ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

A Randomized Controlled Trial of Treatment Distribution and Biofeedback Effects on Speech Production in School-Age Children With Apraxia of Speech

Jonathan L. Preston,[a] Nicole F. Caballero,[a] Megan C. Leece,[a] Dongliang Wang,[b] Benedette M. Herbst,[a] and Nina R. Benway[a]

[a] Syracuse University, NY [b] Upstate Medical University, Syracuse, NY

ARTICLE INFO

Article History:
Received November 1, 2022
Revision received March 13, 2023
Accepted June 8, 2023

Editor-in-Chief: Cara E. Stepp
Editor: Maria Grigos

https://doi.org/10.1044/2023_JSLHR-22-00622

ABSTRACT

Purpose: This study examines how ultrasound biofeedback and intensive treatment distribution affect speech sound generalization during an evidence-based treatment, Speech Motor Chaining, for children with persisting speech errors associated with childhood apraxia of speech (CAS).
Method: In a 2 × 2 factorial randomized controlled trial, children ages 9–17 years meeting CAS criteria were randomized to receive (a) a distributed treatment (20 sessions twice weekly over 10 weeks) or intensive treatment (20 hr in 5 weeks, with 10 hr in Week 1) and (b) treatment with or without biofeedback. Due to the COVID pandemic, some participants were randomized to distributed/intensive telepractice treatment only. The primary outcome was percent target sounds correct on untreated phrases (i.e., generalization) at the 10-week time point. More than 50,000 narrow phonetic transcriptions were analyzed.
Results: Forty-eight participants completed treatment. Intensive treatment significantly increased generalization at all time points. The effect of biofeedback was significant at 5 weeks from the start of treatment but not significant at the primary 10-week time point. However, when comparing each group immediately after their 20 hr of treatment finished, generalization was significantly greater in intensive over distributed treatment and greater in ultrasound over no-ultrasound treatment (with a significant interaction favoring intensive treatment with ultrasound). Only the advantage of intensive treatment remained significant 5 weeks after groups finished treatment. There was no significant difference between face-to-face and telepractice modalities.
Conclusions: When the number of treatment hours is fixed, an intensive schedule of Speech Motor Chaining facilitated greater improvement than a distributed schedule. Ultrasound biofeedback initially accelerated learning, but the benefits may dissipate as treatment continues or after it ends.

Childhood apraxia of speech (CAS) is a speech sound disorder (SSD) that impacts speech sound accuracy, speech movement consistency, transitions between sounds and syllables, and prosody (American Speech-Language-Hearing Association [ASHA], 2007). There is a range of expression of the characteristics of CAS. The presumed locus of deficit is in planning and programming speech movements (ASHA, 2007), although deficits in phonological knowledge are common as well (e.g., Lewis et al., 2018; Nijland, 2009). As there was limited consensus about the definition or characteristics of CAS before 2007 (ASHA, 2007), studies on CAS treatment options have been limited, and there are few randomized controlled trials (Murray et al., 2014). Historically, many children with

Correspondence to Jonathan L. Preston: jopresto@syr.edu. **Publisher Note:** This article is part of the Special Issue: Selected Papers From the 2022 Apraxia Kids Research Symposium. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

CAS remain in speech therapy for several years (Campbell, 1999), and the speech production impairments for individuals with CAS may persist into adolescence or adulthood for some individuals (Carrigg et al., 2016). For example, Lewis et al. (2021) report that 78.6% of adolescents with CAS have residual speech errors as part of their communication profile. There is a need to explore modifications to treatment that may accelerate speech learning to remediate unresolved speech errors, particularly for school-age children who have been in therapy for many years. Therefore, this study is a randomized controlled trial designed to explore two variations that may affect speech motor learning: treatment distribution and visual biofeedback of the tongue with ultrasound imaging. These variations are tested within a standardized treatment program, Speech Motor Chaining.

Motor-Based Treatment for CAS

As CAS is a motor speech impairment, most evidence-based speech treatments for CAS involve motor-based interventions intending to improve speech motor plans (Murray et al., 2014). Many of these interventions integrate prepractice, practice, and feedback principles that are thought to facilitate motor learning (i.e., principles of motor learning [PMLs]; Maas et al., 2008). In this framework, *motor learning*, which is evidenced by retention of motor skills and generalization of motor skills to untrained contexts, is distinguished from *motor acquisition*, which is evidenced by the performance of trained motor skills during practice. In general, there appears to be a trade-off such that the principles aiding acquisition are not necessarily those that most facilitate learning, and vice versa.

It remains unclear, however, whether speech motor interventions should be tailored within a PML framework, which has an empirical footing from decades of nonspeech motor research (Maas et al., 2008). Nonspeech motor learning studies often teach a new motor pattern that is not used outside of training (e.g., a novel finger or arm movement sequence). Motor speech impairments, however, likely require updating a speech movement for an existing erred or unstable motor pattern that the learner will use in their conversational speech between intervention sessions. It may be, then, that the PMLs that optimize nonspeech motor learning differ from those that optimize speech motor learning. Indeed, studies have not consistently revealed that all nonspeech PMLs are equally applicable to speech motor learning (Maas et al., 2012; Maas & Farinella, 2012). For example, Maas et al. (2012) found that the frequency with which children receive feedback on their productions had a variable effect on speech motor learning, such that some children showed an advantage with more frequent feedback and others showed an advantage for less frequent feedback. The inconsistent empirical findings for PML approaches across individuals with CAS suggest that these PMLs may need to be adapted for clinical implementation in speech therapy. Furthermore, these prior studies exploring PMLs in speech treatment have mostly been small in scale (i.e., single-subject experimental designs) and have also framed research questions as a function of a single PML or static parameter (e.g., high vs. low feedback frequency). In contrast, the current study posits that treatment for pediatric motor speech disorders may require conditions that can be viewed as dynamically changing over time to support a progression in the learning process (Guadagnoli & Lee, 2004). Thus, there may be learning considerations that require a *sequenced approach* that initially emphasizes *principles of acquisition* and systematically transitions to include more *PMLs*. We test this hypothesis by studying two treatment adaptations that are contrary to the status quo for speech therapy for CAS but that are supported by logic, theory, and previous small-scale research with school-age children with CAS: *treatment distribution (i.e., practice intensity)* and *feedback type (i.e., use of ultrasound)*.

Treatment Distribution

In the context of schema-based motor learning (Maas et al., 2008), *practice distribution* refers to how practice *trials* are spaced out in time. In this study, however, we more narrowly define *treatment distribution* as the manner in which an equivalent number of *treatment sessions* are spaced out in time. Treatment distribution can be conceptualized along a continuum from massed (i.e., intensive treatment, condensed in time) to distributed (i.e., spread out in time). While delivering treatment in an intensive (massed) fashion is theorized to enhance motor acquisition, motor learning is theorized to be enhanced when practice is distributed in time (Maas et al., 2008). In fact, there is general consensus from studies of nonspeech motor learning (see T. D. Lee & Genovese, 1988, for a meta-analysis) and cognitive recall (Cepeda et al., 2006) that intensive/massed practice may be detrimental to learning and that distributed practice may be beneficial.

However, there may be unique considerations for applying motor learning principles in the context of speech, clinical populations, and children. As mentioned above, speech errors practiced outside of treatment sessions may compete with the successful attempts during treatment because generalization does not occur immediately, suggesting that more intense practice may be preferred over more distributed practice. Furthermore, children with CAS particularly may be highly inaccurate and/or variable in their speech production and may need to

increase their within-session performance before generalization can occur, further indicating that massed practice may be needed to boost performance when skill level is low. A systematic review by Kaipa and Peterson (2016) supports the need for these unique considerations, highlighting that existing evidence supports the use of massed practice over distributed practice for childhood speech disorders. For example, a randomized group study of pediatric speech therapy has shown that massed practice (three sessions per week for 8 weeks) may be superior to distributed practice (one session per week for 24 weeks) for increasing percent consonants correct in children with phonological disorders (Allen, 2013). Note, however, that participants had phonological SSDs rather than CAS. Another, nonrandomized, study (Thomas et al., 2014) lent support to massed practice increasing overall accuracy of probe words rated on three features (i.e., segmental accuracy, lexical stress accuracy, fluency/nonsegregation of words) over a distributed treatment distribution for children with CAS. Finally, Maas et al. (2019) conducted a single-subject experimental design with six children with CAS using an integral stimulation approach in which treatment targets were treated in massed or distributed practice. The authors observed that the majority of children showed an advantage for improved speech motor learning with massed rather than distributed practice. Thus, the minimal research on pediatric speech sound treatment appears to offer preliminary support for massed practice over distributed practice in this population, but these results are contrary to studies of nonspeech learning and have been generally small in scale. This emphasizes the need to test specific considerations for treatment distribution for children with CAS. In addition, these prior studies have not conceptualized treatment as a sequence designed to progress from intensive (massed) practice to boost acquisition followed by distributed practice to facilitate generalization and retention. Here, this intensive-to-distributed sequenced approach (i.e., the intensive condition) will be contrasted against the traditional approach of providing only distributed practice.

Ultrasound Biofeedback for Motor Learning

In the schema-based motor learning context, *feedback type* refers to the amount of detail provided to the learner: Knowledge of results (KR) feedback provides summary feedback about the accuracy of the production (i.e., "correct!" or "not quite"), while knowledge of performance (KP) feedback provides detailed corrective feedback about the production attempt (Maas et al., 2008). While schema theory does not offer a speech-based framework for the content of KP, we believe that effective KP feedback references movement information used to generate speech sounds, as posited in multiple models of speech production (e.g., Directions Into Velocities of Articulators, Bohland et al., 2010; dual-stream model, Hickok, 2014). These models specify that feedforward and feedback loops are recruited during speech development and in online monitoring of speech output to compare the auditory and somatosensory information generated after speech sound production to a speaker's acoustic and somatosensory goals for the speech sound(s). The speaker's feedback loop allows for motor plans to be updated as errors in production are detected. During development, these internal mechanisms facilitate connections between sensory feedback and speech movement targets and refine the underlying speech representations as they become more adultlike. For speech movements that are habitually in error, such as in those with CAS, it is often necessary to teach the child to inhibit an existing erred motor plan—the old speech movement—and thus to learn a new auditory and somatosensory representation for the adultlike speech sound and intervening sound transitions. Drawing a salient contrast between the desired pattern and the erred pattern can be accomplished through enhancing sensory feedback. We hypothesize that adding another sensory feedback modality—visual biofeedback of real-time tongue movement—could aid the speech learning process for individuals whose auditory and somatosensory feedback systems may not have been sufficient to enable learning of standard speech movements (Preston et al., 2013). Visual feedback may help teach children to associate the visual representation with auditory and somatosensory representations required for accurate feedforward (planning/programming) of speech sounds.

Ultrasound biofeedback of the tongue, therefore, provides a novel form of KP feedback. Biofeedback supplements a clinician's verbal feedback and the learner's own auditory and somatosensory feedback. However, to avoid overreliance on KP feedback (Hodges & Franks, 2001), a preferred sequence to maximize speech learning in children with impaired speech motor systems might be to facilitate speech motor acquisition with visual feedback of movements and subsequently reduce and remove this feedback to encourage speech motor learning (Preston et al., 2018; Preston, McAllister, et al., 2019). In this study, this sequenced approach involving frequent ultrasound biofeedback that fades over time will be contrasted against the traditional therapeutic approach of providing treatment without visual feedback of speech movements.

Case studies and single-subject experimental designs have revealed that treatment programs that include articulatory biofeedback such as ultrasound can result in successful speech sound learning in children with CAS (Preston et al., 2013, 2016; Preston, Leece, et al., 2017), as well as in school-age children and young adults with persisting speech sound errors (Adler-Bock et al., 2007;

Cleland et al., 2015, 2019; McAllister Byun et al., 2014; Preston & Leece, 2017). These studies generally have provided evidence that biofeedback of articulatory movements can facilitate greater improvement than a no-treatment condition. However, this study addresses limitations of prior work by comparing two similarly structured treatment programs that differ in exposure to real-time ultrasound biofeedback. To extend the current level of evidence from case series and single-subject experimental designs for CAS (Preston et al., 2013, 2016), there is a need for a randomized comparison of treatment with and without biofeedback to advance the quality of research and employ the proper experimental control (Robey, 2004). Importantly, based on the current state of the literature, it is unknown whether the biofeedback is an operative component in treatment for CAS, and a randomized comparison can help to address this limitation.

