transition Treatment (ReST),[16–20] have used exclusively demonstrated positive treatment effects in specific perceptual measures of articulatory and lexical stress accuracy, including in one randomized controlled trial of children with CAS.[17] A smaller single-case design study demonstrated similar perceptual improvements in four children with CAS who received lower-dose treatment frequency.[18] Telehealth delivery of this intervention also resulted in positive gains in perceptual measures of production accuracy.[19,20]

Perceptual measures are the current standard for clinical diagnosis and treatment of CAS but have limitations in comparison with acoustic measures, which have been shown to be reliable, objective, and sensitive measures of features of motor speech disorders.[21–23] Specifically, acoustic measures of the speech of children with CAS show evidence of a disruption in temporal control of speech, marked by increased duration and reduced variability in duration of speech segments.[15,24–29] Ballard et al. provided strong preliminary evidence as to the utility of acoustic measures as sensitive indices of treatment-related change in CAS in the first single-case design study of TEMPO^{SM},[15] but there is otherwise a lack of acoustic evidence as to the efficacy of CAS treatment. Therefore, the application of these durational acoustic measures in a larger group study is a critical next step to advance our understanding of the specific mechanisms through which TEMPO^{SM} improves prosody in CAS. The current study aims to extend previous work to include an additional acoustic measure (intersegment duration) to provide a more precise measure of changes in segmentation of speech following TEMPO^{SM} intervention. Additionally, this study aims to validate the acoustic measures of lexical stress previously applied by Ballard et al.[15] (increased durational, pitch, and volume contrasts) in a larger group of children with CAS. Perceptual measures of segmentation of speech, lexical stress, and speech sound accuracy are also included as secondary outcome measures to parallel the measures used in previous studies of this treatment approach and the provision of perceptually based feedback during TEMPO^{SM} clinical sessions.

The primary hypotheses are:

(1) Children will demonstrate significantly improved performance in acoustic and perceptual measures of segmentation, lexical stress, and consonant distortions as a result of treatment of trained pseudowords comprised of plosive phonemes, with retention of treatment effects one-month post-treatment.
(2) Treatment effects will generalize to untreated related exemplars, both real words and pseudowords comprised of plosive phonemes, with maintenance up to one-month post-treatment.
(3) Improvements in perceptual speech sound accuracy will not generalize to pseudowords that are untreated and unrelated to the trained set; that is, pseudowords comprised of fricative phonemes.
(4) Acoustic measures of intersegment duration and lexical stress contrast will be significantly correlated with perceptual measures of segmentation and stress accuracy.

Method

This study was approved by the Institutional Review Board for the Protection of Human Subjects in Research at the University of New Hampshire and written informed consent was obtained from both parents and the participating children.

Participants

Children were referred to the clinic by their current treating speech-language pathologists in response to a recruitment advertisement. Families first completed a short telephone screen to collect a case history and schedule an evaluation to confirm eligibility criteria for participation in the study: a) normal hearing; b) no orofacial structural abnormalities, muscle weakness, or altered muscle tone or reflexes; c) native speaker of English, d) no other developmental, neurological, or genetic disorders, and e) diagnosis of CAS. Formal clinical evaluation was then completed by graduate student clinicians under the supervision of qualified clinicians (authors DAR and ASP). Assessments included a pure-tone hearing screening to ensure normal hearing at time of study entry;[30] Clinical Evaluation of Language Fundamentals – Fifth Edition (CELF-5) to assess language skills;[31] Goldman – Fristoe Test of Articulation – Third Edition (GFTA-3) to document speech errors;[32] and a structural and functional oral and speech motor assessment (i.e., Motor Speech Examination, MSE) to determine presence of CAS and rule out structural abnormality or frank dysarthria.[33] The oral structure examination confirmed no evidence of asymmetry, slowness, weakness, reduced movement amplitude, or altered muscle condition (e.g., flaccidity or spasticity). Children deemed eligible for the study also demonstrated no evidence of impaired respiration, phonation, or resonance during the examination.

Expert clinician judgment of the presence/absence of key discriminative perceptual features remains the current gold standard in diagnosis of CAS, as agreed-upon thresholds of presences for key CAS features have yet to be established.[34] Evidence for segmentation, equal stress, and speech sound distortions, which are the three commonly cited key perceptual features in both acquired and childhood apraxia of speech,[1–3,34,35] was assessed during the polysyllable speech tasks of the MSE (e.g. production of words of increasing length, multiple repetitions of three-syllable words, sentences containing multisyllabic words, and conversational speech).[33] This is consistent with recent studies that have shown polysyllable words to be sensitive to these apraxic features.[3,34] Diagnosis of CAS required unanimous independent agreement across two experienced speech-language pathologists (authors ASP and DAR, both certified SLPs with more than 20 years of clinical experience with CAS) as to the presence, in at least three tasks (word-level or above) of the MSE, of all three features of the disorder: 1) segmentation, defined as increased segment and intersegment durations, particularly the perception of syllable segregation and inter-syllabic pauses; 2) equal or reduced stress contrast across syllables in multisyllabic words or phrases; and 3) speech sound distortions, defined as phonetic errors in the production of a phoneme, including voicing distortions, vowel distortions, and distorted substitutions. Clean phonemic substitution errors (i.e. produced without distortion) were not counted as evidence for CAS. While there is no commonly agreed upon operational definition of "presence" of an apraxic feature,[34] we applied the descriptions of Strand et al. from the Apraxia of Speech Rating Scale; that is, a feature was considered present in any given task if a rater judged that it was either "detectable but infrequent" (i.e., this could be a single and clear occurrence in a task), "frequent but not pervasive", or "nearly always evident" (see ref.[36]). To ensure that a child did not receive the diagnosis based on a single occurrence of each feature, we required unanimous agreement that all three features were observed across at least three polysyllabic speaking tasks.

Figure 1 shows participant flow through the study in accordance with TREND guidelines.[37] Of 22 families who completed the preliminary screen, 16 completed evaluations in the clinic. The remaining six families decided not to participate and did not schedule an evaluation. Three children were excluded following evaluation due to a diagnosis of phonological disorder with no evidence of motor programming impairment. Of the 13 children who met study criteria with a confirmed diagnosis of CAS, one family declined to participate in the study, and the remaining 12 children completed the four-week intervention. One of the 12 children was excluded from the group analyses reported here, as behavior prevented collection of some post-treatment measures; individual data for this participant is included in Supplemental Materials.

The 11 participating children ranged in age from 5;10 to 8;4 years;months, as of the first day of treatment (M = 7;1, SD = 0.7 years). Age and sex of each participant, as well as scores from administration of the Clinical Evaluation of Language Fundamentals – Fifth Edition[31] (CELF-5) and Goldman – Fristoe Test of Articulation – Third Edition[32] (GFTA-3) are presented in Table 1. All were enrolled in speech therapy for two or more years up until the start of this study and suspended community-based therapy until completion of the study (i.e. from the first baseline probe until the completion of the one-month retention phase).

Experimental Design

This phase II small-group cohort-control study employed a mixed between- and within-subjects design, with multiple baselines across participants and behaviors; the range of baseline length was varied sufficiently to allow for use of a waiting control group (for example, see ref. [38]) Participants were pseudo-randomly assigned to either the immediate treatment or waiting control group to allow for constraints due to family availability during the treatment periods. Participants in the

Table 1. Participant characteristics.

Participant	Age (year; month)	Sex	Treatment Level (syllables)	Number of Baselines	Group Assignment	GFTA	Receptive Index Score	Expressive Index Score	Core Language Score	Language Content Index
1	6;7	F	3	5	2	58	100	106	102	100
3	6;11	F	3	5	2	57	94	85	93	84
4	6;11	M	3	3	1	40	73	69	70	72
6	5;10	M	3	3	1	42	100	89	96	96
7	7;8	M	3	3	1	60	69	70	66	78
8	8;4	F	3	2	1	40	102	90	87	110
11	8;1	M	4	3	1	94	78	83	80	88
12	7;0	M	3	5	2	60	80	89	87	80
13	6;4	M	4	3	1	61	102	104	102	110
15	7;5	M	3	6	2	65	89	87	86	98
16	7;1	M	4	4	2	62	104	89	93	100
Mean	7;1		3.3	3.8		58	90.1	87.4	87.5	92.4
SD	0;9		0.5	1.3		17	12.9	11.4	11.8	12.8

For Group Assignment, 1 = Immediate Treatment group, 2 = Waiting Control group. GFTA = normed score on Goldman-Fristoe Test of Articulation-Third Edition; the Receptive Index Score, Expressive Index Score, Core Language Score, and Language Content Index are all normed composite scores from the Clinical Evaluation of Language Fundamentals (CELF-5). M = Male; F = Female, SD = standard deviation.

