Research Article

Differences Between School-Age Children With Apraxia of Speech and Other Speech Sound Disorders on Multisyllable Repetition

Nina R. Benway[a] and Jonathan L. Preston[a]

Purpose: The aim of the study was to evaluate whether features of childhood apraxia of speech (CAS) identified in previous literature could be replicated in a sample of school-age children.
Method: A literature review was conducted to identify candidate speech features that have been previously considered when differentiating CAS from other types of speech sound disorders (SSDs). The candidate features recoverable from blinded transcriptions of multisyllable word repetitions (MSWRs) were applied to a cohort of 61 children aged 7–17 years, previously classified as having CAS (n = 21) or non-CAS SSD (n = 40).
Results: One hundred and ninety-four features had been explored in previous literature to assess their ability to differentiate CAS from other SSDs. Fifteen perceptual features were selected from this list to be applied to performance on the MSWR. In this sample, children with CAS differed from children with SSD on the prevalence of voicing changes, percentage of structurally correct words, correct lexical stress, and syllable deletions within a speech corpus derived from the MSWR task.
Conclusion: Although previous literature points to numerous features as differentiating CAS from other SSDs, only a portion of those features were replicated in this sample of school-age children. Features of CAS that affect segmental accuracy, prosody, and word structure may be likely to persist into late childhood and early adolescence.

Childhood apraxia of speech (CAS) is a distinct subtype of pediatric speech sound disorder (SSD) characterized by segmental and prosodic features that are either associated with known neurological conditions (e.g., intrauterine stroke, infections), complex neurobehavioral disorders (e.g., genetic, metabolic), or idiopathic presentation (American Speech-Language-Hearing Association [ASHA], 2007). CAS is characterized by difficulty with the planning of precise, consistent speech movements that are coordinated in space and time, resulting in segmental and prosodic errors; however, there remains no fully validated diagnostic criteria or assessment battery to clinically identify individuals within this population. A nonexhaustive list of consensus speech features included in a technical report of ASHA (2007) include inappropriate lexical stress on multisyllabic words and phrases, segmental inconsistency between multiple repetitions of the same word, and lengthened and disrupted coarticulatory transitions between sounds and syllables. Despite recent improvements in sensitivity and specificity of assessment procedures in specific populations of children with CAS (Strand et al., 2013), expert clinician judgment remains the gold standard of diagnosis for the larger population. There remains a need for empirical support and replication of speech characteristics that may differentiate children with CAS from other types of SSD, including in school-age children. The purpose of this study, therefore, is to elucidate which candidate speech features from the body of SSD differential diagnostic literature retrodict CAS diagnosis from multisyllabic word repetitions produced by school-age children with CAS and SSD.

[a]Department of Communication Sciences & Disorders, Syracuse University, NY

Correspondence to Nina R. Benway: nrbenway@syr.edu

Editor: Peter Meulenbroek

Received September 13, 2019
Revision received January 20, 2020
Accepted February 17, 2020
https://doi.org/10.1044/2020_PERSP-19-00086
Publisher Note: This article is part of the Forum: Advances in Research and Clinical Management of Clinical Apraxia in Speech.

Disclosures
Financial: Nina R. Benway has no relevant financial interests to disclose. Jonathan L. Preston has no relevant financial interests to disclose.
Nonfinancial: Nina R. Benway has no relevant nonfinancial interests to disclose. Jonathan L. Preston has no relevant nonfinancial interests to disclose.

Multisyllables

Multisyllabic word repetition tasks require children to plan and execute segmental and suprasegmental aspects of articulation based on an external phonological representation of a stimulus.[1] Multisyllable speech production tasks are common in assessments, and such tasks have previously been found to differentiate the speech of children with phonological disorders from children with typical phonologies (Lewis & Freebairn, 1992), children with residual speech sound errors from children with typical speech sound skills (Preston & Edwards, 2007), children with SSD from children with typical development and language impairment (Leitão et al., 1997), and children with CAS versus SSD (Lewis et al., 2004; Murray et al., 2015).

The assessment of multisyllabic word productions, which has garnered a strong evidence base (Masso et al., 2018), is essential to detect both the segmental and suprasegmental speech features that characterize CAS. Even though most children with typically developing speech acquire most phonemes of American English by the age of 5 years (McLeod & Crowe, 2018), accurate multisyllabic word production requires appropriate lexical stress patterns and syntagmatic consonant and vowel mastery that continues to develop after the acquisition of phonemes in a canonical form. Multisyllable production is believed to mature in typical speakers in late childhood and is impacted by both segmental and prosodic features of the target word (James et al., 2008); therefore, occasional errors in multisyllable segmental accuracy and timing may also be seen in typically developing children and in children with a range of SSD presentations. This presents a challenge in the differential diagnosis of CAS, as children without speech motor planning deficits will also present with errors on multisyllabic word repetition tasks.