Speech Motor Chaining

We test the effects of treatment distribution and ultrasound biofeedback within the framework of a motor-based treatment, Speech Motor Chaining (Preston, Leece, & Storto, 2019). Speech Motor Chaining is a treatment developed to systematically teach new sounds, sequences, and intersound movements (i.e., transitions) in children with frequent sound errors and is intended to be appropriate for children with mild–moderate CAS who retain sound errors in late childhood and adolescence (e.g., Lewis et al., 2021). Speech Motor Chaining therefore differs from other CAS treatments that are intended for more severely impacted individuals with limited speech (Strand, 2020) or treatments that target prosodic aspects of speech using existing sounds (McCabe et al., 2020). Speech Motor Chaining is adaptive, with an initial emphasis on principles of acquisition and subsequent emphasis on PMLs (as dictated by the child's performance). Speech Motor Chaining targets core syllables, composed of sounds that are of low accuracy for a child, in order to address errors in the context of movement sequences that transition into and out of the target sound. The syllable is emphasized (vs. sounds in isolation) due to evidence that many aspects of speech planning occur in syllabic units (Laganaro et al., 2012; Nijland et al., 2003; Rogers & Storkel, 1998). As performance increases, the emphasis of Speech Motor Chaining adapts to learning, and the core syllables are embedded within longer units (monosyllabic words, multisyllabic words, phrases, and sentences). Several PMLs are incorporated throughout practice, such as increasing prosodic variability and complexity of the movements surrounding the core syllable and modifying the type and frequency of feedback (Preston, Leece, & Storto, 2019).

In addition, although motor learning principles provide the primary framework for the current study, it is important to acknowledge that CAS may also be accompanied by broader linguistic deficits, including poor phonological awareness (Lewis et al., 2004; McNeill et al., 2009c; Nijland, 2009). For example, Lewis et al. (2018) reported in a large cohort study that adolescents with CAS have persisting delays in phonological awareness skills compared to children with other types of SSD. As such, well-rounded treatment programs for CAS likely emphasize both speech motor control and other relevant skills such as phonological awareness to reinforce underlying phonological representations that may contribute to a child's clinical presentation (McNeill et al., 2009a, 2009b). In this study, we assume that phonological awareness training may improve children's understanding of where sounds are within words so they can better generalize the movements they learn in production practice with Speech Motor Chaining.

Telepractice

During the COVID-19 pandemic, face-to-face research was paused for several months, and a subset of participants transitioned to telepractice. While there is a growing evidence base supporting the efficacy of speech therapy delivered via telepractice for articulation and phonological disorders (S. A. S. Lee, 2018; Peterson et al., 2022; Wales et al., 2017), there is a paucity of research on telepractice treatment for CAS. To our knowledge, there is only one published study that investigated the efficacy of telepractice for any CAS treatment. Thomas et al. (2016) used a multiple-baseline across-participants design to examine the efficacy of Rapid Syllable Transition (ReST) treatment for five children aged 5–11 years with CAS. The authors found that children who received ReST via telepractice demonstrated similar acquisition and generalization outcomes to children who had received ReST in prior face-to-face treatment studies. While the results of this study are promising, there is a need for more research on CAS treatment delivered via telepractice.

Functional Impacts of CAS

One clear goal for treating CAS is to improve speech production. Beyond speech production, CAS can have functional impacts in other areas, such as a child's socialization and academic participation (Lewis et al., 2021; Rusiewicz et al., 2017). To date, there is little information on the extent to which children with CAS or their parents report functional effects of treatment. This information could provide insight into any functional changes as a result of treatment and the time course of such changes. Therefore, the current study explores parent- and

child-reported changes that may be associated with the functional impact of treatment.

Purpose and Hypotheses

The primary goal of this study was to examine the extent to which evidence-based treatment outcomes could be influenced by treatment distribution and ultrasound biofeedback. Children were randomized to two different treatment distribution conditions (i.e., intensive vs. distributed) and to two different biofeedback conditions (i.e., with and without ultrasound biofeedback). The study also explored the effects of treatment distribution for face-to-face and telepractice therapy modalities. Study outcomes examined speech improvements and any participant-reported social–emotional impact.

It was hypothesized that an intensive treatment distribution would result in greater speech motor generalization than a more traditional, uniformly distributed treatment schedule (e.g., Namasivayam et al., 2015; Preston, Leece, et al., 2017). In this study, the 5-week–long intensive treatment distribution began with a massed week of 10 hr of treatment that faded to thrice per week in Weeks 2 and 3 and then fading again to a more traditional treatment distribution, twice per week, for Weeks 4 and 5. This contrasted with the distributed treatment condition, in which participants were scheduled twice per week for 10 weeks. It was also hypothesized that the availability of ultrasound biofeedback of the tongue during treatment would result in greater speech motor learning than practice without biofeedback. Furthermore, we examined the combined effects of these conditions, and it was hypothesized that intensive treatment with biofeedback would yield the greatest improvement in speech production accuracy. In addition, we also examine these outcomes when holding constant the time since the end of treatment (i.e., immediately after treatment has concluded and 5 weeks later) to examine retention.

Method

The study was approved by the Syracuse University Institutional Review Board (#17-177). Participants provided written assent, and their parents/guardians provided informed written consent. The trial was registered through ClinicalTrials.gov (NCT03238677). Recruitment began in September of 2017 and continued through May 2022 when funds were exhausted. Recruitment included contact with regional speech-language pathologists (SLPs) through mailings and e-mails to school district personnel and presentations to local SLP groups. Public announcements on local social media groups, radio ads, television spots, and Facebook advertising were also used to raise awareness within the community. Based on prior studies, a power analysis suggested that a sample size of 40 participants (10 per group) would be sufficient to identify significant main effects of biofeedback and of treatment intensity.

Participants

Screening

Eligible participants were American English–speaking children between 9 and 17 years of age. Eligibility for the full evaluation was based on a phone screening in which parents reported a history of CAS or CAS-like speech features, normal hearing, and vision that was normal/corrected to normal. Children with developmental disabilities such as autism, cerebral palsy, or Down syndrome were not eligible, although children with concomitant learning disabilities were not excluded. Families had to agree to have their child randomized to any of the four treatment conditions described below.

Local participants who met phone screening criteria were invited for a screening session to fully assess eligibility at the Syracuse University Speech Production Laboratory. In the cases of participants recruited to complete the study via telepractice (see below), eligibility assessments were completed via Zoom in a participant's residence. Overall, we assessed 96 children. Children were required to meet the following inclusionary criteria: (a) having passed a hearing screening bilaterally at 20 dB at 500, 1000, 2000, and 4000 Hz; (b) Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007) standard score $\geq$ 70; (c) Concepts and Following Directions subtest of the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig et al., 2013) scaled score $\geq$ 4; (d) Goldman-Fristoe Test of Articulation–Third Edition (GFTA-3; Goldman & Fristoe, 2015) percentile score $\leq$ 5; and (e) having demonstrated no frank oral structural anomalies (e.g., submucous cleft) based on an oral structural exam. Additional standardized tasks were administered for descriptive purposes: the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence–Second Edition (Wechsler, 2011), the three subtests of the Phonological Awareness Composite of the Comprehensive Test of Phonological Processing–Second Edition (i.e., Elision, Phoneme Isolation, and Blending Words; Wagner et al., 2013), and the Recalling Sentences subtest of the CELF-5 (Wiig et al., 2013). Descriptive information for the groups at pretreatment is reported in Table 1, and individual data are available in the supplemental dataset on the study's Open Science Framework (OSF) page.

All children who qualified were reported to have a previous history of speech therapy, although none had

Table 1. Pretreatment means ± standard deviations and ranges for 56 children with childhood apraxia of speech randomized to six groups.

Variable	Distributed + No Ultrasound (*n* = 9)	Distributed Telepractice (*n* = 10)	Distributed + Ultrasound (*n* = 10)	Intensive + No Ultrasound (*n* = 9)	Intensive + Telepractice (*n* = 9)	Intensive + Ultrasound (*n* = 9)
Age (years)	10.85 ± 1.7 (9.41–14.67)	11.05 ± 2.4 (9.03–15.48)	12.22 ± 2.1 (9.17–15.44)	10.01 ± 1.1 (9.15–12.61)	11.16 ± 2.3 (9.44–16.67)	10.31 ± 1.0 (9.01–11.61)
PPVT-4 standard score	104.6 ± 12.0 (91–126)	96.8 ± 9.8 (83–115)	100.7 ± 11.9 (82–113)	97.7 ± 13.1 (75–119)	101.9 ± 10.8 (85–119)	112.1 ± 13.2 (95–137)
CELF-5 Concepts and Following Directions scaled score	9.0 ± 1.2 (7–11)	Not administered via telepractice	8.6 ± 3.0 (6–14)	10.6 ± 4.2 (4–16)	Not administered via telepractice	9.9 ± 2.8 (7–16)
CELF-5 Recalling Sentences scaled score	8.6 ± 2.5 (4–11)	4.2 ± 2.1 (1–7)	6.7 ± 3.0 (1–12)	5.9 ± 3.2 (1–11)	3.4 ± 1.7 (1–6)	11.7 ± 5.0 (6–18)
GFTA-3 standard score	43.9 ± 8.8 (40–66)	43.9 ± 8.4 (40–63)	42.3 ± 5.5 (40–57)	44.9 ± 9.7 (40–63)	43.2 ± 7.6 (40–63)	49.8 ± 12.1 (40–73)
CTOPP-2 PA Composite	81.7 ± 13.6 (67–96)	64.3 ± 12.1 (45–88)	72.8 ± 14.1 (56–98)	78.0 ± 13.3 (60–96)	68.7 ± 7.0 (52–75)	90.0 ± 10.0 (91–98)
WASI-II Matrix Reasoning *T* score	45.9 ± 7.8 (33–57)	41.0 ± 9.5 (29–57)	39.0 ± 9.5 (23–55)	44.1 ± 8.1 (28–54)	47.2 ± 9.0 (34–60)	48.3 ± 12.1 (34–67)

Note. Note that GFTA-3 percentiles (< 5) were used for determining eligibility, but because percentiles are range limited, standard scores are summarized here. PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; CELF-5 = Clinical Evaluation of Language Fundamentals–Fifth Edition; GFTA-3 = Goldman-Fristoe Test of Articulation–Third Edition; CTOPP-2 PA Composite = Comprehensive Test of Phonological Processing–Second Edition, Phonological Awareness Composite; WASI-II = Wechsler Abbreviated Scale of Intelligence–Second Edition.

participated in prior studies in our lab. Information about participant race, sex, and ethnicity was collected on a parent questionnaire. Of the 56 eligible participants randomized to treatment, 39 were male and 17 were female. Using National Institutes of Health categories of race and ethnicity, participants were identified as follows: non-Hispanic/non-Latino White (42), Hispanic/Latino White (4), Hispanic/Latino Black or African American (1), American Indian or Alaskan Native (1), and mixed/multiple races (6; self-identified as Black + Italian, Black + White, American Indian + White, Asian + White). Two families declined to provide race/ethnicity information.

Procedure for CAS Diagnosis

The diagnosis of CAS was determined based on children's performance on three tasks, which yielded four total variables. Children were required to meet the CAS criteria on at least two of the four variables. Analysis of performance was based on audio recordings rather than live judgment.

The first CAS eligibility task, the syllable repetition task (Shriberg et al., 2012), required children to imitate 18 nonwords ranging from two to four syllables containing /bɑ/, /mɑ/, /nɑ/, or /dɑ/. From this task, the outcome measure was the percentage of nonwords with added sounds, such as /dɑbɑmɑ/ → /dɑnbɑmɑ/ (i.e., transcoding score, which counts sound additions but not syllable additions; see Shriberg & Lohmeier, 2008). Following Shriberg et al. (2012), if more than 20% of items (i.e., four or more of the 18 nonwords) contained sound additions, the participant met the criteria for CAS on this task.

The second CAS eligibility task, the maximum performance tasks, followed procedures described by Thoonen et al. (1999) and formalized by Rvachew et al. (2005). This task requires children to sustain the sounds /ɑ/, /s/, /f/, and /z/ and repetitions of /mɑmɑ/ for as long as possible, as well as repeat the syllables /pɑ/, /tɑ/, and /kɑ/ and /pɑtɑkɑ/ as quickly as possible. The apraxia score (i.e., 0 = *not apraxic*, 1 = *undetermined*, 2 = *likely apraxic*) is based primarily on the accuracy and rate of repeated /pɑtɑkɑ/ productions (i.e., less than four accurate repetitions of /pɑtɑkɑ/ or ≤ 3.4 syllables per second in /pɑtɑkɑ/ yields a score of "likely apraxic"). If the participant obtained a "likely apraxic" score based on scoring procedures described by Rvachew et al., the participant met the CAS criteria on this task. A dysarthria score is also calculated from the maximum performance tasks, with seven participants receiving a "likely dysarthric" score and three receiving an "undetermined" score, but the dysarthria score was not used to determine eligibility for the study.