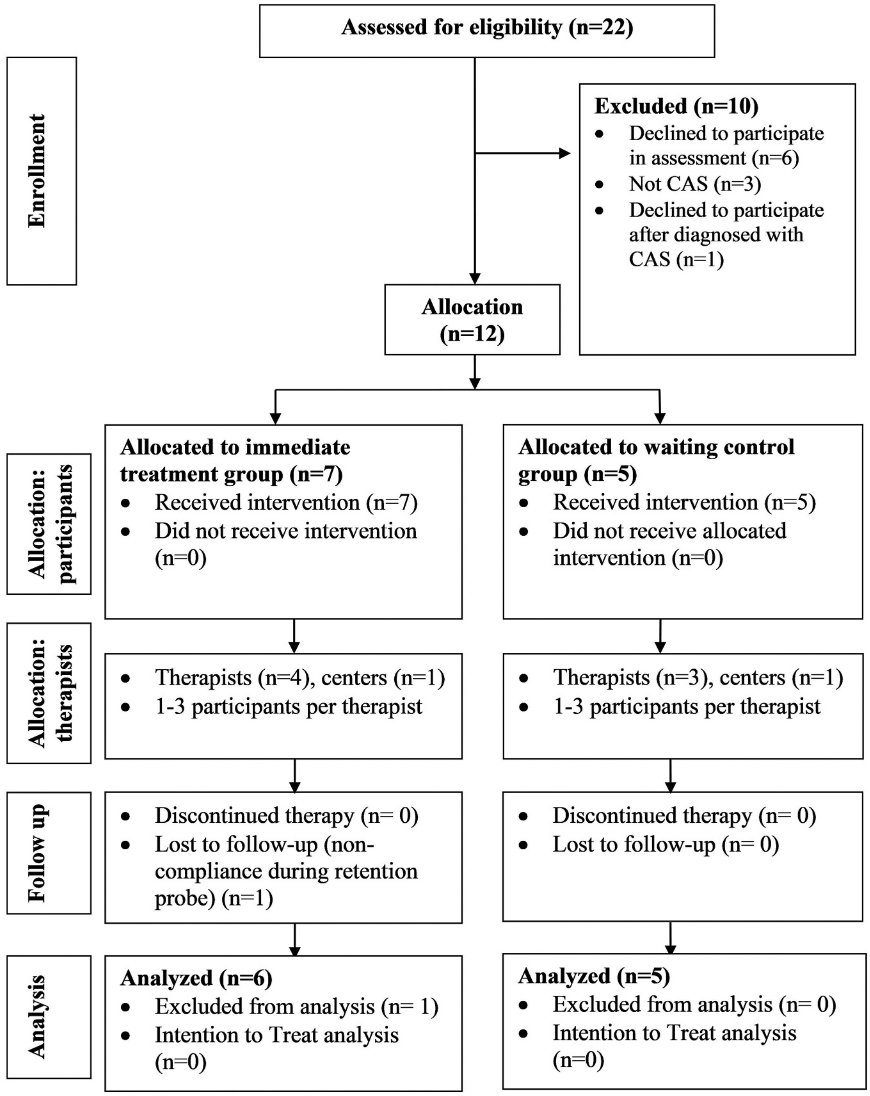

Figure 1. Flow chart of participant enrollment, assignment, and follow-up.

immediate treatment group (i.e. group 1 in Table 1) completed 2–3 baselines in a roughly two-week period while the waiting control group (i.e. group 2 in Table 1) completed 4–6 baseline testing sessions over a period of six weeks. Multiple baseline tests documented level and variability of performance for each child prior to treatment, with staggered baselines across participants used for demonstrating internal validity; that is, testing whether the treated and related behaviors improved significantly only after the administration of the treatment. This waiting control group design was used to establish the effect of the four-week TEMPO^{SM} treatment protocol relative to no treatment; however, the two groups will be reported here as a single group analysis as the waiting control group showed no change in performance on baseline probes administered over the wait-list period. The subsequent entry of the wait list group into the four-week treatment allowed us to double the sample size that was then considered for analysis of changes in speech accuracy for treated and untreated stimulus sets over time. Additional experimental probes were completed immediately post-treatment (*M*= 2.2 days post-treatment, *SD* = 1.8 days), and at one-month post-treatment (*M*= 31.7 days post-treatment, *SD* = 2.3 days) to measure treatment effects, retention, generalization to untreated items, and experimental control.

Stimuli

Treatment stimuli consisted of three- or four-syllable pseudo-words with either a strong-weak (SW) or weak-strong (WS) stress pattern over the first two syllables. Based on performance during the initial evaluation, baseline probes, and pre-practice during the first treatment session, children were assigned to either the three-syllable of four-syllable treatment level (see Table 1) to provide an optimal challenge and functional difficulty level.[39] For the three-syllable stimuli, a list was generated of all 72 possible CVCVCV combinations containing three different plosive consonants (/b/, /t/, and/g/) and three different long vowels (/a/, /i/, /u/) in both SW and WS stress patterns, consistent with Ballard et al. (e.g., TAgibu and taGIbu).[15] A similar list was created for the four-syllable stimuli, again containing plosives (/b/, /t/, and/g/) and vowels (/a/, /i/ , /u/) to generate a comparable list of 72 stimuli (e.g., GItubagi and giTUbagi). From these stimulus lists, 40 syllable strings were randomly selected for treatment (Treated Plosives) in both SW and WS stress conditions (e.g. 20 unique strings, each to be practiced in both SW and WS stress patterns for a total of 40 items). The remaining 32 combinations were left untreated to measure transfer to similar but untreated exemplars in experimental probes (Untreated Plosives). Two additional stimulus sets were created for experimental probes: a Real Words set of three-syllable real words containing the treated sounds (e.g. toboggan, barbecue), and a Fricatives set of three-syllable pseudowords containing fricative sounds (e.g. SHIvasu, shiVAsu).

Baseline and experimental probes were administered by graduate clinicians and contained a total of 120 items, including 20 items (10 SW and 10 WS) randomly selected from each of the four sets above: Treated Plosives, Untreated Plosives, Real Words, and Fricatives. An additional 40 stimuli were selected from other complexity levels; however, performance on these items is not reported here. Stimuli were presented in randomly ordered carrier phrases (e.g., "There's my ____" or "She has a ______."). Four variations of the probe lists were used for experimental probes, with the order counterbalanced and randomized across participants. Each child's treating clinician administered at least one of the child's baseline probes and all subsequent experimental probes.

Treatment

Treatment was administered by trained graduate student clinicians over a four-week treatment period, with four one-hour sessions per week. The treatment period in this current study was extended one week from the original three-week period used in previous studies of this treatment approach.[15–18] Consistent with other motor learning protocols in speech pathology we extended the treatment period to four weeks as this has produced robust learning effects for motor speech disorders (for example, see ref. [40]) and children with CAS are known to require high dosages of therapy.[1,9,41]

Recommendations from the Treatment Fidelity Workgroup of the National Institutes of Health Behavior Change Consortium were implemented to ensure treatment fidelity.[42] Clinicians completed multiple intensive training sessions with first and last authors, including structured practice and role-playing for accurate implementation of the treatment protocol, as documented in the treatment manual. Over 50% of sessions were directly observed by one of three authors to ensure inter-rater reliability. Clinicians also each observed at least one session conducted by another clinician. Any discrepancies with the treatment protocol were addressed during treatment sessions or at frequent supervisory meetings.

Table 2. Principles of motor learning.

Condition	Optimal Motor Learning
Practice amount	Higher number of practice trials > less practice
Practice distribution	Distributed practice > massed
Practice variability	Variable practice on different targets > constant
Practice schedule	Random practice with intermixed targets > blocked
Attentional focus	External focus on effects of movements > internal
Target complexity	Complex sounds and sequences > less complex
Feedback type	Knowledge of results > knowledge of performance
Feedback frequency	Reduced feedback > constant feedback
Feedback timing	Delayed feedback > immediate feedback

Intervention explicitly targeted each of the three features of CAS through repeated productions of multisyllabic pseudowords at a natural speech rate. Correct production was assessed on accuracy across each of the three features of CAS: correct sounds, fluent transitions between syllables, and accurate lexical stress. Twenty stimuli (10 SW and 10 WS) from the Treated Plosive set were randomly selected for each session.

Treatment was structured within a motor learning framework (see Table 2 for a summary of key PML).[13,14] Treatment sessions consisted of (a) Pre-Practice, continuing until the child produced correct responses to five different stimuli with clinician-provided Knowledge of Performance (KP) feedback, modeling, and visual or tactile cues, as necessary; and (b) Practice, consisting of 100 total productions being five productions each of the twenty randomly ordered stimuli. During the Pre-practice stage, clinicians established a reference of correctness for the task, provided explanations of target response accuracy, and ensured motivation and stimulability. The Practice stage adhered to a strict low frequency, delayed feedback schedule with Knowledge of Results (KR) feedback provided on 60% of trials after a 3-second delay, and a 5-second delay between provision of feedback and presentation of the next stimulus. Clinicians used a feedback sheet containing a visual of the three targeted features to refer to each term as they gave KR feedback to the child (e.g. "Nice and smooth [pointing to visual for segmentation], but sounds and stress weren't right [pointing in turn to visuals for sounds and stress]."). Since not all children could read fluently, stimuli were presented auditorily by the clinician, with a three-second delay between the model and the child's production in response to a "Go" signal. This study differed from previous work in that children completed 60 minutes of Pre-practice for the first two sessions and did not begin the Practice stage until the third day of treatment. Pre-practice lasted no more than 15 minutes in subsequent sessions. Per Warren et al. recommendations for standardized descriptions of treatment intensity,[43] the intervention duration in this study was four weeks, the dose frequency was once per day for four days per week, and the treatment dose was 100 practice trials per session, for a cumulative intervention intensity of approximately 1600 trials per child. Children also completed shortened experimental probes immediately prior to the first treatment session in the second, third and fourth week of treatment (data not reported here).

Equipment

All experimental probes and treatment sessions were recorded in a quiet room at 44.1 kHz with Samson XPD1 microphones, positioned 5 cm from child's mouth. Recordings were saved in WAV format.

Dependent Measures

About ten percent of responses (519 of 4980 tokens) in baseline and experimental probes could not be analyzed due to technical issues or extraneous noise (e.g. recording failure, background noise, yelling or laughing during production) or errors in clinician models (e.g. transposing syllables).

Acoustic Measures

Acoustic analyses of pre- and post-treatment Treated Plosive, Untreated Plosive, and Real Word stimuli were completed by trained research assistants using Praat signal-processing software.[44] Segmentation was measured as intersegment duration (ISD), defined as the time from the last glottal pulse, as indicated by the end of F1 and F2 in the wideband spectrogram, to the onset of the plosive burst in the following syllable. For some items, an intersegment duration could not be marked due to either omission or severe distortion of the plosive consonant that made it impossible to distinguish start and end points of segments. In real words containing non-plosive sounds (e.g. bandanna, bicycle), intersegment duration was only measured preceding plosives.

Acoustic measures of lexical stress included duration (ms), maximum fundamental frequency (Hz), and maximum vocal intensity (dB) for the vowels of strong and weak plosive syllables in Treated Plosive, Untreated Plosive, and Real Word stimuli. Vowel duration was measured between the first and last glottal pulse of the vocalic nucleus, as indicated by energy extending through F1 and F2 displayed on the wideband spectrogram, and using fundamental frequency, formant, and intensity contours generated by the Praat software.[15,45] Maximum intensity and fundamental frequency of each vowel were generated automatically using Praat algorithms. The pairwise variability index (PVI) of each variable was calculated using Equation (1) to provide a normalized comparison of the strong and weak syllable in each stimulus:

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

where d is the duration of the k^{th} syllable.[15,46] A higher PVI value reflects increased contrast in lexical stress, whereas a PVI of zero indicates equal stress across syllables. PVI were calculated for duration (PVI(dur)), fundamental frequency (PVI(f_0)), and intensity (PVI(I)).

Real words where either strong or weak syllables did not contain a plosive (e.g. pineapple, tomato) were not included in PVI analyses. In cases where children added an extra syllable, the syllables that best fit the intended stress pattern were analyzed. Stimuli in which the child did not repeat the intended stress target were excluded from analysis of PVI (1% of analyzed files, or 34 of 2902 tokens), as were stimuli where omission or devoicing of the weak syllable prevented measurement of vowel duration (approximately 25% of analyzed stimuli, or 692 of 2902 tokens). Syllable omission rates varied between participants, as well as by stimulus type and time point: weak syllable omission occurred on 32% of WS plosive pseudowords at baseline (range across children was 0–75%) and 16% at post-treatment (range 0–79%), and 15% of SW plosive pseudowords at pre-treatment (range 0–64%) and 1% at post-treatment (range 0–8%). No strong syllables were omitted, and omission rates were lower for real words than pseudowords. Omission rates for each participant are available in Supplementary Materials.

Perceptual Measures

Perceptual measures of accuracy on each of the three features of CAS (segmentation, equal stress, and speech sound distortions) in tokens from pre- and post-treatment experimental probes were scored by a group of eight trained graduate students and clinical faculty. Segmentation was scored based on the perception of increased segment or intersegment durations, or inter-syllable pauses. Accuracy on this criterion required fluent transitions between all syllables with the child matching the clinician's production of the item, without perception of hesitation or pause. For scoring of lexical stress accuracy, productions were scored as incorrect for equal or reduced stress contrasts across syllables. Consistent with the acoustic analyses, scorers did not score this variable if the child omitted one of the first two syllables in a stimulus or if the child did not repeat the intended stress target for an item (i.e. produced a WS pseudoword instead of SW; again less than 1% of total files).

Speech sound accuracy was scored based on the presence or absence of consonant distortions – plosive sounds in Plosive and Real Word stimulus sets and fricative sounds in the Fricative stimulus set. Transposition errors were not counted as distortions if the substituted consonant was produced without distortion. Distorted substitutions were scored as errors (e.g., nasalization or spirantization of plosives). This measure differed from previous work (e.g. ref. [17]) who included vowel production in their scoring of speech-sound accuracy. This change was made to explicitly compare plosive accuracy to the untreated fricative phonemes included for experimental control, although feedback on vowel accuracy was provided during intervention.

The eight perceptual scorers were all native English speakers with little prior experience with CAS. Scorers completed two training sessions with first and last authors prior to scoring where they reviewed examples of correct and incorrect productions in each feature and scored sample tokens. Scorers each scored at least one complete data set from one child, with some tokens included twice for intra-rater reliability calculations. All perceptual scorers were blinded to time point of samples and study hypotheses. Order of presentation was randomized within each stimulus set, with Fricative, Real Word, and combined Treated and Untreated Plosive stimulus sets scored separately. Scorers were instructed to listen to each sample twice for each of the three criteria, for a total of six times for each item, using personal headphones in a quiet room. In cases where a child added an extra syllable, scorers were instructed to score the syllables that best fit the target (e.g., score only the last three syllables if "giTAgibu" was produced for "TAgibu").

Reliability

For acoustic measures, a randomly selected ~15% of each scorer's samples were rescored by a second rater to calculate inter-rater reliability. Reliability was high for inter-rater comparison of both intersegment duration and vowel duration measures (ICC 0.945 and 0.948 respectively, $p < .001$, absolute agreement on single measures). The average point-to-point differences for intersegment and vowel duration measures were 11.1 ms (SD = 18.5 ms) and 11.9 ms (SD = 18.7 ms), respectively.

Each rater also re-measured these durations in a randomly selected ~15% of their samples for intra-rater comparisons.

Average differences were 9.6 ms (SD = 19.3 ms) for intersegment duration and 9.5 ms (SD = 18.1 ms) for vowel duration. Intra-class correlation coefficients for the combined nine raters were also high (.955 and .958 respectively, $p < .001$, absolute agreement on single measures). Intra-rater and inter-rater reliability for acoustic durational measures are comparable to those reported in similar studies.[27,47,48]

For perceptual measures, intra-rater reliability was calculated from a random 10% of each scorer's total samples. Point-to-point agreement for intra-rater comparison was 82% for segmentation, 76% for stress, and 85% for distortion count. A second listener also rescored 10% of each scorer's files to calculate inter-rater reliability. Point-to-point agreement on each measure was 77% for segmentation, 67% for stress, and 75% for distortions. Overall, Cohen's κ values demonstrate moderate to substantial intra-rater agreement (Segmentation: $\kappa = .60$, $p < .001$; Stress: $\kappa = .52$, $p < .001$; Distortions: $\kappa = .68$, $p < .001$) and fair to moderate strength of agreement for inter-rater agreement (Segmentation: $\kappa = .40$, $p < .001$; Stress: $\kappa = .34$, $p < .001$; Distortions: $\kappa = .41$, $p < .001$).[49]

Data Analysis

The strength of the association between acoustic and perceptual measures of segmentation and stress was calculated using Spearman correlations. SW and WS perceptual stress measures were collapsed onto a 3-point ordinal scale, where 1 was accurate WS stress, 2 was inaccurate stress production (i.e., equal stress across syllables), and 3 was accurate SW stress. Correlations were conducted for PVI measures of duration, intensity, and frequency. For segmentation, the binary perceptual measure of accuracy was compared with the maximum intersegment duration measure for each production. A significance level of 0.05 was set *a priori* for all statistical analyses.

Descriptive statistics and boxplots were generated for each acoustic variable (see Supplemental Materials), with SW and WS stress patterns treated as separate outcome measures. Based on visual inspection of data distribution followed by manual review of a sample of audio files and spectrograms, outliers greater than two standard deviations outside the mean for each participant were determined to represent uncharacteristic performance (e.g. child speaking with non-habitual voice, presence of background noise) and excluded from further statistical analyses (less than 4% of analyzed data for each variable). Outliers were removed separately for pseudo- and real words, due to significant differences in pre-treatment performance between these stimulus sets. An additional five real word PVI(dur) values (less than 1% of total real word PVI values) were manually removed by visual inspection alone, with consensus between the first and last author. In these cases, the two standard deviation criterion was not used as these cases had fewer than 20 data points due to high degrees of syllable omission. All statistical models were run both with and without outliers to confirm the effect of outlier removal for analytic transparency. Results for the analyses with outliers removed are reported here (see Supplementary Materials for a comparison of analyses with and without outliers).

Generalized linear mixed models (GLMM) with the participant as a random intercept were used to account for the clustering of repeated measures within subjects over time during our analyses of both baseline stability and treatment effect. GLMM offer a flexible statistical approach for assessing longitudinal data, and are accepted as the state-of-the art for modeling complex data structures (e.g., unbalanced designs, repeated measures within subjects).[50–52] This method ensures that standard errors are not underestimated, as can occur when Ordinary Least Squares (OLS) models are applied to clustered data structures.[53] GLMM enable the user to adjust for random or repeated variables and to specify different covariance structures.[54,55] For the acoustic analyses, GLMM with a normal distribution and an identity link function were used (continuous outcome). For perceptual analyses, GLMM with a binomial distribution and a logit link function were used (dichotomous outcome).

For both acoustic and perceptual measures, baseline stability was first assessed using GLMM. No statistically significant differences across baseline time points were observed ($p > .05$) for any outcome measures. This demonstration of baseline stability, whether baselined over two weeks or six weeks, allowed for the pooling of data across all children to generate a single group average for the baseline time-point for the subsequent analyses of treatment effects.

For analyses of treatment effect and generalization, models initially included fixed effects for treatment phase (baseline, immediate post-treatment, and one-month retention) and stimulus set (treated plosive pseudowords, untreated plosive pseudowords, and untreated real words), and their interaction. Interaction terms were removed from final models if they were not statistically significant. Analyses for perceptual measures included calculation of odds ratios (OR), reflecting exponentiated beta coefficients, with 95% confidence intervals. Calculation of odds ratios for dichotomous outcomes is considered best practice in biostatistics, as it facilitates the interpretation of relative risk and offers an intuitive measure of effect size.[56] Post-hoc testing used Bonferroni adjustments for multiple comparisons.

Following the model-building approach recommended by Bryk & Raudenbush[51] and Burnham & Anderson,[57] we ran the models for each outcome with the random intercept for subject, main effects for Phase and Set, and the interaction of Phase*Set. Model fit was assessed using Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC; see ref. [58]) If the model without interactions produced the best fit, then interaction terms were removed from the final model (e.g. for the PVI models, Table 3). Notably, we did not remove Phase and Set terms, even in cases where they were not statistically significant, because we felt that these were essential to the evaluation of treatment effect and generalization and warranted consistent adjustment across models. Cross-level interactions and random slopes are not included in any of the models. An unstructured covariance matrix for random effects was utilized.

Table 3 a.Summary of GLMM for a) intersegment duration b) PVI for duration (dur), fundamental frequency (f0), and intensity (I) in SW stimuli and c) PVI(dur), PVI(f0), and PVI(I) for WS stimuli. Tables show F-tests for main effects and interaction; β-value, standard error, and p-value, for phase (Baseline, Post-treatment, and 1-month Retention) and set (Treated and Untreated pseudoword and Real word) comparisons; **ICC** – the Intraclass Correlation Coefficient denotes the variability accounted for by the "between-subject" factor with respect to the overall variability in the model; and fit statistics: **AIC** – Akaike's Information Criterion – An estimator of the of out-of-sample prediction error and the relative quality of statistical models for a given set of data. In estimating the amount of information lost by a model, AIC deals with the trade-off between the fit of the model and the simplicity of the model; **AICC**- The AIC Corrected can be used for smaller sample sizes; **BIC** – Bayesian Information Criterion – A criterion for model selection among a finite set of models. The model with a lower BIC is considered better. When fitting models, it is possible to increase the likelihood (sometimes reflected in a better fit) by adding parameters, although this may result in overfitting the model. BIC and AIC both attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC.