The third CAS eligibility task was a polysyllable picture naming task consisting of 80 words modeled after Gozzard et al. (2006). Pictures of polysyllabic words (e.g., cheeseburger, escalator) were presented, and the participants' responses were recorded as WAV files. There were two variables derived from this third task: lexical stress percent correct and number of syllable segregations. Raters reviewed the set of 80 items twice: first to rate lexical stress accuracy (i.e., correct number of syllables and correct stress pattern) and then for the presence or absence of syllable segregation (i.e., an inappropriate pause or gap

in the middle of the word). The criterion for CAS on the lexical stress measure was < 70% accuracy, as children with typical speech are generally above this threshold (Gozzard et al., 2006) and children with CAS are typically below it (McCabe et al., 2014; Murray et al., 2015; Shriberg, Aram, & Kwiatkowski, 1997). The criterion for CAS on the syllable segregation measure was the presence of segregated syllables on > 3% of polysyllabic words, based on data from Murray et al. (2015).

To be eligible for the study, children had to meet CAS criteria on at least two of the four variables extracted from these three tasks. Furthermore, to ensure reliability, the performance on the three tasks was measured independently by two raters (at least one SLP plus a speech-language pathology graduate student trained in the protocol). If there was disagreement regarding whether the child met CAS eligibility requirements, then a third rater was required to independently score the child's performance on the three tasks to determine eligibility. Therefore, performance on these tasks, rather than one clinician's subjective clinical impression, was required to determine eligibility. We also completed the Mayo-10 feature checklist (Shriberg et al., 2011) for each participant to facilitate comparison of this study's population to previous studies of CAS. While the Mayo-10 data were not used to determine study eligibility, all participants who met our eligibility criteria also met CAS criteria on the Mayo-10 checklist (i.e., demonstrated at least four CAS features across at least two different tasks). Interested readers can find participant Mayo-10 checklists in the supplemental dataset on the study's OSF page.

Diagnostic reliability. Diagnostic reliability was analyzed for the 88 assessed children who met the receptive language inclusionary criteria. At the participant level, the first two raters agreed on whether the participant met the operationalized CAS criteria for 79 of 88 participants; therefore, nine participants required a third rater to determine whether the child met CAS criteria. At the task level, the first two raters agreed on the binary judgment of meeting the task criteria for 88% of participants on the syllable repetition task, 91% of participants on the maximum performance tasks, 87% of participants on lexical stress accuracy, and 80% of participants on syllable segregation.

Study Design

The study's 2 × 2 factorial design was fully realized for participants in face-to-face treatment (but see the COVID-19 Modifications section below regarding the inability to test ultrasound biofeedback for telepractice participants). Half of the eligible participants who were seen face-to-face were randomized to receive production practice with ultrasound biofeedback, and half were randomized to receive production practice without any biofeedback. Furthermore, half of the participants were randomized to an intensive schedule (i.e., 20 hr of treatment in 5 weeks, beginning with an intensive week with 10 hr of therapy), whereas the remaining half of participants were randomized to a distributed schedule (i.e., 20 hr of treatment at a rate of two visits per week for 10 weeks). This resulted in four groups: Distributed + Ultrasound, Intensive + Ultrasound, Distributed + No Ultrasound, and Intensive + No Ultrasound. Randomization was completed using concealed envelopes. The randomization scheme was completed by the first author using random.org to generate blocks of four such that each treatment condition was represented once in each sequence of four envelopes. The assessors and treaters were masked (i.e., blinded) to the randomization order, and random assignments were distributed to the clinician and participant only once a participant was deemed eligible. Participant enrollment, randomization, and completion are summarized in Figure 1. A total of 96 children were assessed, 56 were eligible and randomized to treatment conditions, and 48 completed treatment.

All clinicians held a master's degree and were either certified SLPs or, in the case of a Clinical Fellow working toward certification, supervised by a licensed clinician. A total of seven different clinicians conducted treatment. Face-to-face treatment occurred in the lab or, for two participants for whom the travel distance was a burden, in a quiet room in a public library located between the lab and the participant's home. Treatment for telepractice participants (described below) occurred via secure Zoom connection. For fidelity monitoring, treatment sessions were audio-recorded or screen-recorded (in the case of sessions with ultrasound or with telepractice).

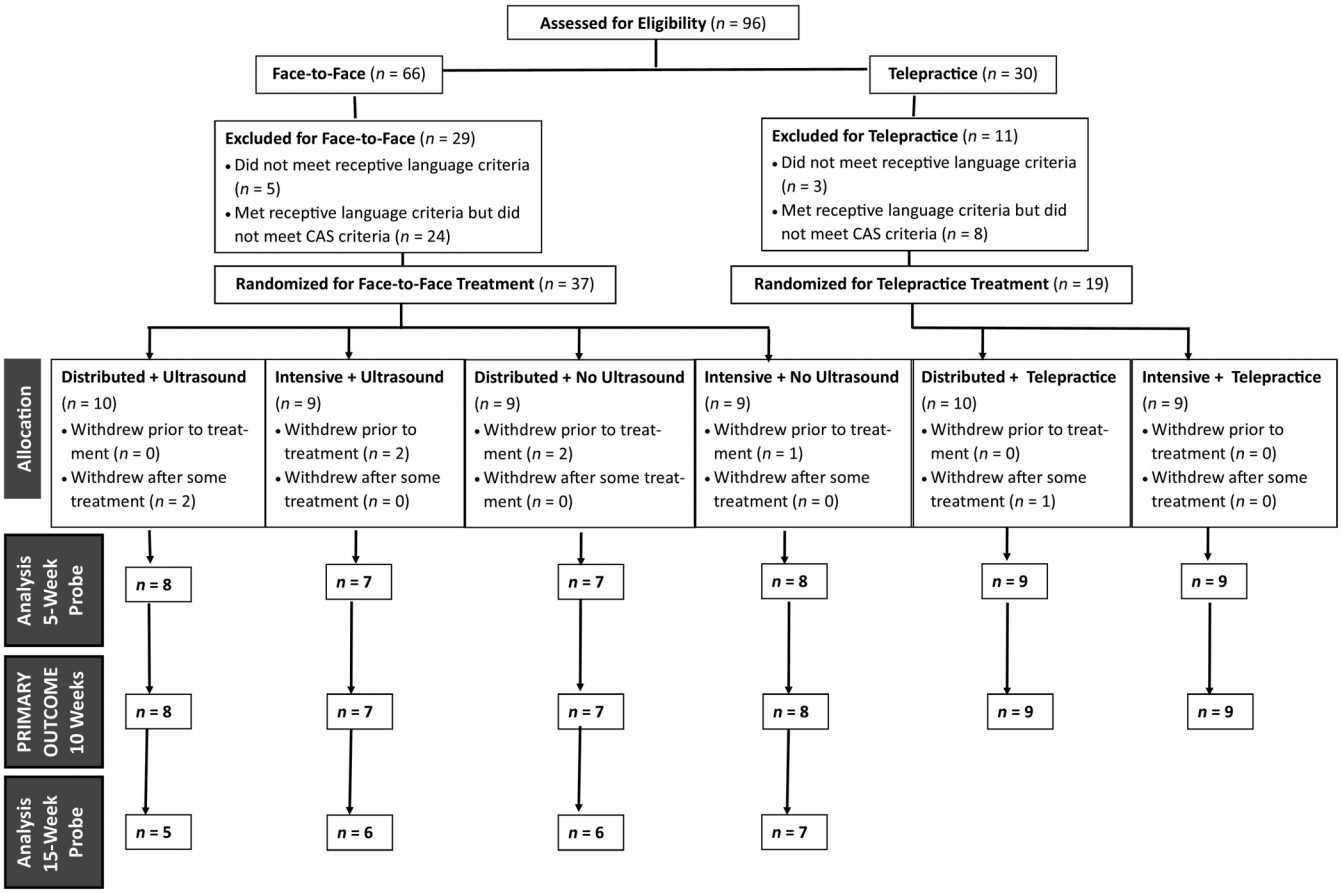
Common Aspects of Treatment in All Conditions

Treatment target selection criteria and the treatment session structure were common to each condition, as outlined below.

Treatment target selection. For each participant, we identified three target sound–positions. Eligible targets were consonants of American English, fixed in a syllable position in order to situate the target within a movement context rather than in isolation (e.g., /l/ onset, /s/ coda). Two targets were treated each session, and a third was introduced to replace an original target if criteria were met for mastery (see below).

A stimulability task was administered to assist with target selection. Before treatment, consonant–vowel and vowel–consonant stimulability was assessed for the sounds

Figure 1. Consolidated Standards of Reporting Trials table depicting participant enrollment, randomization, and follow-up. CAS = childhood apraxia of speech.



that the clinician perceived to be in error during the speech assessment tasks (e.g., GFTA-3, polysyllable picture naming). Up to six sound–positions were initially assessed (e.g., /l/ onset, /s/ coda). For each sound–position assessed, the child imitated these sounds in five different syllables, following the clinician's verbal model. Each syllable was repeated 3 times (e.g., for /l/ onset, three repetitions each of /li/, /le/, /lo/, /lu/, and /lʌ/).

The clinician identified the three least stimulable sound–positions from those that were assessed on pretreatment stimulability probes. The two sound–positions with the lowest stimulability were selected by the clinician as treatment targets, for each participant (if all three of the least stimulable sound–positions were at 0%, the selected treatment targets were two sound–positions that were different in manner and place of articulation; otherwise, it was possible to have two sound–positions targeting the same phoneme). The third sound–position target was held in reserve for future treatment in the event that a participant met the study's mastery criteria for a treated target (i.e., > 80% accuracy in sentences in two consecutive sessions) before the 20 sessions were finished. For each sound–position target, four syllable-level variants were chosen for practice. For example, for the target sound–position /l/ onset, four syllable-level variants might include /lo/, /laɪ/, /klV/, and /plV/; for /s/ coda, four variants might include /is/, /aɪs/, /ɪks/, and /æst/. Within a session, the goal was to teach movement transitions using a variety of adjacent sounds surrounding the target sound–position. If phonotactically plausible, the clinician was encouraged to select both singletons and clusters. When singletons were chosen, the SLP was instructed to include at least one potentially facilitative vowel context (e.g., a low-back vowel adjacent to /ɹ/, a high-front vowel adjacent to /s/) and to vary the vowels (i.e., include both monophthongs and diphthongs, avoid vowels that are adjacent in the quadrilateral).

Furthermore, to monitor changes in stimulability at each time point, the stimulability probes were also administered at probe visits. Stimulability was scored independently by three listeners who were masked to treatment condition and to time point using audio recordings. Each

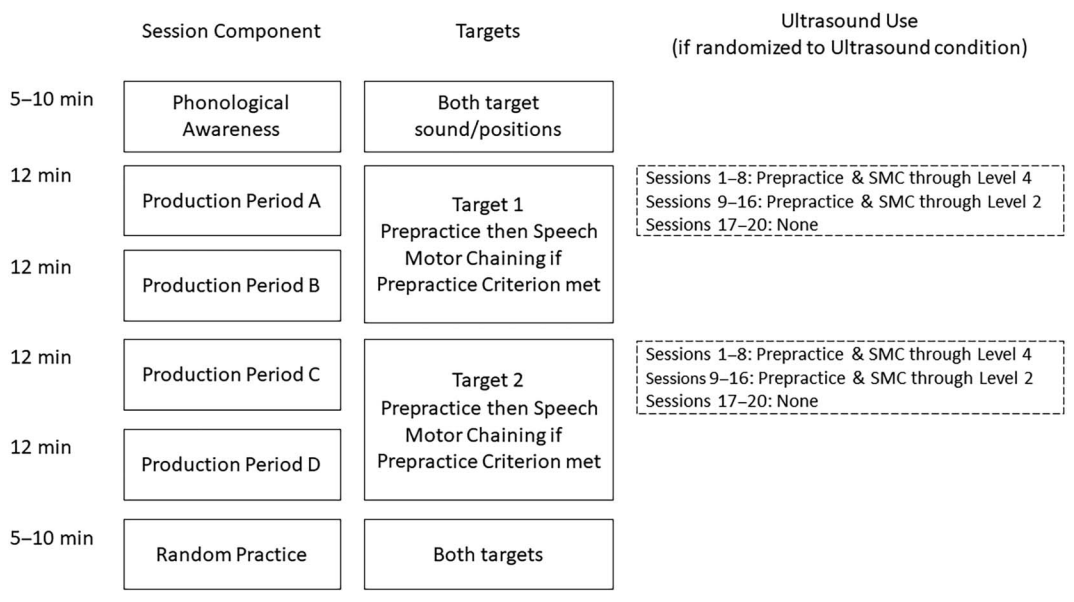
syllable was rated as correct (1) or incorrect (0) based on accuracy of the entire syllable (both the consonant and the adjacent vowel). The outcome measure of the stimulability probe was percent correct for the target sound–position for the session, averaged over three listeners. The intraclass correlation coefficient (ICC) for all stimulability data was .894, 95% CI [.873, .912]. When separated out by telepractice condition, the ICC for face-to-face data was .932, 95% CI [.915, .947], and the ICC for telepractice was .827, 95% CI [.773, .870]. This indicates good-to-excellent reliability in the stimulability data (Koo & Li, 2016).