	Outcome: Intersegment Duration		
Effect	**β**	**SE**	***p***
Intercept	133.78	9.15	<0.001
Phase[1] Post Test-Baseline	-28.54	3.76	<0.001
Phase[1] Retention-Baseline	-32.52	3.84	<0.001
Set[2] Treated-Untreated	1.17	2.81	0.68
Set[2] Treated-Real	-22.59	2.91	<0.001
Interactions[3]: Phase (Post-Test) x Set (Untreated)	-4.56	5.34	0.39
Phase (Post-Test) x Set (Real)	15.03	5.84	0.01
Phase (Retention) x Set (Untreated)	-1.95	5.49	0.72
Phase (Retention) x Set (Real)	12.80	5.96	0.03
F Tests	Phase: $F_{2,2536} = 101.75$, $p<0.001$		
	Set: $F_{2,2534} = 15.15$, $p<0.001$		
	Phase*Set: $F_{4,5234} = 3.78$, $p=0.005$		
ICC	0.177		
Fit Statistics	Model 1- With Interactions (Best Fit):		
	2 Log Likelihood		58904.9
	AIC (Smaller is Better)		58928.9
	AICC (Smaller is Better)		58929.0
	BIC (Smaller is Better)		58933.7
	Null Model Likelihood Ratio Test		
	DF	Pr > ChiSq	Chi-Square
	2	920.04	<.0001
	Model 2- (Without Interactions):		
	2 Log Likelihood		58920.0
	AIC		58936.0
	AICC		58936.0
	BIC		58939.2
	Null Model Likelihood Ratio Test		
	DF	Chi-Square	Pr > ChiSq
	2	918.82	<.0001

Solutions are for Fixed Effects
[1]Baseline Phase is the referent category
[2]Treated Psuedowords Set is the referent category
[3]Baseline Phase- Treated Psuedowords Set is the referent category

Table 3b

	Outcome: PVI(dur) (SW)			**PVI(f_0) (SW)**			**PVI(I) (SW)**		
Effect	**β**	**SE**	***p***	**β**	**SE**	***p***	**β**	**SE**	***p***
Intercept	45.12	5.94	<0.001	6.82	1.75	0.002	4.47	0.72	<0.001
Phase[1] Post Test-Baseline	18.62	2.90	<0.001	-2.05	0.98	0.036	-1.47	0.51	0.004
Phase[1] Retention-Baseline	15.25	3.01	<0.001	-3.51	0.99	<0.001	-1.22	0.52	0.020
Set[2] Treated-Untreated	-1.19	2.74	0.665	-0.38	0.92	0.678	0.24	0.48	0.609
Set[2] Treated-Real	-2.83	2.91	0.332	-0.23	0.97	0.812	2.20	0.51	<0.001
F Tests	Phase: $F_{2,1149} = 26.03$, $p<0.001$			Phase: $F_{2,1129} = 6.87$, $p=0.001$			Phase: $F_{2,1143} = 5.35$, $p=0.005$		
	Set: $F_{2,1147} = 0.47$, $p=0.634$			Set: $F_{2,1127} = 0.09$, $p=0.916$			Set: $F_{2,1141} = 10.79$, $p<0.001$		
ICC	0.165			0.167			0.080		

(Continued)

Table 3 (Continued).

Effect	Outcome: PVI(dur) (SW)			$PVI(f_0)$ (SW)			PVI(I) (SW)		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Fit Statistics	Model 1- Without Interactions (Best Fit) -2 Log Likelihood 11827.5 AIC (Smaller is Better) 11843.5 AICC (Smaller is Better) 11843.6 BIC (Smaller is Better) 11846.7 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 188.78 <0.001 Model 2- With Interactions: -2 Log Likelihood 11825.0 AIC (Smaller is Better) 11849.0 AICC (Smaller is Better) 11849.2 BIC (Smaller is Better) 11853.7			Model 1- Without Interactions (Best Fit): -2 Log Likelihood 9108.2 AIC (Smaller is Better) 9124.2 AICC (Smaller is Better) 9124.3 BIC (Smaller is Better) 9127.4 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 125.48 <0.001 Model 2- With Interactions: -2 Log Likelihood 9104.4 AIC (Smaller is Better) 9128.4 AICC (Smaller is Better) 9128.7 BIC (Smaller is Better) 9133.2			Model 1- Without Interactions (Best Fit): -2 Log Likelihood 7726.0 AIC (Smaller is Better) 7742.0 AICC (Smaller is Better) 7742.1 BIC (Smaller is Better) 7745.2 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 56.06 <0.001 Model 2- With Interactions: -2 Log Likelihood 7724.5 AIC (Smaller is Better) 7748.5 AICC (Smaller is Better) 7748.7 BIC (Smaller is Better) 7753.2		

Solutions are for Fixed Effects
[1]Baseline Phase is the referent category
[2]Treated Psuedowords Set is the referent category
[3]Baseline Phase- Treated Psuedowords Set is the referent category

Table 3c

Effect	Outcome: PVI(dur) (WS)			$PVI(f_0)$ (WS)			PVI(I) (WS)		
	β	SE	*p*	β	SE	*p*	β	SE	*p*
Intercept	-55.60	5.03	<0.001	-5.67	0.98	<.0001	-2.08	0.48	<0.001
Phase[1] Post Test-Baseline	-13.57	3.09	<0.001	2.69	0.92	0.004	2.03	0.53	<0.001
Phase[1] Retention-Baseline	-9.19	3.06	0.003	3.43	0.92	<0.001	0.52	0.53	0.322
Set[2] Treated-Untreated	-5.30	2.79	0.058	0.96	0.84	0.252	-0.21	0.48	0.658
Set[2] Treated-Real	-17.33	3.16	<0.001	3.71	0.94	<0.001	-0.73	0.54	0.179
F Tests	Phase: $F_{2,906} = 11.20$, $p<0.001$ Set: $F_{2,905} = 15.21$, $p<0.001$			Phase: $F_{2,906} = 8.74$, $p<0.001$ Set: $F_{2,905} = 8.04$, $p<0.001$			Phase: $F_{2,915} = 7.50$, $p<0.001$ Set: $F_{2,914} = 0.92$, $p=0.401$		
ICC	0.133			0.047			0.018		
Fit Statistics	Model 1- Without Interactions: -2 Log Likelihood 9217.6 AIC (Smaller is Better) 9233.6 AICC (Smaller is Better) 9233.8 BIC (Smaller is Better) 9236.8 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 121.69 <0.001 Model 2- With Interactions -2 Log Likelihood 9214.7 AIC (Smaller is Better) 9238.7 AICC (Smaller is Better) 9239.0 BIC (Smaller is Better) 9243.5			Model 1 – Without Interactions: 2 Log Likelihood 6968.6 AIC (Smaller is Better) 6984.6 AICC (Smaller is Better) 6984.8 BIC (Smaller is Better) 6987.8 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 27.34 <0.001 Model 2- With Interactions -2 Log Likelihood 6962.5 AIC (Smaller is Better) 6986.5 AICC (Smaller is Better) 6986.9 BIC (Smaller is Better) 6991.3			Model 1 – Without Interactions: -2 Log Likelihood 6013.5 AIC (Smaller is Better) 6029.5 AICC (Smaller is Better) 6029.7 BIC (Smaller is Better) 6032.7 Null Model Likelihood Ratio Test DF Chi-Square Pr > ChiSq 2 6.17 0.046 Model 2- With Interactions -2 Log Likelihood 6013.3 AIC (Smaller is Better) 6037.3 AICC (Smaller is Better) 6037.7 BIC (Smaller is Better) 6042.1		

Solutions are for Fixed Effects
[1]Baseline Phase is the referent category
[2]Treated Psuedowords Set is the referent category

Results

Segmentation

Acoustic Measures

For acoustic measures of intersegment duration, Generalized Linear Mixed Models (GLMM) with random intercepts for subject revealed a significant treatment phase effect ($F(2, 5243) = 101.92$, $p< .001$), stimulus set effect ($F(2, 5242) = 15.17$, $p< .001$), and treatment phase by set interaction ($F(4, 5242) = 3.78$, $p= .005$). Treatment phase and stimulus set comparisons are summarized in Table 3, with the sign preceding the β-value indicating the direction of the effects (e.g. a negative β-value indicates a decrease in intersegment duration between indicated treatment phases, or the desired treatment effect). Adjusted means for each stimulus set across the three treatment phases are also plotted in Figure 2.

Bonferroni post-hoc comparisons revealed significant decreases in intersegment duration between baseline and both post-treatment time-points, controlling for set ($p< .001$). The significant treatment phase by set interaction was driven by significant differences between pseudoword and real word sets at baseline ($p< .001$); however, within each stimulus set, decreases in intersegment duration between baseline and retention were significant ($p< .001$), indicating maintenance of treatment effects across the retention period for treated and untreated pseudo- and real words.

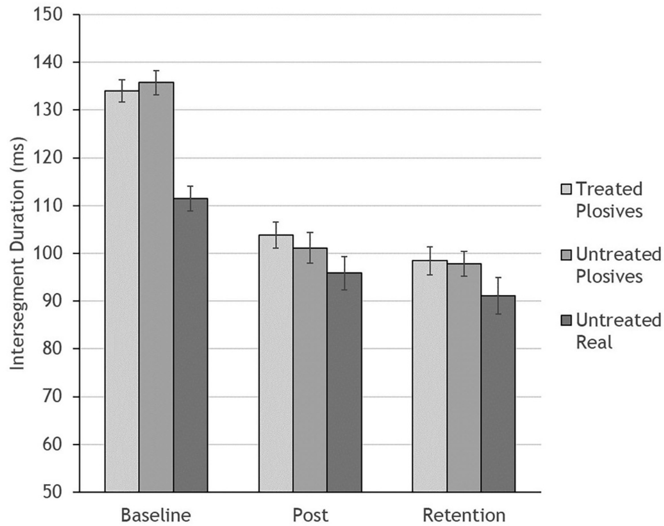

Figure 2. Mean intersegment duration for the combined group of 11 participants, with each of three stimulus sets (treated plosive pseudowords, untreated plosive pseudowords, and untreated real words) graphed separately across baseline, immediate post-treatment (Post), and one-month retention time-points. Errors bars show standard error.

Table 4. Summary of perceptual outcome measures. Odds ratios, 95% confidence intervals, standard error, and *p*-values are shown for each phase comparison (Baseline, Post-treatment, and 1-month Retention) and set comparison (Treated and Untreated Plosives and Real Word).

	Phase							Set						
Outcome Measure	Post-Baseline			Retention-Baseline			F Statistics	Untreated-Treated			Real-Treated			F Statistics
	OR (CI)	SE	*p*	OR (CI)	SE	*P*		OR (CI)	SE	*p*	OR (CI)	SE	*p*	
Segmentation	1.74 (1.41, 2.13)	0.10	***<.001	1.99 (1.61, 2.44)	0.11	***<.001	$F_{2,3351} = 27.65$, ***$p< .001$	1.00 (0.82, 1.23)	0.10	.99	1.65 (1.37, 2.00)	0.10	***<.001	$F_{2,3351} = 18.13$, ***$p< .001$
Stress (SW)	1.51 (1.11, 2.06)	0.16	**.009	1.85 (1.34, 2.55)	0.16	***<.001	$F_{2,1510} = 8.29$, ***$p< .001$	0.87 (0.64, 1.18)	0.15	.36	1.11 (0.83, 1.48)	0.15	.49	$F_{2,1510} = 1.33$, $p = .27$
Stress (WS)	1.77 (1.27, 2.45)	0.17	***<.001	1.84 (1.33, 2.53)	0.16	***<.001	$F_{2,1363} = 10.00$, ***$p< .001$	1.09 (0.80, 1.50)	0.16	.57	1.24 (0.92, 1.66)	0.15	.15	$F_{2,1363} = 1.06$, $p = .35$
Distortions (Plosive)	2.17 (1.76, 2.68)	0.11	***<.001	1.94 (1.57, 2.40)	0.11	***<.001	$F_{2,3346} = 35.09$, ***$p< .001$	0.95 (0.77, 1.16)	0.10	.62	2.27 (1.87, 2.76)	0.10	***<.001	$F_{2,3346} = 49.58$, ***$p< .001$
Distortions (Fricative)	1.03 (0.65, 1.63)	0.24	.92	1.43 (0.92, 2.24)	0.23	.12	$F_{2,1069} = 1.27$, $p= .28$							

* ($p< .05$), ** ($p< .01$), *** ($p< .001$); Post = immediate post-treatment, Retention = one-month post-treatment, Treated = treated plosive pseudowords, Untreated = untreated plosive pseudowords, Real = untreated real words, OR = odds ratio, CI = 95% confidence interval, SE = standard error, SW = strong-weak stress pattern, WS = weak-strong stress pattern.

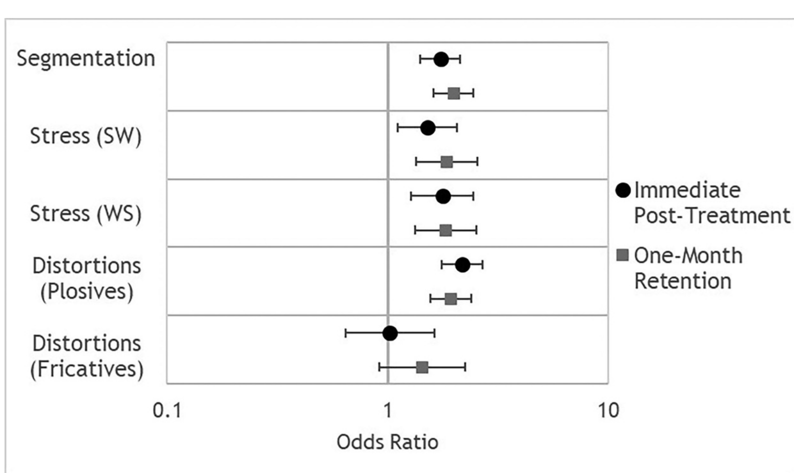

Figure 3. Odds ratios for combined 11 participants for each perceptual variable, showing the likelihood of accuracy in each variable at immediate post-treatment and one-month retention time points as compared to performance at baseline. Error bars show 95% confidence intervals.

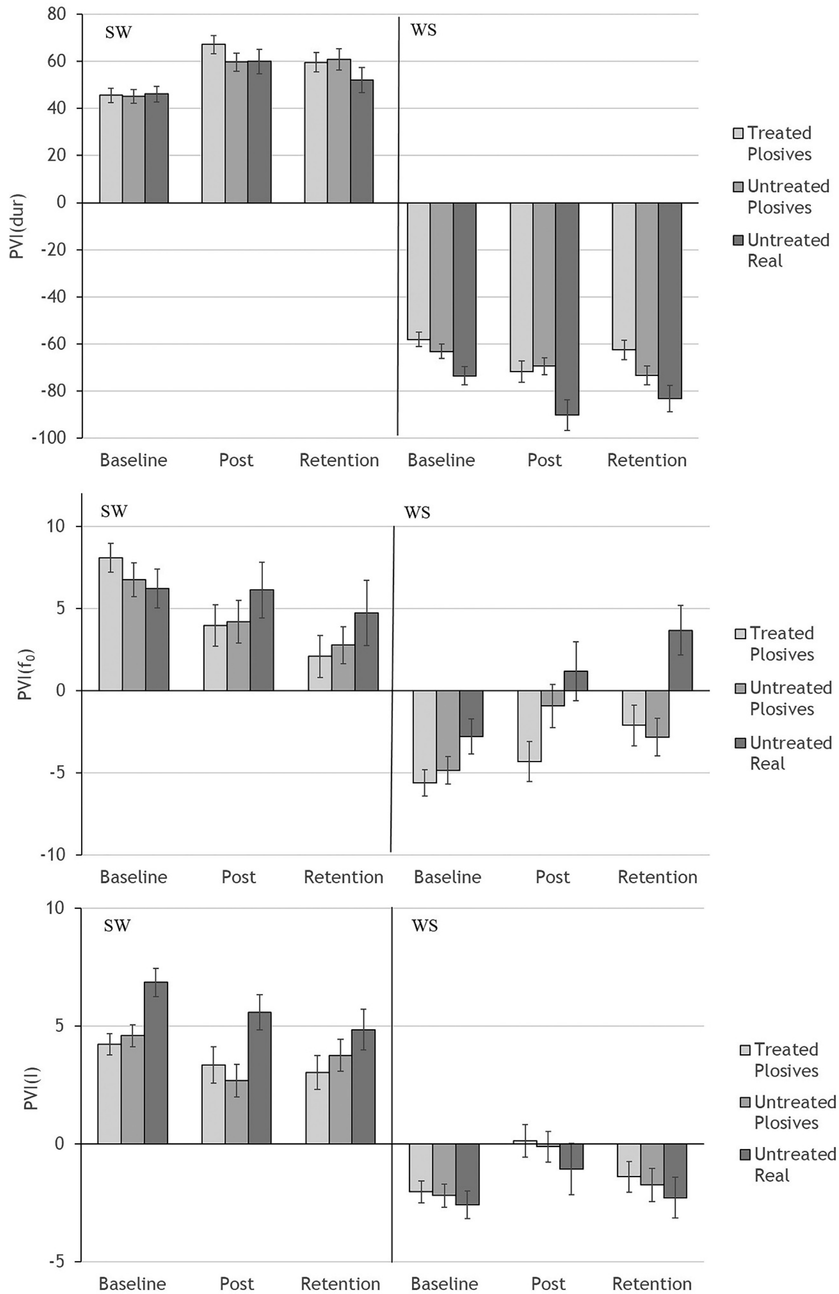

Figure 4. Mean pairwise variability indices of strong and weak vowel duration contrasts (PVI(dur), top panel), maximum fundamental frequency contrasts (PVI(f_0), middle panel), and maximum intensity contrasts (PVI(I), bottom panel), for the combined group of 11 participants. Each stimulus set (treated plosive pseudowords, untreated plosive pseudowords, and untreated real words) is graphed separately across baseline, immediate post-treatment (Post), and one-month retention time points. Means for strong-weak (SW; positive values) and weak-strong (WS; negative values) stimuli are graphed separately for each variable, with error bars showing standard error.

Treatment phase and stimulus set comparisons are summarized in Table 3, with the sign preceding the β-value indicating the direction of the effects (e.g. a negative β-value indicates a decrease between indicated treatment phases). Adjusted group means for each stimulus set across the three treatment phases are also shown in Figure 2.

Perceptual Measures

GLMM with random intercepts for subject revealed a significant phase effect for perceptual measures of segmentation ($F(2, 3351) = 27.65$, $p < .001$), with odds ratios (OR) of 1.985 (95% CI 1.615–2.440, $p < .001$) at retention. This illustrates that, controlling for stimulus set, participants at retention were roughly twice as likely to show favorable outcome scores (i.e. produce stimuli with fluent transitions between syllables) as compared to baseline. There was also a significant effect for stimulus set ($F(2, 3351) = 18.13$, $p < .001$), driven by more accurate performance on real words (OR of 1.654, 95% CI 1.365–2.004, $p < .001$, compared to treated pseudowords as referent). Odds ratios for performance on all perceptual variables at each post-treatment time-point, as compared to baseline performance, are summarized in Table 4. Odds ratios are also presented visually in Figure 3. An OR greater than 1.0 (with a 95% CI that does not overlap 1.0) indicates a statistically significant increase in the outcome relative to baseline, while an OR less than 1.0 (with a 95% CI that does not overlap 1.0) indicates a statistically significant decrease in the outcome measure relative to baseline.