Treatment session structure. Participants received individual therapy sessions. Each session began with 10 trials (~5–10 min) of phonological awareness training followed by four 12-min periods of structured sound practice (Periods A, B, C, and D) and concluded with 5–10 min of random practice. The first target (e.g., /l/ onset) was practiced in Periods A and B, while the second target (e.g., /s/ coda) was practiced in Periods C and D. The target practice order alternated in each session (e.g., one session treated /l/ onset in Periods A and B and /s/ coda in Periods C and D; the next session treated /s/ coda in Periods A and B and /l/ onset in Periods C and D). Timed components of the session were monitored with a digital timer. Figure 2 depicts the session structure. Videos demonstrating the components of treatment are available on the study's OSF page.

Phonological awareness training. Sessions began with 10 trials of phonological awareness training, modeled after prior studies (McNeill et al., 2009c; Moriarty & Gillon, 2006). Phonological awareness training centered around the targets chosen for production practice. Children manipulated blank colored blocks (LEGO DUPLO) to demonstrate understanding of sound addition, deletion, or substitution based on the clinician's instructions. Children began with a word containing a sound in error (e.g., for target /l/ onset, /lif/), and they then engaged in 10 practice trials (e.g., change /lif/ to /lit/, change /lit/ to /lits/, change /lits/ to /luts/). No speech was required by the participant, only manipulation of the blocks. Performance-related feedback was provided after every trial. The training included five levels of increasing complexity. Levels 1–4 included one sound change in one-, two-, three-, and four-syllable items. Level 5 included two sound changes in two-syllable items (e.g., change /lifmus/ to /limes/). Criterion to step up to the next level at the next session was at least nine of 10 correct. Criterion to step down a level at the next session was 0–5 correct. Criterion to stay at the same level at the next session was 6–8 correct. Phonological awareness training was discontinued if the participant met the step-up criterion (at least nine correct) at Level 5 in two sessions. A datasheet, available on the study's OSF page, was used to guide the phonological awareness training portion of the session.

Production prepractice. During the prepractice (elicitation) phase, clinicians aimed to establish a reference for correctness of the target sequences, which involved clinical cueing to facilitate stimulability for production of the four variants of the target (e.g., for target /l/ onset, /lo/, /laɪ/, /klV/, and /plV/). SLPs were free to use a variety of strategies (e.g., imitation, gestures, verbal cues, analogies, shaping) to achieve accurate productions of the variants

Figure 2. Overview of treatment sessions. SMC = Speech Motor Chaining.



so long as the strategies were congruent with the treatment condition. As illustrated in Figure 2, Period A began with elicitation of the first target, and Period C began with elicitation of the second target. Prepractice continued until the performance criteria were met of all four syllable-level variants produced correctly 3 times each (i.e., 12 total correct productions). Once this criterion was met, practice began with Speech Motor Chaining (described below). Therefore, prepractice could take all 24 min devoted to a sound–position if the performance criterion was not met, or prepractice could be completed in approximately 1 min if the child was highly successful. Cueing used during prepractice included direct imitation, verbal articulatory cues (i.e., clearly discussing the different parts of the tongue that are required to form the sound), shaping procedures from facilitative contexts, and visual reference to articulation. In the ultrasound condition, visual reference involved biofeedback with the tongue during Periods A and C. When no ultrasound was used, visual reference included static electropalatography images, anatomical images from the Internet, sagittal magnetic resonance images, or animations from Seeing Speech (Lawson et al., 2018).

Production practice with Speech Motor Chaining. Once the prepractice criterion of 12 total correct imitated productions was achieved for the four syllable variants of a target, Speech Motor Chaining practice began for the remainder of the allotted time for that sound–position. Procedures followed those described by Preston, Leece, and Storto (2019) and were guided by a datasheet in Microsoft Excel (available on the study's OSF page). Each of the four syllable variants treated in prepractice (e.g., /aɪs/) was used as the core movement pattern around which the production practice chains were constructed. Chains are designed to increase in complexity from Level 1 (syllables, e.g., /aɪs/) to Level 2 (monosyllabic words, e.g., "nice"), Level 3 (multisyllabic words, e.g., "nicely"), Level 4 (phrases, e.g., "talked nicely"), and Level 5 (child-generated sentence, e.g., make up a sentence with "nicely"). Practice occurred at each level in blocks of six consecutive trials, with items elicited via imitation. If the child achieved at least five of six correct at the level, they progressed to the next highest level. If they produced fewer than five correct in a level, they began at Level 1 (syllables) for the next variant (e.g., /is/). If a participant mastered all five levels, the chain was discontinued the next session and replaced with a new chain; otherwise, the same chains were practiced from session to session.

Note that Speech Motor Chaining is more than just practice with stimuli that increase in contextual linguistic complexity for the target syllable–position. As a participant advances through the levels, several PMLs are also manipulated to change the practice emphasis from acquisition (the focus of the lower levels) to motor learning (the focus of the higher levels). As guided by the Speech Motor Chaining datasheet, the clinician's *feedback frequency* was systematically faded (from feedback on five of six trials at Level 1 to three of six trials at Level 5), while the *feedback type* changed from primarily KP at Level 1 to primarily KR at Level 5 (see Preston, Leece, & Storto, 2019). *Practice variability* was manipulated in higher levels by introducing variable prosody (i.e., loud, slow, fast, question, exclamation, neutral; see Preston, Leece, et al., 2017). During half of trials in each block, participants were also prompted to provide a self-judgment of their accuracy before any clinician feedback was offered.

Random practice. The final 5 min of each session (or 10 min, if the phonological awareness goal was met) consisted of random practice. First, the clinician selected the practice items: the highest linguistic level for each chain that was mastered (at least five correct) from Periods B and D. If the participant did not get past prepractice, only syllables were targeted during random practice. On average, 30 trials were elicited each session in random practice. Before each trial, the participant rolled dice or spun a spinner to determine prosodic variation (i.e., loud, slow, fast, question, exclamation, or neutral) and then imitated the target utterance using the selected prosodic cue. Regarding *feedback type*, delayed KR was provided on every other trial except for syllable targets, which received immediate KP, under the assumption that more feedback was needed if the child did not advance out of prepractice. The datasheet was used to guide the clinician as to the practice items and the frequency and type of verbal feedback.

The response definition for correct productions throughout all of prepractice, Speech Motor Chaining practice, and random practice required the target syllable to be fully correct. All elements of the core syllable (i.e., consonant[s] and adjacent vowel) were required to be fully correct with no sound additions, distortions, omissions, or substitutions. Transitions between the sounds within the target syllable were required to be clear (not distorted) and uninterrupted, but a slow response was not scored as an error as it is typical for children to slow down when focusing on their productions.

Experimental Treatment Conditions

Participants were assigned to one of two treatment distribution conditions (intensive or distributed) and to one of two biofeedback conditions (no ultrasound or ultrasound), as outlined below.

Intensive. Children assigned to the intensive treatment condition completed 20 hr of treatment in a 5-week period. Ten hours of treatment occurred in Week 1 (5 days, each containing two 60-min sessions). The intensity was

reduced such that participants attended a 60-min session three times per week in Weeks 2 and 3 as well as two times per week in Weeks 4 and 5. This condition was designed to represent a sequenced approach that transitions from an emphasis on motor acquisition to motor learning.

Distributed. Children assigned to the distributed treatment condition completed 20 hr of treatment in a 10-week period. Two 60-min sessions were scheduled to occur twice weekly. Clinicians were instructed that it was preferred to avoid scheduling sessions on back-to-back days if possible, although this was not required due to logistical constraints.

Ultrasound. Participants assigned to the ultrasound condition used real-time ultrasound imaging of the tongue in a fading fashion, as shown in Figure 2. Ultrasound visual feedback was included during Periods A and C (the first 12 min working on a target sound–position each session), whereas ultrasound was not available in Periods B and D to encourage independent practice without visual feedback. When the ultrasound was used in Periods A and C, it was available only when the participant was in prepractice and when the participant practiced at Speech Motor Chaining Levels 1 (syllables) through 4 (phrases). After 8 hr of treatment, ultrasound was faded such that it was only available as children practiced Levels 1 and 2 to encourage independent practice without visual support at more complex levels. After 16 hr of treatment, ultrasound visual feedback was faded out completely and was not used during the final four sessions. Therefore, this condition was also designed to represent a sequenced approach that transitions from an emphasis on motor acquisition to motor learning.

When ultrasound biofeedback was used, the clinician was free to choose sagittal or coronal view depending on the nature of the target sound–position and the type of errors being treated. For example, lateralized distortions of sibilants were generally treated in coronal view (Preston et al., 2014), and errors on tongue tip elevation for alveolars such as /t/ and /l/ could be treated in sagittal view, but /ɹ/ errors could be treated in either sagittal or coronal view depending on the movements being targeted (Preston, McAllister Byun, et al., 2017). The clinician was free to draw tongue shape templates on the screen for the client to match or would draw targets on the screen as cues for the child to try to hit (e.g., an arrow pointing to the elevation of the tongue tip to achieve the initial movement of /laɪ/). During biofeedback trials, children were prompted to watch the screen and self-monitor their movements. Clinicians were able to reference the visual display when providing verbal KP feedback (e.g., "I didn't see that groove when you said /is/" or "Good! Your tongue tip went up when you started *line*!"). When the ultrasound was used, either the clinician held the probe or (more commonly) the participant was instructed on how to hold the probe beneath the chin. Ultrasound biofeedback was conducted using either an Echo Blaster 128 (Telemed Medical Systems) or a MicrUs (Telemed Medical Systems) ultrasound, with the display shown in Echo Wave II 4.0.0 software on a Dell computer with a 24-in. monitor.

No ultrasound. Children in the no-ultrasound condition completed prepractice and Speech Motor Chaining components of sessions as described above without the aid of any real-time visual feedback of the tongue, as is the case in traditional therapy approaches.

COVID-19 Modifications

Face-to-face data collection was paused in March 2020 due to the COVID-19 pandemic. Two participants who were in the Intensive + No Ultrasound treatment at the time transitioned to telepractice for their last two treatment sessions. Participants were mailed a headset microphone to ensure consistent audio quality, and the sessions were conducted via Zoom. For both participants, the 5-week and 10-week probes were also collected via Zoom. One additional participant declined to transition to telepractice and withdrew after Session 14.

Furthermore, in May 2020, new participants were recruited via social media and were randomized to receive treatment via telepractice following the same intensive and distributed schedules described above (hereafter, Intensive + Telepractice and Distributed + Telepractice). It was not possible to randomize participants to receive ultrasound biofeedback by telepractice. Assessment and treatment was delivered via Zoom, with both the clinician and the participant using an IPD IPH-165 binaural noise-canceling microphone headset. Procedures for assessment were similar to face-to-face procedures with the following exceptions: We relied on parent report of normal hearing rather than using a hearing screening, oral structural exams were limited to what was visible on the participant's web camera, and we were not able to administer the CELF-5 Concepts and Following Directions. During treatment, the structure of prepractice, Speech Motor Chaining practice, and random practice were the same as face-to-face treatment. The phonological awareness training component was adapted to be administered with PowerPoint. Rather than manipulating colored blocks, the participants were given remote control access to the clinician's screen and dragged colored squares into empty boxes. If the clinician was not able to give remote control access, the participant verbally identified the block that needed to be changed and the type of change (e.g., "take away the green block," "change the purple block").

When returning to face-to-face face treatment in November 2020, the room setup was revised to include a 6-ft distance between the SLP and the participant. A clear mask was worn by the clinician, and a surgical mask was worn by the child. However, progress-monitoring probes were collected without a mask in a soundproof booth, as was done before the pandemic.

Clinician Training and Treatment Fidelity

Intensive one-on-one training in the treatment procedures was provided by the first author. A treatment manual (available on the study's OSF page) was provided to the clinicians, and videos of Speech Motor Chaining and ultrasound biofeedback from previous studies were used to demonstrate the procedures. Direct observation and feedback were provided to the treating clinicians after the first two sessions, with biweekly check-ins to address questions that arose. Treatment fidelity was completed by research assistants who reviewed audio or video of the entire treatment session for four sessions per participant (20% of sessions). The research assistants were provided with an individual training session, a manual to guide fidelity scoring, a 30-min training video, and a check-in with the first or third author following training. Feedback on treatment fidelity from these sessions was provided to the treating clinician on an ongoing basis so they were aware of areas where they were relatively stronger or weaker with respect to adherence to the protocol as a part of ongoing clinician training.

Specific elements within treatment sessions were evaluated for fidelity during phonological awareness training, prepractice, Speech Motor Chaining practice, and random practice. As shown in Table 2, fidelity was high, exceeding 87% for all elements measured for all groups.

Fidelity of Scheduling

Treatment scheduling was an important component of the study design. For those who completed all treatment sessions, the number of days from the first to the last treatment session was 76 (SD = 14) for Distributed + No Ultrasound, 74 (SD = 7) for Distributed + Ultrasound, 70 (SD = 4) for Distributed + Telepractice, 37 (SD = 4) for Intensive + No Ultrasound, 33 (SD = 3) for Intensive + Ultrasound, and 32 (SD = 2) for Intensive + Telepractice.

Fidelity of Ultrasound Biofeedback

Use of ultrasound biofeedback during sessions was a second important component of the design. The researcher recorded the minutes spent using ultrasound biofeedback during face-to-face treatment sessions. These estimates were summed across each participant. For any given participant, time spent with ultrasound could theoretically range from 0 min (in the case of No-Ultrasound Sessions or Ultrasound Sessions 17–20) to 384 min (24 min in each session across Sessions 1–16). For participants who completed treatment, the minutes spent with ultrasound was 0 (SD = 0) for Distributed + No Ultrasound, 351 (SD = 38) for Distributed + Ultrasound, 0 (SD = 0) for Intensive + No Ultrasound, and 332 (SD = 26) for Intensive + Ultrasound.

Outcome Measures

Progress-Monitoring Speech Probes

To measure outcomes, speech production probe data were collected 1 week before treatment, after 5 weeks (after 20 hr of therapy for the intensive group and 10 hr of therapy for the distributed group), after 10 weeks (after both groups had completed 20 hr of therapy), and after 15 weeks (to assess retention after treatment was discontinued). The 10-week probe was the primary outcome point for this study. Face-to-face probes were collected in a single-walled soundproof booth with a Sennheiser MKE 2 lapel microphone, and telepractice probes were collected over Zoom. Probe sessions were assigned a random four-digit number so that raters who scored and phonetically transcribed the probe sessions were masked to participant number, time point, and treatment condition.

Performance on untreated phrase probes was used to assess the primary outcomes. Participants imitated 20 prerecorded multitarget phrases (10 spoken by a male adult, 10 spoken by a female adult). Each phrase included the target sound pattern 2 times per stimulus (e.g., for /l/ onset, "leave the location"), resulting in 40 attempts at each sound–position.

The untreated phrase probes were scored using the Phon phonetic transcription and database system (Hedlund & Rose, 2022). For each session, three separate transcribers (students in speech-language pathology or SLPs) independently completed narrow phonetic transcriptions. Written and oral instructions were provided on how to complete the transcriptions. Diacritics were used to denote clinical distortions (e.g., lateralized sibilants, derhotacized /ɹ/), but diacritics were not used to indicate normal allophonic variation. Transcribers met with the first author at least monthly to review difficult items they had flagged and to discuss questions about the procedures. Across all participants, probe sessions, sound–positions, and listeners, more than 50,000 narrow transcriptions of words were generated for the primary analysis in this project. Transcribers were allowed to discard an utterance if they judged it to be of insufficient audio quality to transcribe (e.g., background noise). Each probe for each sound was intended to contain 40 targets to be scored, but missing data (due to discarded trials from poor audio quality,

Table 2. Treatment session description and fidelity for the treatment groups.

Variable	Distributed + Ultrasound	Distributed + No Ultrasound	Distributed + Telepractice	Intensive + Ultrasound	Intensive + No Ultrasound	Intensive + Telepractice
Average number per session of Speech Motor Chaining structured practice trials in Periods A–D and random practice	230 (195)	248 (201)	238 (146)	438 (167)	336 (226)	212 (211)
Cumulative intervention intensity	5,050 (3,516)	4,923 (3,806)	4,404 (3,332)	8,689 (2,070)	6,560 (4,092)	4,218 (3,860)
Self-monitoring percent agreement between the participant and clinician during Speech Motor Chaining	70.5 (13.2)	69.7 (15.6)	69.4 (15.6)	82.0 (8.3)	75.5 (8.2)	67.4 (17.4)
Fidelity for percentage of trials per session during phonological awareness training in which performance-related feedback was provided	98.87 (6.2)	93.46 (16.9)	100 (0)	97.32 (6.3)	99.62 (2.0)	99.68 (1.8)
Fidelity for percentage of sessions in which the clinician provided a visual referent during prepractice (rated as yes/no)	91.82	87.80	100	91.52	92.63	99.68
Fidelity for percentage of trials per session in which the appropriate feedback type was provided (i.e., KP/KR, as specified by the datasheet for both SMC and random practice)	98.65 (2.7)	97.95 (5.9)	98.84 (4.0)	97.87 (5.7)	97.80 (7.6)	97.29 (6.2)
Fidelity for percentage of trials per session in which the clinician prompted the participant to self-evaluate as scripted	92.49 (25.8)	99.73 (0.7)	100 (0)	98.13 (7.8)	99.42 (2.0)	100 (0)
Fidelity for percentage of trials per session in which prosodic cues were provided as scripted	99.53 (1.8)	99.49 (0.8)	100 (0)	93.30 (19.9)	99.17 (3.3)	100 (0)

Note. Twenty percent of sessions were rated for fidelity from each participant. Means are shown in each cell, and standard deviations are in parentheses. Cumulative intervention intensity is the total number of Speech Motor Chaining structured practice plus random practice trials (prepractice trials were not tallied). KP = knowledge of performance; KR = knowledge of results; SMC = Speech Motor Chaining.

error in data collection that resulted in fewer utterances elicited, or the participant not having said the target word) resulted in a mean of 39.7 scored targets per transcriber (median = 40, minimum = 34).

The primary outcome measure for the untreated phrase probes was percent correct for their target sound–positions. These narrow transcriptions were queried from the entire narrowly transcribed database using Phon syntax and postprocessed in Python to create a dataset that contained information about each individual probe production of the treated sound, for all participants. This dataset is available on the study's OSF page and included unmasked session information (e.g., participant, data collection time point), production information (e.g., target orthography), International Phonetic Alphabet (IPA) target transcriptions, and IPA actual transcriptions. Target transcriptions and actual transcriptions were compared for each production, and any instance where the main IPA symbol and combining diacritics for the actual transcription was not identical to the target transcription was scored as "incorrect." Each participant received a probe score reflecting the mean percent correct for the target sound–position within that session, averaged across three listeners (Shriberg, Austin, et al., 1997). The dataset containing listener-averaged accuracy is also available on the study's OSF page.

Reliability on untreated phrase probes. Percent correct was computed for each target sound–position for each probe session, for each listener. ICC estimates were calculated to examine interrater reliability using SPSS v27. The ICC (1, k) based on mean ratings (k = 3) for all untreated phrase data using one-way random-effects model was .926, 95% CI [.913, .938]. When separated out by modality, the ICC for face-to-face data was .942, 95% CI [.928, .953], and the ICC for telepractice was .895, 95% CI [.862, .921]. This indicates good-to-excellent reliability in the primary outcome data.

Participant- and Parent-Reported Outcomes

Finally, participants and their parents independently completed surveys before treatment and right after

completing 20 hr of treatment to report on the impact of the child's speech impairment on social and emotional functioning. Questions from a social–emotional questionnaire previously used with children with residual speech errors (Hitchcock et al., 2015) were adapted to inquire more broadly about the impact of CAS. The modifications from the original parent survey included two new questions ("My child is reluctant to read in class," "My child is reluctant to ask or answer questions in class"). Questions from the child-facing survey were modifications of the parent survey and also included questions about whether the child felt as though they knew how to correct their speech errors ("I'm not sure what to do to make my speech sound clear"). Two additional questions were included for children at posttreatment to reflect on the treatment experience: "Coming here helped make my speech clearer" and "I liked coming to speech lessons here." The questions are available on the study's OSF page. Ten 5-point Likert scale responses (from *strongly disagree* to *strongly agree*) were rated by the child participant (e.g., "I have been teased or bullied because of the way my speech sounds"), and 13 items were rated separately by the parent (e.g., "My child has been teased or bullied because of the way his/her speech sounds"). These were numerically coded and centered (i.e., from −2 for *strongly disagree*, 0 for *neutral* to +2 for *strongly agree*). A social–emotional impact score was calculated by averaging responses to the questions separately for children and parents. Higher scores for all items indicated greater socioemotional impact reported from the child or parent. Differences between pretreatment and posttreatment scores were quantified, separately for children and parents, using Wilcoxon signed-ranks tests.

Finally, as is customary during clinical trials, the participants and families were queried after 10 and 20 hr of treatment to examine perception of any side effects of the treatment. The responses were transcribed verbatim from audio recordings. After treatment, participants were also asked open-ended questions about what they liked and did not like about the treatment as well as what helped them and did not help them during treatment.

Primary Data Analysis

Linear mixed-effects models were applied to assess the impact of treatment distribution (intensive vs. distributed), biofeedback (ultrasound vs. no ultrasound), and modality (teletherapy vs. face-to-face) on percent target sounds correct from untreated phrase probes. The dependent variable was the percentage of correct speech sounds at each time point (Weeks 0, 5, 10, and 15). Fixed effects included groups as the combination of treatment distribution, biofeedback and modality, time points, and their interaction. Participants were included as random effects. Appropriate correlation structures between observations from a single subject were selected via visual examination of fitted residual variance–covariance matrix and comparing Akaike information criterion (AIC) values. Time points were modeled as a categorical variable instead of a continuous variable with smaller AIC values. The final model reported here includes each participant's baseline (pretreatment) accuracy at Week 0 as a covariate, using heterogeneous first-order autoregressive variance and correlation structure. No missing values were imputed, and no multiplicity adjustment was applied. Restricted maximum likelihood was used to estimate covariance parameters, and Kenward and Roger's method was used for approximating degrees of freedom (Kenward & Roger, 1997). Specific contrasts were constructed to test the difference between treatment groups in terms of the improvement between baseline and Weeks 5, 10, and 15, respectively. Interaction terms at each time point between treatment distribution and biofeedback were also tested. The primary analysis was a comparison of change in percent correct from pretreatment to the 10-week period (after all participants had completed 20 hr of therapy). Additional analyses were conducted after 5 weeks, in which the intensive group had completed 20 hr and the distributed group had completed 10 hr of treatment, and also at 15 weeks to assess retention following the conclusion of treatment. Furthermore, we examined the effects of treatment in an analysis that holds constant number of treatment sessions rather than number of weeks (i.e., short-term generalization following 20 hr of therapy, plus retention 5 weeks later). All statistical analyses and effect size calculations were performed using SAS (v9.4) statistical software. All tests of significance were claimed at two-tailed $\alpha = .05$.

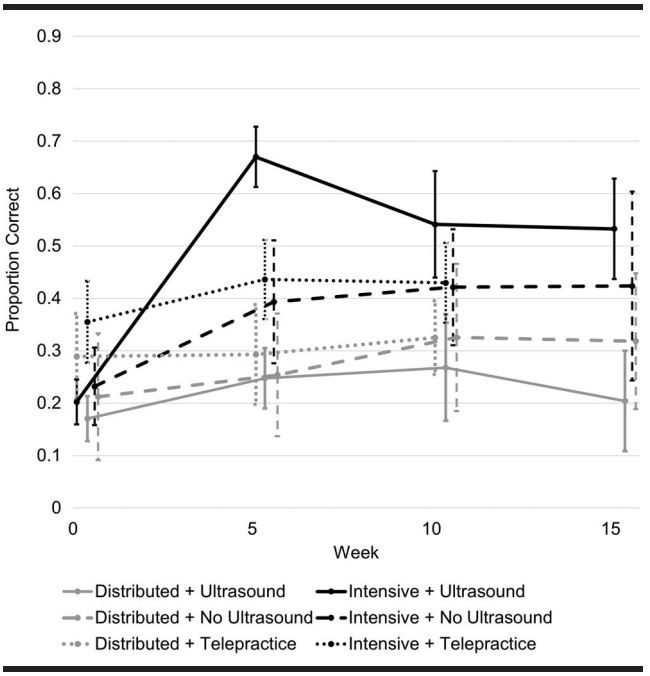
Results

Impact of Treatment Distribution and Biofeedback on Untreated Phrases

Primary Analysis: 10-Week Probe of Untreated Phrases

Figure 3 shows the raw mean accuracy and standard deviations across groups at each time point. The main outcome was measured at 10 weeks from the start of treatment, when both intensive and distributed conditions had completed the treatment package (20 hr of treatment). Note that this included a 5-week no-treatment period for the intensive groups once treatment had finished. Table 3 shows the results of the contrasts between the treatment conditions on the untreated phrase probes.