Lexical Stress

Durational Contrast Acoustic Measures

GLMM revealed a significant treatment phase effect for PVI(dur) for both SW and WS stimuli (SW: $F(2, 1149) = 26.03$, $p < .001$; WS: $F(2, 906) = 11.20$, $p < .001$). There was a significant effect for stimulus set for WS stimuli only (SW: $F(2, 1147) = 0.47$, $p = .63$; WS: $F(2, 905) = 15.21$, $p < .001$). Bonferroni post-hoc comparisons revealed significant improvements in durational contrasts (PVI(dur)) between baseline and post-treatment time points, controlling for set (post-treatment: $p < .001$ for both stress patterns; retention: SW, $p < .001$; WS, $p = .008$). There were no significant differences between post-treatment and retention time points (SW: $p = 1.00$; WS: $p = .68$), indicating maintenace of treatment effects across the retention period. Treatment phase and stimulus set comparisons are summarized in Table 3, for SW and WS stimuli respectively. Again, the sign preceding the β-value indicating the direction of the effects (e.g. a positive β-value indicates a more positive PVI following treatment, the desired treatment effect for SW stimuli; a negative β-value indicates a more negative PVI, or the desired treatment effect for WS stimuli). Adjusted means for PVI(dur) for both SW and WS stimuli are plotted in Figure 4.

Acoustic Measures of Pitch and Volume Contrasts

Treatment phase was also significant for PVI(f_0) and PVI(I) variables for both stress patterns, although counter to the hypothesized direction (i.e. children demonstrated decreased pitch and volume contrasts post-treatment, see Figure 4; PVI(f_0) SW: $F(2, 1129) = 6.87$, $p = .001$; PVI(I) SW: $F(2, 1143) = 5.35$, $p = .005$; PVI(f_0) WS: $F(2, 906) = 8.74$, $p < .001$; PVI(I) WS: $F(2, 915) = 7.50$, $p = .001$). Stimulus set was significant for PVI(I) SW and PVI(f_0) WS variables (PVI(I) SW: $F(2, 1141) = 10.79$, $p < .001$; PVI(f_0) WS: $F(2, 905) = 8.04$, $p < .001$). Treatment phase and stimulus set comparisons for pitch and intensity contrasts are summarized in Table 3. Reanalysis of outcome measures with and without outliers revealed some changes [in post-hoc comparisons] for PVI(f_0), which are detailed in Supplemental Materials.

Perceptual Measures of Stress

GLMM revealed a significant treatment phase effect for perceptual stress accuracy for both SW and WS patterns (SW: $F(2, 1510) = 8.29$, $p < .001$; WS: $F(2, 1363) = 10.00$, $p < .001$), with OR of 1.847 (95% CI 1.337–2.552, $p < .001$) and 1.835 (95% CI 1.331–2.531, $p < .001$) at one-month retention timepoint for SW and WS stress patterns, respectively. This illustrates that, controlling for stimulus set, participants were significantly more likely to produce stimuli with accurate lexical stress at retention than at baseline. ORs for both immediate post-treatment and one-month retention time points, as compared to baseline, are again summarized in Table 4 and shown visually in Figure 3. Stimulus set was not significant for either stress pattern (SW: $F(2, 1510) = 1.33$, $p = .27$; WS: $F(2, 1363) = 1.06$, $p = .35$), indicating no significant differences in performance on treated and untreated stimulus sets.

Distortions

Perceptual Measures

GLMM revealed a significant treatment phase effect for perceptual measures of distorted plosive sounds ($F(2, 3346) = 35.09$, $p < .001$) with OR of 1.940 (95% CI 1.567–2.401, $p < .001$), indicating that subjects were roughly twice as likely to produce stimuli with no plosive distortions at the one-month retention timepoint compared to baseline. There was also a significant effect for set ($F(2, 3346) = 49.58$, $p < .001$). OR calculations revealed this difference was due to higher accuracy on real words as compared to the treated psuedoword stimuli, with OR of 2.272 (95% CI 1.872-2.758, $p < .001$).

A separate GLMM for perceptual measures of distortions in the fricative stimulus set revealed no significant phase effect ($F(2, 1069) = 1.27$, $p = .28$), with OR of 1.431 (95% CI 0.915–-2.239, $p = .12$) at the one-month retention timepoint. This indicates that there was no significant change in number of distorted fricative pseudwords across treatment phases, or that children were as likely to produce fricative distortions at baseline and post-treatment timepoints. OR for both plosive and fricative sounds are plotted in Figure 3 and summarized in Table 4.

Acoustic-Perceptual Correlations

Spearman correlation analysis revealed a significant correlation between acoustic and perceptual measures of segmentation ($r_s = -0.445$, $p < .001$). The correlations between acoustic (i.e., PVI) and perceptual measures of stress accuracy were also moderate to strong and highly significant (PVI(dur): $r_s = 0.702$, $p < .001$; PVI(f_0): $r_s = 0.280$, $p < .001$; PVI(I): $r_s = 0.335$, $p < .001$).

Discussion

This study investigated the efficacy of TEMPO^{SM} in improving acoustic measures of segmentation and stress contrasts in a group of eleven children with CAS. Critically, participants demonstrated stable performance across baselines on all acoustic and perceptual outcome measures, which allowed for examination of treatment effects. Results replicate previously reported treatment effects, with evidence of generalization and retention of improvements across both acoustic and perceptual variables. This work adds to the growing literature on the positive effects of this treatment approach in remediation of the primary features of CAS,[15–18] and provides additional support for the application of PML in apraxia of speech intervention (for a review, see ref. [59–62]). Details of principles critical to successful learning (i.e., generalization and retention) are summarized in the methods and include intensive, high-frequency, randomized practice and delayed, reduced frequency, knowledge of results feedback.

The novel use of an acoustic measure of intersegment duration substantiates the efficacy of this intervention in reducing segmentation of speech, a prominent feature of CAS. The acoustic measures of segmentation and stress used in this study, both of which were significantly correlated with perceptual ratings, provide a more sensitive measure of change than binary perceptual measures of prosodic accuracy. Changes in acoustic measures can precede changes in perceptual measures and capture smaller increments of improvement than may be perceptually measurable. In addition, acoustic measures provide more fine-tuned detail that allows for provision of an underlying explanation of perceptual measures. Therefore, these improvements in specific acoustic variables provide strong evidence to support the efficacy of TEMPO^{SM} as an intervention for CAS. Ultimately, the development and clinical implementation of semi-automated acoustic measurements will allow for more accurate, sensitive measures of the speech of individuals with motor speech disorders in future clinical practice.

This study was also unique in the inclusion of four children with below average receptive language scores (range 69–80 Receptive Language Index, CELF-5). These participants all demonstrated substantial improvements across the three features of the disorder, with generalization and maintenance of treatment effects (see Supplemental Material for individual data). Evidence that children with language impairments can successfully engage in an intensive speech motor programming treatment suggests a broad applicability of TEMPO^{SM} to a large number of children with co-occurring CAS. It may be that the use of nonsense syllables minimizes the demands placed on the language system.

Improvements in Segmentation

Acoustic measures of intersegment duration showed reduced segmentation following treatment, consistent with our finding of perceptually "smoother" speech post-treatment with fluent transitions between syllables. Combined, these acoustic and perceptual findings demonstrate that TEMPO^{SM} improves speech motor programming skill in children with CAS. Children with CAS have a deficit in the programming and storage of motor units in INT prior to speech initiation (see ref. [6]); consequently, speech production for these children is limited as they can only hold shorter and less complex motor programs in this impaired buffer prior to execution, resulting in segmentation of speech. The use of pseudowords in TEMPO^{SM} simulates novel word learning processes to allow for isolated and targeted focus on the motor programming system. Critically, practice of novel syllable strings trains the reorganization of multiple speech units into a single motor program, as it is this concatenation process that is hypothesized to be impaired in apraxia of speech.[6] As speech motor learning occurs during treatment, shorter segments – such as gi, ta, and bu – are combined into a multisyllable unit such as gitabu. Significant decreases in perceptual and acoustic measures of segmentation in both treated and untreated stimuli provide empirical support for the hypothesized mechanism of action of the treatment in establishing more efficient organization of speech motor programs. The observed generalization also provides strong support for the hypothesis that treatments that successfully target the underlying mechanism improve overall speech production, beyond the specific stimuli targeted in treatment sessions.

Improvements in Stress Contrasts

Participants demonstrated significant improvements in stress contrasts for both SW and WS stress patterns, as indexed by acoustic PVI(dur) measures. These improvements were also observed in secondary perceptual measures of stress accuracy. Our findings of stronger correlations between perceptual ratings of stress accuracy and PVI(dur) corroborate previous findings that durational contrasts play a dominant role in the perception of atypical stress production in children with CAS.[15] Unexpectedly, children overall demonstrated significantly reduced contrast for PVI(f_0) and PVI(I) measures following treatment despite significant improvements in perceptual and durational measures of stress. This finding differs from previous work by Ballard et al. where changes in durational contrasts following this intervention were accompanied by improved intensity and pitch contrasts.[15] This discrepancy is likely due to the high degree of variability in the presentation of the disorder across participants, the complex interaction between these three prosodic features in the perception of accurate stress production, and the influence of the specific elicitation task used (see ref. [47] for a discussion of task effects). Children with CAS may compensate for a limited motor programming buffer capacity through strategic tradeoffs in resource allocation, as has been demonstrated in both typical children and adults with acquired apraxia.[63,64] This strategic allocation of resources explains the significant variability in children with CAS, as children may differ in their attention to specific features of speech (e.g. segmental versus prosodic accuracy), or even in their focus on a particular feature in each production of a word. Many of the children in this study appeared to compensate for difficulty controlling durational contrasts with an overreliance on frequency and intensity to mark stress at baseline, compared to typical speakers.[47,65] The decreased pitch and volume contrasts following treatment may reflect the normalization of their stress production to use duration as the primary indicator of stress.

Overall, participants demonstrated more difficulty with production of WS stress contrasts, consistent with later development of this stress pattern in typical English-speaking children. Production of SW patterns is typically mastered around age three, while production of WS stress contrasts continues to develop through at least age eleven.[47,65] The increased difficulty of WS words may reflect a trochaic bias resulting from increased exposure to the SW stress pattern in English or the increased physiological demands required for production of WS stimuli. The WS pattern may be particularly difficult for children with CAS because it requires more dramatic durational contrasts, with reported PVI(dur) for typical adults up to twice that for the SW pattern.[47] In fact, PVI(dur) measures in WS words have been proposed as a diagnostic criterion for acquired apraxia of speech, as this stress pattern is more sensitive to subtle impairments in control of temporal contrasts.[48] This is logical because apraxia, in both childhood and acquired forms, is a disruption in temporal control of speech marked by increased duration and reduced variability in duration of speech segments. Therefore, PVI(dur) in WS words may also be the most sensitive indicator of treatment of prosodic deficits in CAS.