Figure 3. Mean percent target sounds correct in untreated phrases, by group and by time (±1 standard error).



We first hypothesized that intensive treatment would yield greater speech motor learning than distributed treatment. At the primary 10-week outcome point, there was a significant main effect such that the intensive groups showed larger improvements in percent target sounds correct on untreated phrases than the distributed groups (p = .028, η^2 = .108). This confirmed our first hypothesis.

We also hypothesized that children who received ultrasound biofeedback would show greater improvement in percent target sounds correct on untreated phrases than children who did not receive ultrasound biofeedback. At the primary 10-week outcome point, the main effect comparing ultrasound versus no-ultrasound conditions was not statistically significant (p = .132, η^2 = .052). Thus, there was insufficient evidence to support our second hypothesis at the main outcome point.

In addition, there was no significant interaction between treatment distribution and biofeedback conditions at the 10-week outcome point (p = .125, η^2 = .054). There was also no significant difference between the changes in percent target sounds correct on untreated phrases for telepractice and face-to-face groups at the 10-week outcome point (p = .187, η^2 = .04).

Secondary Analysis: 5-Week Probe of Untreated Phrases

As shown in Figure 3, the intensive groups show larger increases from pretreatment to the 5-week time point than the distributed groups, as would be expected because the intensive groups received 20 hr of treatment in that time while the distributed groups received only 10 hr of treatment. When comparing the improvement in speech sound accuracy at the 5-week point on untreated phrase probe accuracy, there was a significant main effect of treatment distribution (p = .0005, η^2 = .26) such that the intensive groups (who had received 20 hr of treatment) showed greater improvement than the distributed groups (who had received 10 hr of treatment) in percent target sounds correct on untreated phrases. There was also a significant main effect such that the ultrasound groups outperformed the no-ultrasound groups at the 5-week probe (p = .005, η^2 = .179). The interaction between intensity and ultrasound was significant (p = .034, η^2 = .106), suggesting that, at the 5-week probe, the benefit of ultrasound biofeedback may be greater when delivered in an intensive treatment program than in distributed treatment. There was no significant difference between the changes in percent target sounds correct on untreated phrases for telepractice and face-to-face groups at the 5-week outcome point (p = .317, η^2 = .025).

Secondary Analysis: 15-Week Probe of Untreated Phrases

To examine longer-term retention, we also conducted probes at 15 weeks (10 weeks after the intensive groups had completed treatment and 5 weeks after the distributed groups had completed treatment). As shown in Figure 1, this analysis is limited to the participants in each group for whom we collected 15-week follow-up data. When comparing the improvements in percent target sounds correct at the 15-week point on untreated phrases, there was a significant main effect of intensity such that the intensive groups showed greater improvement from baseline than the distributed groups (p = .002, η^2 = .200). There was no main effect of biofeedback (p = .365, η^2 = .019) and no significant interaction between treatment distribution and biofeedback (p = .071, η^2 = .074). This suggests that we had evidence to support the effects of intensive treatment over distributed treatment for longer-term retention, but we did not find evidence to support the effects of ultrasound over no ultrasound on longer-term retention.[1]

[1]Out of clinical interest, we also conducted and analyzed a model that was limited to only participants who completed face-to-face treatment and had /ɹ/ targets (which are commonly targeted by clinicians in treatment and have been studied to the greatest extent in prior research on biofeedback). Identical conclusions were reached for this subsample regarding the main effects and interactions at each time point. A summary of that supplemental analysis is available on the study's OSF page.

Table 3. Comparisons of interest at each time point with 95% confidence interval (CI) from a linear mixed model.

Time (weeks)	Contrast (first condition minus second condition)	Estimate (change score difference)	Lower CI	Upper CI	*df*	*t* value	*p* value	η^2
5	Distributed – intensive	–.190	–.290	–.088	40.1	–3.78	.0005*	.26
5	No ultrasound – ultrasound	–.162	–.271	–.052	40.6	–2.98	.005*	.179
5	Interaction of biofeedback and practice distribution	.235	.0181	.451	40.5	2.19	.034*	.106
5	Telepractice – face-to-face	–.063	–.187	.062	39.9	–1.01	.317	.025
10	Distributed – intensive	–.137	–.259	–.015	42.7	–2.27	.028*	.108
10	No ultrasound – ultrasound	–.100	–.232	.031	43.6	–1.54	.132	.052
10	Interaction of biofeedback and practice distribution	.202	–.059	.463	43.0	1.56	.125	.054
10	Telepractice – face-to-face	–.100	–.249	.050	43.1	–1.34	.187	.040
15	Distributed – intensive	–.220	–.355	–.085	43.1	–3.28	.002*	.200
15	No ultrasound – ultrasound	–.061	–.197	.074	43.3	–0.91	.365	.019
15	Interaction of biofeedback and practice distribution	.124	–.011	.259	43.1	1.85	.071	.074

Note. Estimate = difference in the proportion of target sounds correct on untreated phrases from pretreatment to the indicated Time (weeks) between stated conditions. Estimates are from a linear mixed model with heterogeneous first-order autoregressive correlation and baseline accuracy as a covariate.

*Significant at $p < .05$.

Secondary Analysis: Treatment Effects and Retention While Holding Constant Time Since Completion of Treatment

Whereas the above analyses held constant time since the start of treatment, we also examine the effects of treatment when holding constant the time since the end of treatment. We first examine the effects of treatment during the probe immediately after 20 sessions were completed (i.e., at the 5-week time point for the intensive group and at the 10-week time point for the distributed group). Results, shown in Table 4, indicate greater improvement in untreated phrase accuracy for children who received 20 sessions in intensive treatment over 20 sessions of distributed treatment ($p = .009$, $\eta^2 = .129$) and greater improvement for untreated phrases for children who received 20 sessions in the ultrasound condition over 20 sessions of no ultrasound ($p = .025$, $\eta^2 = .094$). This effect was qualified by a significant interaction between practice distribution and biofeedback ($p = .022$, $\eta^2 = .097$), suggesting the greatest benefit was for children who receive 20 hr of ultrasound in intensive treatment.

To explore retention, we next examined the effects of treatment during the probe 5 weeks after 20 sessions were completed (i.e., at the 10-week time point for the intensive group and at the 15-week time point for the distributed group). There remained significantly greater improvement in untreated phrase accuracy 5 weeks after

Table 4. Comparisons of interest at each point relative to conclusion of therapy.

Period	Contrast	Estimate (change score difference)	Lower CI	Upper CI	*df*	*t* value	*p* value	η^2
Immediate posttreatment	Distributed (at 10 weeks) vs. intensive (at 5 weeks)	–.152	–.264	–.041	50.5	–2.74	.009	.129
Immediate posttreatment	No ultrasound vs. ultrasound	–.139	–.260	–.018	51.8	–2.32	.025	.094
Immediate posttreatment	Interaction of biofeedback and practice distribution	.280	.041	.518	51.7	2.35	.022	.097
5 weeks posttreatment	Distributed (at 15 weeks) vs. intensive (at 10 weeks)	–.159	–.273	–.045	51.5	–2.8	.007	.132
5 weeks posttreatment	No ultrasound vs. ultrasound	–.099	–.223	.025	52.9	–1.6	.115	.046
5 weeks posttreatment	Interaction of biofeedback and practice distribution	.204	–.041	.45	52.5	1.67	.101	.05

Note. Estimate = difference in the proportion of target sounds correct on untreated phrases between stated conditions (change from pretreatment to the indicated time point). Estimates are from a linear mixed model with heterogeneous first-order autoregressive correlation and baseline accuracy as a covariate.

treatment in the intensive group than in the distributed group ($p = .007$, $\eta^2 = .132$). However, there was no longer a significant difference between the ultrasound and no-ultrasound groups ($p = .115$, $\eta^2 = .046$) and no interaction between practice distribution and biofeedback ($p = .101$, $\eta^2 = .05$).

Secondary Outcomes: Participant- and Parent-Reported Outcomes

There were 32 children who completed both pretreatment and posttreatment questionnaires. As shown in Table 5, children's average and median social–emotional impact scores did not appear to differ before or after treatment. A Wilcoxon signed-ranks test was used to quantify this observation; no significant pretreatment to posttreatment differences were found ($S = -78$, $p = .1095$). However, a significant change was observed in impact scores for the 15 parents with pretreatment and posttreatment data ($S = -51.5$, $p = .0002$) with a report of reduced social–emotional impact following treatment. In addition, children's responses to the statement, "I'm not sure what to do to make my speech sound clear," were analyzed to interpret the degree to which the children indicated a better understanding of methods to improve the clarity of their speech production. Results indicated a significant difference between responses to this question before and after treatment, with the median response shifting from "neutral" (0) pretreatment to "disagree" (−1) posttreatment (Wilcoxon signed-ranks $S = -85.5$, $p = .0010$). On the posttreatment questionnaire, children responded to the statement "Coming here helped make my speech clearer" with a median score of 2 (*strongly agree*). In addition, children responded to the statement "I liked coming to speech lessons here" with a median score of 1.25 (between *agree* and *strongly agree*).

Open-ended questions were asked to all participants during selected probe sessions to obtain information about possible side effects during treatment and aspects of the treatment that participants did or did not like. The reported side effects were mild inconveniences such as missing out on other extracurricular activities while in treatment sessions. One exception was a report from a participant in the Distributed + Telepractice treatment condition who indicated that treatment made their "throat hurt." Numerous positive effects were reported, such as "I got better at speaking," "It helped my speech," and "It helped me to pronounce the words I can't pronounce." In addition, several positive comments were made about the usefulness of the ultrasound images from participants who were randomized to that condition. Complete responses are posted on the study's OSF page.

Descriptive Performance

Although the primary outcome was related to the untreated probes, we also explored the within-session performance of the treatment groups. Supplemental description of performance on phonological awareness training, within-session practice, and changes in stimulability are provided in supplemental materials on the study's OSF page. In addition, the average number of Speech Motor Chaining structured practice trials in Periods A–D and random practice each session, as well as average cumulative intervention intensity (Warren et al., 2007), is shown in Table 2. The large standard deviations in average practice trials per session are due to the increasing range of number of trials from Session 1 to Session 20. Variation between sessions, participants, and groups are due in large part to the amount of time children spent in prepractice to establish initial correct productions of the targets (prepractice trials were not tallied due to the dynamic nature of prepractice). Furthermore, during Speech Motor Chaining

Table 5. Pretreatment and posttreatment scores on Social Emotional Questionnaire and additional survey questions.

Variable	Pretreatment score		Posttreatment score	
	Mean	Median	Mean	Median
Child impact score (average of responses to 10 child questions)	–0.43	–0.6	–0.42	–0.6
Parent impact score (average of responses to all 13 parent questions)	0.52	0.58	–0.06	–0.08
Child responses to "I'm not sure what to do to make my speech sound clear"	0.14	0	–1.00	–1
Child responses to posttreatment question: "Coming here helped make my speech clearer"	N/A	N/A	1.35	2
Child responses to posttreatment question: "I liked coming to speech lessons here"	N/A	N/A	0.90	1.25

Note. Scores are coded on a 5-point ordinal scale and were numerically centered from –2 for *strongly disagree* to 0 for *neutral* and +2 for *strongly agree*. N/A = not applicable.

structured practice, participants were prompted to self-evaluate their productions on 50% of trials (before the clinician provided feedback). Agreement between the participant and the clinician on the accuracy of productions is shown in Table 2.

Discussion

This study addressed the hypotheses that both intensive treatment distribution and ultrasound biofeedback would affect speech sound learning in school-age children with CAS. The sample size, although modest, appears to be the largest treatment study on CAS to date. Comparison between the conditions at Week 5 showed that 20 treatment sessions resulted in more improvement in speech sound accuracy than 10 sessions. More important, however, when the number of treatment sessions is held constant (as was the case at our primary 10-week outcome time point), greater speech generalization and retention is observed when 20 treatment sessions were provided in an intensive format (starting with a 1-week "boot camp" followed by a titrated schedule for 4 weeks and then 5 weeks of no treatment) than when the same number of sessions were provided in a traditional distributed schedule of 2 times per week for 10 weeks. In addition to evidencing the advantage of intensive treatment over distributed treatment at our primary outcome time point, this study provides evidence that the advantage of intensive treatment remains in the weeks following the conclusion of treatment for all groups.

In addition, ultrasound biofeedback was found to be temporarily advantageous to facilitate generalization in both intensive and distributed formats, but after the end of treatment, the advantage over non-biofeedback treatment may be reduced. In other words, the amount of generalized speech motor learning demonstrated immediately following intensive treatment appears to be greater in children who received biofeedback than children who did not receive biofeedback, but after additional time has passed after the end of treatment, the level of learning appears statistically similar whether or not ultrasound was used. This suggests that biofeedback enables short-term generalization but treatment likely needs to continue or adapt to maximize the gains brought by biofeedback.

Theoretical Implications

The schema-based PML framework (Maas et al., 2008), a nonspeech framework that has broadly been used to guide investigation of several treatments for CAS, predicts greater generalized motor learning with distributed treatment than with massed (intensive) treatment and greater generalized motor learning without biofeedback than with biofeedback. However, we predicted that the different considerations may be warranted in the unique context of motor learning (a) in children (b) with low skill level (c) on a speech task. The results of this study, which are generally in line with prior studies on massed versus distributed treatment in childhood speech disorders (Allen, 2013; Maas et al., 2019; Thomas et al., 2014), show that an intensive approach that first emphasizes acquisition through a week of massed treatment before fading to a distributed treatment schedule yields greater speech motor learning than uniformly distributed treatment alone for children with persisting speech errors associated with CAS. There are several possible reasons for this. First, it is plausible that the principles derived from research on *nonspeech* motor learning do not apply to *speech* motor learning in children with CAS and that revised theoretical frameworks emphasizing different principles for speech motor learning are needed. Second, motor learning may be a function of success during an accumulation of practice attempts (Mattar & Ostry, 2007), and it may be difficult for children with CAS to achieve sufficient cumulative successful practice due to use of the erred motor plan when speaking between treatment sessions. Because it is in the early stages of practicing a skill that this interference is thought to be most influential (Luft & Buitrago, 2005), an intensive program may help to *minimize the between-practice use of the erred motor plan* and enable a more favorable ratio of correct productions (primarily during treatment sessions) to incorrect productions (during treatment and outside of treatment) during intensive treatment. In essence, closely spaced sessions may help to limit the opportunity for between-session interference of the updated motor plan with the erred motor plan. Traditional treatment schedules such as two sessions per week may be too distributed in time for efficient updating of the motor plan, allowing for forgetting or learning decay between sessions (Murre & Dros, 2015).

With respect to ultrasound biofeedback, at the primary outcome point (10 weeks) and at the follow-up point (15 weeks), the slight difference between the treatment conditions with and without ultrasound biofeedback did not reach statistical significance. It is possible, then, that ultrasound biofeedback may not provide much overall advantage in treatment for children with CAS. However, the main effect of ultrasound at the 5-week time point suggests that ultrasound biofeedback may be temporarily advantageous for children during treatment but that this advantage may be lost in the weeks after treatment ends. This evidence suggests that speech motor learning may be affected by different timescales of measurement—that is, short-term improvement for children receiving ultrasound biofeedback (i.e., the 5-week probe) indicates that

biofeedback provided a boost at the beginning of treatment. However, biofeedback did not adequately guard against interference or forgetting over a longer timescale after treatment has ended (as evidenced at the 10-week and 15-week time points). Although we found no evidence that detailed feedback of the tongue impedes learning, as predicted by some nonspeech motor learning literature (Hodges & Franks, 2001), we found that biofeedback supported an initial gain that was not fully sustained over time. There are several possible reasons. One reason might be that we did not provide enough biofeedback in this study to foster long-term retention of new motor plans (i.e., less than half of practice attempts included biofeedback during sessions, and the biofeedback was front-loaded across the sessions and then faded), as research on ultrasound biofeedback in children with residual speech errors has indicated that more biofeedback practice trials may yield greater outcomes than fewer practice trials with biofeedback (Preston et al., 2018). Alternatively, it is possible that biofeedback only facilitates a short-term improvement in articulatory knowledge that is not sustained because children need further practice without biofeedback to fully establish the new motor plan to enable permanent change. This could be framed within a feedforward/feedback model of speech motor control (Guenther & Vladusich, 2012) as biofeedback perhaps facilitating the acquisition of corrected feedback of auditory/somatosensory target maps but not directly facilitating the feedforward selection of the updated motor plan, which might require repeated practice for habituation. Another possibility is that the biofeedback provides a novel element that enhances motivation early on during treatment and during early probes, but the motivational effects dissipate as novelty wears off. Finally, it is also plausible that biofeedback interacts with participant-level characteristics, as biofeedback may be best suited for individuals who are less stimulable (McAllister et al., 2022).

Child- and Parent-Reported Outcomes

The lack of change in the children's socioemotional impact scores before and after treatment suggests that 5–10 weeks of treatment may not be enough to significantly improve a child's self-perceptions of the social–emotional impact of their speech difficulties. This finding highlights that it may take much longer than 5–10 weeks to observe measurable changes in child-reported socioemotional functioning. Results may point to a potential need for further treatment or for counseling strategies to address the socioemotional impact of CAS. In contrast to the children's self-report, the parents' responses to the questionnaire indicated a perceived reduction in social–emotional impact following treatment. This could reflect either an important effect of treatment that parents observe that children fail to notice or a placebo effect, as parents were not masked to the fact that their child was receiving treatment. As only 15 parents provided both pretreatment and posttreatment responses, inferences about their perception of socioemotional impact should be interpreted with caution.

Clinical Implications

There are several clinical implications from this study. First, the Speech Motor Chaining treatment program described here, which has freely available resources and can be delivered with high fidelity, can be considered an evidence-based option for the treatment of persisting speech errors in the context of CAS. There may be a combination of factors, such as explicit KP feedback, many practice trials, systematic introduction of motor learning principles, or phonological awareness training, that contribute to the overall effectiveness of the approach.

The primary clinical recommendation stemming from these findings is that for school-age children with CAS, treatment may be more effective if it is delivered, at least temporarily, in a more intensive fashion than in a distributed fashion. This recommendation aligns with prior work suggesting the benefits of more closely spaced practice sessions (Kaipa & Peterson, 2016; Maas et al., 2019). Although there are clear practical challenges with scheduling intensive treatment followed by a titrated schedule, the evidence herein raises the question if such programs may ultimately be more cost-effective and reduce caseload sizes if children progress through treatment faster. Studies on overcoming implementation barriers and on cost-effectiveness of such treatment models in various settings will be important future work.

The clinical implications for the use of ultrasound biofeedback are less clear, as no statistical difference between biofeedback and non-biofeedback groups was observed at our study's predefined primary time point (10 weeks). Statistically significant differences, however, in the main effect of biofeedback were seen at 5 weeks, indicating that the general advantage of biofeedback may be an initial boost that is not sustained if treatment continues or is withdrawn. This suggests that if ultrasound is used initially to teach new articulatory movements, withdrawing treatment too soon may eliminate any initial gains, even in a treatment protocol where ultrasound was faded during the last four sessions to purposefully facilitate generalization. Furthermore, in our analysis that examined outcomes at the probe point immediately after the completion of 20 sessions (i.e., after 5 weeks for the intensive group and after 10 weeks for the distributed group), there was a significant advantage of ultrasound over no ultrasound, as well as a significant interaction suggesting that the group that had both ultrasound and intensive

treatment made the greatest progress. Thus, ultrasound can facilitate generalization. However, because this advantage of ultrasound over no ultrasound is not significant 5 weeks later, additional considerations may be needed to facilitate retention.

A holistic view of these implications would suggest that intensive treatment maximizes motor learning compared to more distributed schedules for children and adolescents with persisting errors associated with CAS, and there may be some possible benefits of including biofeedback for short-term gains. Because this study was designed to compare distributed and intensive schedules following a fixed number of sessions (20), we cannot extrapolate what the differences between groups might have looked like if more intervention was provided (e.g., if the Intensive + Ultrasound group had more than 5 weeks of treatment). Further research extending over a longer duration may be needed to determine if the initial momentum of an intensive schedule with ultrasound biofeedback, visualized in Figure 3, can be maintained or extended with a more individualized and adaptive treatment arc, instead of the rigid treatment plans required for experimental control during a randomized controlled trial.

Caveats and Limitations

There are several caveats that warrant attention. Diagnostically, "expert" judgment is common for clinical diagnosis in CAS (ASHA, 2007); however, we specifically avoided using clinical impression because it is unclear how reliable this procedure is, and it would be impossible for future work to replicate the inclusionary criteria if it relied on clinical impression. Therefore, while our diagnostic procedures are based on methods from prior studies and showed good interrater reliability (as well as validity with respect to the Mayo-10 criteria), the approach differs from common clinical practice. It is also the case that, to limit variability, children with comorbid conditions such as autism and hearing loss were excluded. Therefore, the results are generalizable to the subset of children with CAS who met our inclusionary criteria.

The interruptions due to the COVID-19 pandemic resulted in protocol modifications and fewer face-to-face participants than expected. We therefore had an imbalance in participants in the ultrasound versus no-ultrasound conditions, which limits statistical power. While the no-ultrasound treatment procedures were portable to telepractice, and there were no significant differences between the telepractice and face-to-face groups, it is unknown to what extent issues such as videoconferencing fatigue during the pandemic, behavior management, client–clinician rapport, audio transmission issues, and other factors may differ in children who completed treatment via telepractice. Furthermore, caution should be taken in interpreting the lack of difference between face-to-face and telepractice treatment because participants were not randomized to face-to-face or telepractice sessions, limiting the conclusions that might be drawn from such a comparison.

Finally, the treatment program here is one that primarily emphasizes speech sound accuracy, practicing from syllables through sentences. Although prosodic variation is included in treatment, the current structure of Speech Motor Chaining is one that teaches children to accurately produce speech sounds, and there is not a direct focus on teaching other features such as lexical stress. Other approaches, such as ReST training (McCabe et al., 2020, 2023; Thomas et al., 2016), would likely be more appropriate to address other elements of speech production.

Conclusions

Children with CAS were found to improve their speech sound accuracy with 20 sessions of a Speech Motor Chaining treatment program. There was greater improvement observed among participants who began treatment with a 1-week intensive "boot camp" than among participants who steadily received 20 treatment sessions at a rate of twice per week. This advantage of intensive treatment was seen when comparing participants after an equivalent number of sessions, after an equivalent number of weeks, and in the weeks after both groups had experienced a period of no treatment. Ultrasound biofeedback initially enhanced treatment outcomes for participants, but the advantage of biofeedback was not sustained at the primary outcome time point. Future research on translational strategies to implement the current findings in clinical settings, as well as other treatment options, will be valuable to reduce the impact of CAS on school-age children.

Data Availability Statement

De-identified data, data analysis code, supplemental description of performance, and supplemental treatment materials (videos, datasheets) are available on the study's Open Science Framework page (https://osf.io/r872p/).

Acknowledgments

This study was supported by National Institutes of Health (NIH) Grant R15DC016426 (principal investigator: J. L. Preston), the Syracuse University College of Arts

and Sciences, and a Syracuse University Collaboration for Unprecedented Success and Excellence COVID-Relief grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. The authors express gratitude to study participants and their families, as well as to Raya Campeas, Claire Cooke, Olivia Flaherty, Kelly Garcia, Leah Haas, Harris, Rachel Koury, Maya Kumar, Morgan McLenithan, Kerry McNamara, Kylee Peterson, Cora Pokrifka, Caitlin (Vose) Raaz, Stephanie Reeves, Hanna Schmitz, Emily Sargent, Martine Schultheiss, Greta Sjolie, Kelsey Spradlin, and Connor Walsh for assisting with the project.

References

Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology, 16*(2), 128–139. https://doi.org/10.1044/1058-0360(2007/017)

Allen, M. M. (2013). Intervention efficacy and intensity for children with speech sound disorder. *Journal of Speech, Language, and Hearing Research, 56*(3), 865–877. https://doi.org/10.1044/1092-4388(2012/11-0076)

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. https://www.asha.org/policy/tr2007-00278/

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

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

Carrigg, B., Parry, L., Baker, E., Shriberg, L. D., & Ballard, K. J. (2016). Cognitive, linguistic, and motor abilities in a multigenerational family with childhood apraxia of speech. *Archives of Clinical Neuropsychology, 31*(8), 1006−1025. https://doi.org/10.1093/arclin/acw077

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Cleland, J., Scobbie, J. M., Roxburgh, Z., Heyde, C., & Wrench, A. (2019). Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback. *Journal of Speech, Language, and Hearing Research, 62*(2), 229–246. https://doi.org/10.1044/2018_JSLHR-S-17-0360

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics, 29*(8–10), 575–597. https://doi.org/10.3109/02699206.2015.1016188

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Pearson.