Participants in this study demonstrated considerable difficulty producing all syllables in WS stimuli at baseline, with weak syllable omission in up to 75% of pseudoword stimuli for some participants at baseline (see Supplemental Material for individual data). This mimics the progression, albeit delayed, of typical speakers. Toddlers frequently omit unstressed syllables, with a shift around age three toward equal stress across syllables as weak syllable deletion decreases.[66,67] Further analysis of the frequency of syllable omission might be a better initial indicator of progress for some children since inclusion of the weak syllable is a prerequisite to production of appropriate durational contrasts. Children with CAS likely require more than four weeks of intervention to reach typical performance for the increased durational contrasts necessary for WS stimuli.

Improvements in Speech Sound Accuracy

Perceptual results supported our hypothesis that treatment of plosive pseudowords would result in reduced distortions of plosive sounds in both treated and untreated stimuli, but that these improvements would not generalize to a set of untreated fricative pseudowords. This untreated fricative stimulus set was employed to provide additional experimental control, as previous treatment studies of adults with acquired apraxia of speech have shown generalization within phonemes of same manner of production, but not to phonemes of other manners of production.[68–71] Within the framework of Schema Theory, phonemes of the same manner have been hypothesized to share a single Generalized Motor Program (GMP), which governs the general muscle tension and force pattern for that speech movement; movement parameters are then varied to define the specific muscle groups involved (i.e. the location of articulation).[68,72] In addition to establishing experimental control, the stable performance on the untreated fricative set reported here supports the use of Schema Theory to conceptualize speech motor programming with implications for the selection of treatment stimuli in clinical practice.

Limitations and Future Directions

The acoustic measures reported herein provide evidence of the efficacy of the intervention in remediating specific suprasegmental deficits, with evidence of increased syllable stress contrasts and reduced inter-syllable pauses (segmentation of speech) following intervention. Future work would benefit from an expanded set of outcome measures to include an acoustic measure of speech-sound accuracy, such as voice onset time or vowel formants, to provide sensitive measures of voicing or vowel distortions. An additional limitation of this work is the reliability of perceptual measures, especially judgments of stress accuracy. Future work should include more rigorous training of perceptual scorers and obtain scores for each child from multiple raters to mitigate against inter-rater variability in the perception of these speech features. Additionally, it may be preferable to apply a five-point perceptual scale like that used by Ballard et al.,[15] particularly to improve reliability for those items that fall close to the perceptual boundaries between correct and incorrect productions and capture smaller increments of change following treatment. Regardless, the perceptual results support the corresponding acoustic results, which themselves showed strong inter-rater reliability, and attest to the fact that change in various measures of the children's speech is occurring.

This study was also limited by the relatively small and homogenous group of participants, and therefore, may not be appropriate for all children with CAS. Nevertheless, combined with previous studies, there is now a strong body of work demonstrating the efficacy of this approach in the treatment of CAS. Future steps should include a randomized clinical trial to determine the factors that influence individual response to treatment and establish ideal dosage and treatment intensity to achieve robust generalization to spontaneous connected speech. In addition, future studies should investigate the efficacy of subsequent courses of TEMPO^{SM}, as most children with CAS will require more than a single course of intervention to normalize speech motor programming capabilities.

Another crucial next step in CAS research is the use of neuroimaging to establish the efficacy of treatments in normalizing brain networks in children with CAS. Neuroimaging studies of apraxia have primarily focused on adults with the acquired form of the disorder, but there remains a critical need to establish differences in brain networks in children with CAS and determine the specific neural systems underlying the disorder. Neuroimaging work should also characterize how the treatment induces neural plasticity ultimately allowing for its optimization.

Conclusions

Acoustic measurements demonstrate overall reduced segmentation and improved stress contrasts for eleven children following four weeks of intensive TEMPO^{SM} intervention, with generalization to untreated syllable strings and real words and retention of these treatment effects one-month post-treatment. These results replicate and extend previous studies of this approach and support its efficacy in treatment of CAS.

Acknowledgments

The authors thank the children and families who participated; graduate clinicians Roisin Bermingham, Kathy-Ann Murray, Rachel Platt, Megan Rounds, Emily Schultz, and Savanah Snook; Jill Thorson and Bill Rogers for assistance with programs for acoustic and perceptual analyses; IMPROV Lab research assistants who completed acoustic analyses; and our perceptual scorers.

Disclosure Statement

The authors report no conflicts of interest. Portions of this work were presented at the 2018 Conference on Motor Speech in Savannah, Georgia, and the 2018 American Speech-Language Hearing Association

Convention in Boston, Massachusetts, as well as forming part of the first author's Master's thesis at the University of New Hampshire.

ORCID

Hilary E. Miller http://orcid.org/0000-0002-0244-7119
Kirrie J. Ballard http://orcid.org/0000-0002-9917-5390

References

1. American Speech-Language Hearing Association. Childhood Apraxia of speech [Technical Report]. 2007. www.asha.org/policy .
2. McNeil MR, Robin DA, Schmidt RA. Apraxia of speech: definition and differential diagnosis. In: McNeil MR, editor. Clinical management of sensorimotor speech disorders. New York (NY): Thieme; 2009. p. 249–68.
3. Murray E, McCabe P, Heard R, Ballard KJ. Differential diagnosis of children with suspected childhood apraxia of speech. J Speech Lang Hear Res. 2015;58(1):43–60. doi:10.1044/2014_JSLHR-S-12-0358.
4. Klapp ST. Reaction time analysis of two types of motor preparation for speech articulation: action as a sequence of chunks. J Mot Behav. 2003;35(2):135–50. doi:10.1080/00222890309602129.
5. Klapp ST. Motor response programming during simple and choice reaction time: the role of practice. J Exp Psychol Hum Percept Perform. 1995;21(5):1015–27. doi:10.1037/0096-1523.21.5.1015.
6. Maas E, Robin DA, Wright DL, Ballard KJ. Motor programming in apraxia of speech. Brain Lang. 2008;106(2):107–18. doi:10.1016/j.bandl.2008.03.004.
7. Wright DL, Robin DA, Rhee J, Vaculin A, Jacks A, Guenther FH, Fox PT. Using the self-select paradigm to delineate the nature of speech motor programming. J Speech Lang Hear Res. 2009;52 (3):755–65. doi:10.1044/1092-4388(2009/07-0256).
8. Rogers MA, Storkel HL. Planning speech one syllable at a time: the reduced buffer capacity hypothesis in apraxia of speech. Aphasiology. 1999;13(9–11):793–805. doi:10.1080/026870399401885.
9. Campbell T. Functional treatment outcomes in young children with motor speech disorders. In: Caruso A, Strand E editors. Clinical management of sensorimotor speech disorders in children. New York (NY): Thieme Medical Publishers; 1999. p. 394.
10. Rusiewicz HL, Maize K, Ptakowski T. Parental experiences and perceptions related to childhood apraxia of speech: focus on functional implications. Int J Speech Lang Pathol. 2017;20(5):569–80. doi:10.1080/17549507.2017.1359333.
11. Carrigg B, Baker E, Parry L, Ballard KJ. Persistent speech sound disorder in a 22-year-old male: communication, educational, socio-emotional, and vocational outcomes. Perspect School Based Issues. 2015;16(2):37–49. doi:10.1044/sbi16.2.37.
12. Murray E, McCabe P, Ballard KJ. A systematic review of treatment outcomes for children with childhood apraxia of speech. Am J Speech Lang Pathol. 2014;23(3):486–504. doi:10.1044/2014_AJSLP-13-0035.
13. Schmidt RA, Lee TD. Motor learning and performance: from principles to application. Champaign (IL): Human Kinetics; 2013.
14. Maas E, Robin DA, Austermann Hula SN, Freedman SE, Wulf G, Ballard KJ, Schmidt RA. Principles of motor learning in treatment of motor speech disorders. Am J Speech Lang Pathol. 2008;17 (3):277–98. doi:10.1044/1058-0360(2008/025).
15. Ballard KJ, Robin DA, McCabe P, McDonald J. A treatment for dysprosody in childhood apraxia of speech. J Speech Lang Hear Res. 2010;53(5):1227–45. doi:10.1044/1092-4388(2010/09-0130).
16. McCabe P, Macdonald-d'silva AG, Van Rees LJ, Ballard KJ, Arciuli J. Orthographically sensitive treatment for dysprosody in children with childhood apraxia of speech using ReST intervention. Dev Neurorehabil. 2014;17(2):137–46. doi:10.3109/17518423.2014.906002.
17. Murray E, McCabe P, Ballard KJ. A randomized controlled trial for children with childhood apraxia of speech comparing Rapid Syllable Transition treatment and the Nuffield Dyspraxia Programme–third edition. J Speech Lang Hear Res. 2015;58 (3):669–86. doi:10.1044/2015_JSLHR-S-13-0179.
18. Thomas DC, McCabe P, Ballard KJ. Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech: the effect of lower dose-frequency. J Commun Disord. 2014;51:29–42. doi:10.1016/j.jcomdis.2014.06.004.
19. Thomas DC, McCabe P, Ballard KJ, Lincoln M. Telehealth delivery of Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech. Int J Lang Commun Disord. 2016;51 (6):654–71. doi:10.1111/1460-6984.12238.
20. Thomas DC, McCabe P, Ballard KJ, Bricker-Katz G. Parent experiences of variations in service delivery of Rapid Syllable Transition (ReST) treatment for childhood apraxia of speech. Dev Neurorehabil. 2018;21(6):391–401. doi:10.1080/17518423.2017.1323971.
21. Kent RD, Kim Y-J. Toward an acoustic typology of motor speech disorders. Clin Linguist Phon. 2003;17(6):427–45. doi:10.1080/0269920031000086248.
22. Kent RD, Weismer G, Kent JF, Vorperian HK, Duffy JR. Acoustic studies of dysarthric speech: methods, progress, and potential. J Commun Disord. 1999;32(3):141–85. doi:10.1016/S0021-9924(99)00004-0.
23. Allison KM, Hustad KC. Acoustic predictors of pediatric dysarthria in cerebral palsy. J Speech Lang Hear Res. 2018;61(3):462–78. doi:10.1044/2017_JSLHR-S-16-0414.
24. Nijland L, Maassen B, Van Der Meulen S, Gabreëls F, Kraaimaat FW, Schreuder R. Planning of syllables in children with developmental apraxia of speech. Clin Linguist Phon. 2003;17(1):1–24. doi:10.1080/0269920021000050662.
25. Shriberg LD, Green JR, Campbell TF, McSweeny JL, Scheer AR. A diagnostic marker for childhood apraxia of speech: the coefficient of variation ratio. Clin Linguist Phon. 2003;17(7):575–95. doi:10.1080/0269920031000138141.
26. Munson B, Bjorum EM, Windsor J. Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. J Speech Lang Hear Res. 2003;46(1):189–202. doi:10.1044/1092-4388(2003/015).
27. Shriberg LD, Campbell TF, Karlsson HB, Brown RL, McSweeny JL, Nadler CJ. A diagnostic marker for childhood apraxia of speech: the lexical stress ratio. Clin Linguist Phon. 2003;17(7):549–74. doi:10.1080/0269920031000138123.
28. Velleman SL, Shriberg LD. Metrical analysis of the speech of children with suspected developmental apraxia of speech. J Speech Hear Res. 1999;42(6):1444–60. doi:10.1044/jslhr.4206.1444.
29. Shriberg LD, Aram DM, Kwiatkowski J. Developmental apraxia of speech: III. A Subtype marked by inappropriate stress. J Speech Lang Hear Res. 1997;40(2):313–37. doi:10.1044/jslhr.4002.313.
30. American Speech-Language Hearing Association. Childhood hearing screening [Practice Portal]. https://www.asha.org/practice-portal/professional-issues/childhood-hearing-screening/ .
31. Wiig EH, Semel E, Secord WA. Clinical evaluation of language fundamentals. 5th ed. Bloomington (MN): Pearson; 2013.
32. Goldman R, Fristoe M. Goldman-Fristoe test of articulation. 3rd ed. Bloomington (MN): Pearson; 2015.
33. Duffy JR. Motor speech disorders: substrates, differential diagnosis, and management. St Louis (MO): Elsevier Health Sciences; 2013.
34. Murray E, Iuzzini-Seigel J, Maas E, Terband H, Ballard KJ. Differential diagnosis of childhood apraxia of speech compared to other speech sound disorders: a systematic review. Am J Speech Lang Pathol. 2021;30(1):279–300. doi:10.23641/asha.
35. Shriberg LD, Potter NL, Strand EA. Childhood apraxia of speech in children and adolescents with galactosemia. Paper presented at: American Speech-Language-Hearing Association National Convention; 2009 Nov; New Orleans (LA).
36. Strand EA, Duffy JR, Clark HM, Josephs KA. The apraxia of speech rating scale: a tool for diagnosis and description of apraxia of speech. J Commun Disord. 2014;51:43–50. doi:10.1016/j.jcomdis.2014.06.008.

37. Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–66. doi:10.2105/AJPH.94.3.361.
38. Gebhart C, Erlacher D, Schredl M, Billiard MM. Moderate exercise plus sleep education improves self-reported sleep quality, daytime mood, and vitality in adults with chronic sleep complaints: a waiting list-controlled trial. Sleep Disord. 2011:1–10. doi:10.1155/2011/809312.
39. Guadagnoli MA. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. J Mot Behav. 2004;36(2):212–24. doi:10.3200/JMBR.36.2.212-224.
40. Ramig LO, Countryman S, Thompson LL, Horii Y. Comparison of two forms of intensive speech treatment for Parkinson disease. J Speech Hear Res. 1995;38(6):1232–51. doi:10.1044/jshr.3806.1232.
41. Strand EA, Stoeckel RE, Baas BS. Treatment of severe childhood apraxia of speech: a treatment efficacy study. J Med Speech Lang Pathol. 2006;14:297–307.
42. Bellg AJ, Resnick B, Minicucci DS, Ogedegbe G, Ernst D, Borrelli B, Hecht J, Ory M, Orwig D, Czajkowski S. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium. Health Psychol. 2004;23(5):443–51. doi:10.1037/0278-6133.23.5.443.
43. Warren SF, Fey ME, Yoder PJ. Differential treatment intensity research: a missing link to creating optimally effective communciation interventions. Mental Retardation Dev Disabilities. 2007 July;236:228–36. doi:10.1002/mrdd.
44. Boersma P, Weenink D. Praat 6.0.32 [Computer software]; 2017.
45. Kent RD, Read C. The acoustic analysis of speech. San Diego (CA): Singular; 1992.
46. Low EL, Grabe E, Nolan F. Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. Lang Speech. 2000;43(4):377–401. doi:10.1177/00238309000430040301.
47. Ballard KJ, Djaja D, Arciuli J, James DGH, Van Doorn J. Developmental trajectory for production of prosody: lexical stress contrastivity in children ages 3 to 7 years and in adults. J Speech Lang Hear Res. 2012;55(6):1822–35. doi:10.1044/1092-4388(2012/11-0257).
48. Ballard KJ, Azizi L, Duffy JR, McNeil MR, Halaki M, O'Dwyer N, Layfield C, Scholl D, Vogel AP, Robin DA. A predictive model for diagnosing stroke-related apraxia of speech. Neuropsychologia. 2016;81:129–39. doi:10.1016/j.neuropsychologia.2015.12.010.
49. Landis JR, Koch GG. An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33(2):363–74. doi:10.2307/2529786.
50. Arnau J, Balluerka N, Bono R, Gorostiaga A. General linear mixed model for analyzing longitudinal data in developmental research. Percept Mot Skills. 2010;110(2):547–66. doi:10.2466/PMS.110.2.547-566.
51. Bryk AS, Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Newbury Park (CA): Sage; 2002.
52. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J Educ Behav Stat. 1998;23(4):323–55. doi:10.3102/10769986023004323.
53. Bell BA, Ene M, Smiley W, Schoeneberger JA A multilevel model primer using SA(R) PROC MIXED. Proceedings of the SAS(R) Global Forum 2011 Conference. Denver, CO; 2011.
54. Jones A, Huddleston E. The MIXED procedure. SAS/STAT (R); 2009.
55. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Oliver S. SAS for mixed models. Cary (NC): SAS; 2006.
56. Norton EC, Dowd BE, Maciejewski ML. Odds ratios- Current best practice and use. J Am Med Assoc. 2018;320(1):84–85. doi:10.1001/jama.2017.3091.
57. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33(2):261–304. doi:10.1177/0049124104268644.
58. Müller S, Scealy JL, Welsh AH. Model selection in linear mixed models. Stat Sci. 2013;28(2):135–67. doi:10.1214/12-STS410.
59. Ballard KJ, Wambaugh JL, Duffy JR, Layfield C, Maas E, Mauszycki SC, McNeil MR. Treatment for acquired apraxia of speech: a systematic review of intervention research between 2004 and 2012. Am J Speech Lang Pathol. 2015;24(2):316–37. doi:10.1044/2015_AJSLP-14-0118.
60. Bislick LP, Weir PC, Spencer KA, Kendall D, Yorkston KM. Do principles of motor learning enhance retention and transfer of speech skills? A systematic review. Aphasiology. 2012;26(5):709–28. doi:10.1080/02687038.2012.676888.
61. Murray E, Iuzzini-Seigel J. Efficacious treatment of children with childhood apraxia of speech according to the international classification of functioning, disability and health. Perspect ASHA Spe Interest Groups. 2017;2(2):61–76. doi:10.1044/persp2.sig2.61.
62. Maas E, Gildersleeve-Neumann CE, Jakielski KJ, Stoeckel RE. Motor-based intervention protocols in treatment of Childhood Apraxia of Speech (CAS). Curr Dev Disord Rep. 2014;1(3):197–206. doi:10.1007/s40474-014-0016-4.
63. Clark HM, Robin DA. Generalized motor programme and parameterization accuracy in apraxia of speech and conduction aphasia. Aphasiology. 1998;12(7–8):699–713. doi:10.1080/02687039808249567.
64. Clark HM, Robin DA, McCullagh G, Schmidt RA. Motor control in children and adults during a non-speech oral task. J Speech Lang Hear Res. 2001;44(5):1015–25. doi:10.1044/1092-4388(2001/080.
65. Arciuli J, Ballard KJ. Still not adult-like: lexical stress contrastivity in word productions of eight- to eleven-year-olds. J Child Lang. 2017;44(5):1274–88. doi:10.1017/S0305000916000489.
66. Salidis J, Johnson JS. The production of minimal words: a longitudinal case study of phonological development. Lang Acquis. 1997;6(1):1–36. doi:10.1207/s15327817la0601_1.
67. Young EC. An analysis of young children's ability to produce multisyllabic English nouns. Clin Linguist Phon. 1991;5(4):297–316. doi:10.3109/02699209108985898.
68. Austermann Hula SN, Robin DA, Maas E, Ballard KJ, Schmidt RA. Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. J Speech Lang Hear Res. 2008;51(5):1088–113. doi:10.1044/1092-4388(2008/06-0042).
69. Ballard KJ, Maas E, Robin DA. Treating control of voicing in apraxia of speech with variable practice. Aphasiology. 2007;21(12):1195–217. doi:10.1080/02687030601047858.
70. Rubow RT, Rosenbek JC, Collins M, Longstreth D. Vibrotactile stimulation for intersystemic reorganization in the treatment of apraxia of speech. Arch Phys Med Rehabil. 1982;63:150.
71. Wambaugh JL, West JE, Doyle PJ. Treatment for apraxia of speech: effects of targeting sound groups. Aphasiology. 1998;12(7/8):731–43. doi:10.1080/02687039808249569.
72. Schmidt RA. A schema theory of discrete motor skill learning. Psychol Rev. 1975;82(4):225–60. doi:10.1037/h0076770.