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation–Third Edition*. Pearson.

Gozzard, H., Baker, E., & McCabe, P. (2006). Children's productions of polysyllables. *ACQuiring Knowledge in Speech, Language and Hearing, 8*(3), 113–116.

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Hedlund, G., & Rose, Y. (2022). *Phon* (Version 3.4). https://www.phon.ca/

Hickok, G. (2014). Towards an integrated psycholinguistic, neurolinguistic, sensorimotor framework for speech production. *Language, Cognition and Neuroscience, 29*(1), 52–59. https://doi.org/10.1080/01690965.2013.852907

Hitchcock, E. R., Harel, D., & McAllister Byun, T. (2015). Social, emotional, and academic impact of residual speech errors in school-aged children: A survey study. *Seminars in Speech and Language, 36*(4), 283–294. https://doi.org/10.1055/s-0035-1562911

Hodges, N. J., & Franks, I. M. (2001). Learning a coordination skill: Interactive effects of instruction and feedback. *Research Quarterly for Exercise and Sport, 72*(2), 132–142. https://doi.org/10.1080/02701367.2001.10608943

Kaipa, R., & Peterson, A. M. (2016). A systematic review of treatment intensity in speech disorders. *International Journal of Speech-Language Pathology, 18*(6), 507–520. https://doi.org/10.3109/17549507.2015.1126640

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983–997. https://doi.org/10.2307/2533558

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Laganaro, M., Croisier, M., Bagou, O., & Assal, F. (2012). Progressive apraxia of speech as a window into the study of speech planning processes. *Cortex, 48*(8), 963–971. https://doi.org/10.1016/j.cortex.2011.03.010

Lawson, E., Stuart-Smith, J., Scobbie, J. M., & Nakai, S. (2018). *Seeing Speech: An articulatory web resource for the study of phonetics*. University of Glasgow.

Lee, S. A. S. (2018). The treatment efficacy of multiple opposition phonological approach via telepractice for two children with severe phonological disorders in rural areas of West Texas in the USA. *Child Language Teaching and Therapy, 34*(1), 63–78. https://doi.org/10.1177/0265659018755527

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59*(4), 277–287. https://doi.org/10.1080/02701367.1988.10609373

Lewis, B. A., Benchek, P., Tag, J., Miller, G., Freebairn, L., Taylor, H. G., Iyengar, S. K., & Stein, C. M. (2021). Psychosocial comorbidities in adolescents with histories of childhood apraxia of speech. *American Journal of Speech-Language Pathology, 30*(6), 2572–2588. https://doi.org/10.1044/2021_ajslp-21-00035

Lewis, B. A., Freebairn, L., Tag, J., Benchek, P., Morris, N. J., Iyengar, S. K., Taylor, H. G., & Stein, C. M. (2018). Heritability and longitudinal outcomes of spelling skills in individuals with histories of early speech and language disorders. *Learning and Individual Differences, 65,* 1–11. https://doi.org/10.1016/J.LINDIF.2018.05.001

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

Services in Schools, 35(2), 122–140. https://doi.org/10.1044/0161-1461(2004/014)

Luft, A. R., & Buitrago, M. M. (2005). Stages of motor skill learning. *Molecular Neurobiology, 32*(3), 205–216. https://doi.org/10.1385/mn:32:3:205

Maas, E., Butalla, C. E., & Farinella, K. A. (2012). Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 21*(3), 239–257. https://doi.org/10.1044/1058-0360(2012/11-0119)

Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(2), 561−578. https://doi.org/10.1044/1092-4388(2011/11-0120)

Maas, E., Gildersleeve-Neumann, C., Jakielski, K., Kovacs, N., Stoeckel, R., Vradelis, H., & Welsh, M. (2019). Bang for your buck: A single-case experimental design study of practice amount and distribution in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 62*(9), 3160–3182. https://doi.org/10.1044/2019_JSLHR-S-18-0212

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Mattar, A. A., & Ostry, D. J. (2007). Neural averaging in motor learning. *Journal of Neurophysiology, 97*(1), 220–228. https://doi.org/10.1152/jn.00736.2006

McAllister, T., Eads, A., Kabakoff, H., Scott, M., Boyce, S., Whalen, D. H., & Preston, J. L. (2022). Baseline stimulability predicts patterns of response to traditional and ultrasound biofeedback treatment for residual speech sound disorder. *Journal of Speech, Language, and Hearing Research, 65*(8), 2860–2880. https://doi.org/10.1044/2022_JSLHR-22-00161

McAllister Byun, T., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57*(6), 2116−2130. https://doi.org/10.1044/2014_JSLHR-S-14-0034

McCabe, P., Macdonald-D'Silva, A. G., van Rees, L. J., Ballard, K. J., & Arciuli, J. (2014). Orthographically sensitive treatment for dysprosody in children with childhood apraxia of speech using ReST intervention. *Developmental Neurorehabilitation, 17*(2), 137–145. https://doi.org/10.3109/17518423.2014.906002

McCabe, P., Preston, J. L., Evans, P., & Heard, R. (2023). A pilot randomized control trial of motor-based treatments for childhood apraxia of speech: Rapid Syllable Transition Treatment and ultrasound biofeedback.. *American Journal of Speech-Language Pathology, 32*(2), 629–644. https://doi.org/10.1044/2022_AJSLP-22-00144

McCabe, P., Thomas, D. C., & Murray, E. (2020). Rapid Syllable Transition Treatment—A treatment for childhood apraxia of speech and other pediatric motor speech disorders. *Perspectives of the ASHA Special Interest Groups, 5*(4), 821–830. https://doi.org/10.1044/2020_persp-19-00165

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009a). Effectiveness of an integrated phonological awareness approach for children with childhood apraxia of speech (CAS). *Child Language Teaching and Therapy, 25*(3), 341–366. https://doi.org/10.1177/0265659009339823

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009b). A longitudinal case study of the effects of an integrated phonological awareness program for identical twin boys with childhood apraxia of speech (CAS). *International Journal of Speech-Language Pathology, 11*(6), 482–495. https://doi.org/10.3109/17549500902842583

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009c). Phonological awareness and early reading development in childhood apraxia of speech (CAS). *International Journal of Language & Communication Disorders, 44*(2), 175–192. https://doi.org/10.1080/13682820801997353

Moriarty, B. C., & Gillon, G. T. (2006). Phonological awareness intervention for children with childhood apraxia of speech. *International Journal of Language & Communication Disorders, 41*(6), 713–734. https://doi.org/10.1080/13682820600623960

Murray, E., McCabe, P., & Ballard, K. J. (2014). A systematic review of treatment outcomes for children with childhood apraxia of speech. *American Journal of Speech-Language Pathology, 23*(3), 486–504. https://doi.org/10.1044/2014_AJSLP-13-0035

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(1), 43–60. https://doi.org/10.1044/2014_JSLHR-S-12-0358

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE, 10*(7), Article e0120644. https://doi.org/10.1371/journal.pone.0120644

Namasivayam, A. K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., Maassen, B., Kroll, R., & van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech. *International Journal of Language & Communication Disorders, 50*(4), 529–546. https://doi.org/10.1111/1460-6984.12154

Nijland, L. (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics, 23*(3), 222–239. https://doi.org/10.1080/02699200802399947

Nijland, L., Maassen, B., van der Meulen, S., Gabreels, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 17*(1), 1–24. https://doi.org/10.1080/0269920021000050662

Peterson, L., Savarese, C., Campbell, T., Ma, Z., Simpson, K. O., & McAllister, T. (2022). Telepractice treatment of residual rhotic errors using app-based biofeedback: A pilot study. *Language, Speech, and Hearing Services in Schools, 53*(2), 256–274. https://doi.org/10.1044/2021_LSHSS-21-00084

Preston, J. L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology, 22*(4), 627–643. https://doi.org/10.1044/1058-0360(2013/12-0139)

Preston, J. L., & Leece, M. C. (2017). Intensive treatment for persisting rhotic distortions: A case series. *American Journal of Speech-Language Pathology, 26*(4), 1066–1079. https://doi.org/10.1044/2017_AJSLP-16-0232

Preston, J. L., Leece, M. C., & Maas, E. (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience, 10,* Article 440. https://doi.org/10.3389/fnhum.2016.00440

Preston, J. L., Leece, M. C., McNamara, K., & Maas, E. (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology, 26*(3), 840–852. https://doi.org/10.1044/2017_AJSLP-16-0155

Preston, J. L., Leece, M. C., & Storto, J. (2019). Tutorial: Speech Motor Chaining treatment for school-age children

with speech sound disorders. *Language, Speech, and Hearing Services in Schools, 50*(3), 343–355. https://doi.org/10.1044/2018_LSHSS-18-0081

Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., & Whalen, D. H. (2018). Treatment for residual rhotic errors with high- and low-frequency ultrasound visual feedback: A single-case experimental design. *Journal of Speech, Language, and Hearing Research, 61*(8), 1875–1892. https://doi.org/10.1044/2018_JSLHR-S-17-0441

Preston, J. L., McAllister, T., Phillips, E., Boyce, S., Tiede, M., Kim, J. S., & Whalen, D. H. (2019). Remediating residual rhotic errors with traditional and ultrasound-enhanced treatment: A single-case experimental study. *American Journal of Speech-Language Pathology, 28*(3), 1167–1183. https://doi.org/10.1044/2019_AJSLP-18-0261

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., & Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *Journal of Visualized Experiments, 2017*(119), Article e55123. https://doi.org/10.3791/55123

Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., & Maas, E. (2014). Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research, 57*(6), 2102–2115. https://doi.org/10.1044/2014_JSLHR-S-14-0031

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37*(5), 401–411. https://doi.org/10.1016/j.jcomdis.2004.04.003

Rogers, M. A., & Storkel, H. L. (1998). Reprogramming phonologically similar utterances: The role of phonetic features in premotor encoding. *Journal of Speech, Language, and Hearing Research, 41*(2), 258–274. https://doi.org/10.1044/jslhr.4102.258

Rusiewicz, H. L., Maize, K., & Ptakowski, T. (2017). Parental experiences and perceptions related to childhood apraxia of speech: Focus on functional implications. *International Journal of Speech-Language Pathology, 20*(5), 569–580. https://doi.org/10.1080/17549507.2017.1359333

Rvachew, S., Hodge, M., & Ohberg, A. (2005). Obtaining and interpreting maximum performance tasks from children: A tutorial. *Journal of Speech-Language Pathology and Audiology, 29*(4), 146–157.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: III. A subtype marked by inappropriate stress. *Journal of Speech, Language, and Hearing Research, 40*(2), 313–337. https://doi.org/10.1044/jslhr.4002.313

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708–722. https://doi.org/10.1044/jslhr.4004.708

Shriberg, L. D., & Lohmeier, H. L. (2008). *The syllable repetition task* [Technical report no. 14]. http://www.waisman.wisc.edu/phonology/(14)

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 26*(5), 445–482. https://doi.org/10.3109/02699206.2012.655841

Shriberg, L. D., Potter, N. L., & Strand, E. A. (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech, Language, and Hearing Research, 54*(2), 487–519. https://doi.org/10.1044/1092-4388(2010/10-0068)

Strand, E. A. (2020). Dynamic Temporal and Tactile Cueing: A treatment strategy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 29*(1), 30–48. https://doi.org/10.1044/2019_AJSLP-19-0005

Thomas, D. C., McCabe, P., & Ballard, K. J. (2014). Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech: The effect of lower dose-frequency. *Journal of Communication Disorders, 51,* 29–42. https://doi.org/10.1016/j.jcomdis.2014.06.004

Thomas, D. C., McCabe, P., Ballard, K. J., & Lincoln, M. (2016). Telehealth delivery of Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech. *International Journal of Language & Communication Disorders, 51*(6), 654–671. https://doi.org/10.1111/1460-6984.12238

Thoonen, G., Maassen, B., Gabreels, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13*(1), 1–23. https://doi.org/10.1080/026992099299211

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. R. (2013). *Comprehensive Test of Phonological Processing–Second Edition*. Pro-Ed.

Wales, D., Skinner, L., & Hayman, M. (2017). The efficacy of telehealth-delivered speech and language intervention for primary school-age children: A systematic review. *International Journal of Telerehabilitation, 9*(1), 55–70. https://doi.org/10.5195/IJT.2017.6219

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II)*. Pearson.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Pearson.