Perceptual Speech Features

Narrow phonetic transcription, acoustic analysis, and kinematic analysis are three different methods for encoding speech information for later use in clinical decisions such as differential diagnosis, treatment planning, and progress monitoring. Narrow phonetic transcription requires a transcriber to parse a speech stream into a series of phones represented by a standardized set of main symbols and diacritics (e.g., International Phonetic Association, 2015). Acoustic analysis necessitates frequency, intensity, or duration measurements of the complex wave associated with speech. Kinematic studies involve the use of instrumentation (e.g., Smith, 2006) to record displacements, timing, and movement patterning of the articulators. All three methods of analysis have been found to capture the speech features of children with SSD and CAS (e.g., Case & Grigos, 2016; Shriberg et al., 2010a, 2010b).

These three methodologies have been utilized to distinguish the speech of children with CAS from those who do not have CAS (Terband et al., 2019). Kinematic studies have shown that movement variability and temporal control differ between children with CAS and children with SSD, particularly as word length increases (Grigos et al., 2015). Additionally, acoustic analysis of speech features, such as inappropriate between-word pauses, may be one method for differentiating CAS from SSD and indexing the severity of CAS (Shriberg et al., 2017a, 2017b, 2017c). The current exploration, however, focuses on perceptual features that can be recovered through narrow transcription of segmental and suprasegmental features in order to emphasize the connection between perceptual features and functional clinical outcomes.

While acoustic measurements are able to capture fine-grained artifacts of disordered speech production, the process of phonetic transcription involves generalizing variable acoustic input to a perceptual "category"—such as a phoneme—rather than "identifying" the exact properties of an acoustic signal (e.g., Holt & Lotto, 2010). Thus, acoustic analysis has the potential to capture between-speakers and within-speaker variability that has no functional impact on human speech perception. Furthermore, phonetic transcription is a tool that may readily be at the hands of clinicians, whereas acoustic or kinematic analyses remain uncommon in clinical settings. Features that can be captured during narrow phonetic transcription have, by definition, perceptual significance and may have a more direct impact on functional clinical outcomes, such as target selection and intelligibility.

Perceptual speech features have been commonly explored in the CAS differential diagnostic literature. For example, Murray et al. (2015) sought to identify perceptual features of CAS in a sample of individuals 4–12 years of age referred by community-based speech-language pathologists. Their analysis showed that percentage of lexical stress matches and occurrences of syllable segregation on the Single-Word Test of Polysyllables (Gozzard et al., 2008) accounted for 82% of the variance in CAS diagnoses in the sample. When individuals with comorbid conditions other than CAS were excluded from the sample, a model including percentage of phonemes correct and accuracy on diadochokinetic tasks as well as percentage of lexical stress matches and occurrences of syllable segregation accounted for 91% of the variance of CAS diagnosis in the sample. Thus, speech features such as lexical stress match and percent phonemes correct are characteristics that should continue to be explored to differentiate CAS from other types of SSD.

Purpose

The purpose of this study was to consolidate and extend the literature related to the differential diagnosis of CAS by determining if there are perceptual features

[1]Historically within the literature, words with more than one syllable have been referred to as "multisyllables" (e.g., Dollaghan et al., 1993) or "polysyllables" (e.g., Klein, 1981) without a clearly defined distinction between the two terms; this article will use the more common term "multisyllables" while summarizing the literature related to both multisyllabic and polysyllabic productions, except when the term "polysyllable" is used as a proper noun.

embedded in multisyllabic word repetitions that can distinguish between an existing clinical sample of school-age children with CAS and non-CAS SSD. The following research questions are addressed:

1. What perceptual features of multisyllabic word repetitions are candidate features for differential diagnosis of CAS and non-CAS SSD within the literature?
2. How well do candidate perceptual features, extracted from multisyllabic word repetitions, discriminate between groups of children with CAS diagnosis and another subtype of SSD in children and adolescents between the ages of 7 and 17 years?

Method: Research Question 1

To answer the first research question, a review of existing peer-reviewed, published differential diagnostic literature was undertaken to identify features that have been previously investigated for distinguishing CAS or comorbid CAS (referred to as "CAS+") from other subtypes of SSD. Candidate features were included in the review regardless of whether the original author(s) found statistical or qualitative support for this purpose. Additionally, this review did not limit candidate features based on the statistical analysis methods utilized in the original studies. Studies utilizing only a typically developing comparison group were not included in this review; only those which contrasted CAS from other types of speech disorders were included. The methodology for the review is summarized in Figure 1.

The literature review included all CAS diagnostic studies known to the authors, studies cited in the references of those articles, and a literature search via Google Scholar; however, a full systematic review was not undertaken. This search yielded 422 possible candidate features from 15 peer-reviewed investigations of CAS (or an equivalent term, such as "developmental apraxia of speech," "developmental verbal dyspraxia," or "apraxia of speech"). Additional speech features (n = 5) observed in ongoing projects at the Syracuse University Speech Production Lab were included as exploratory features, resulting in a total of 427 candidate features to begin the exploration. This sum included exact duplicates in which the same set of features with the same operationalization were utilized in multiple group comparisons within the same study (e.g., to differentiate CAS vs. SSD and CAS+ vs. SSD) or within a series of studies by the same authors. A total of 233 exact duplicates were removed from the list of potential candidate features. The remaining (n = 194) nonexact duplicate candidate features were filtered to 135 items that could be measured perceptually through narrow transcriptions; all nonperceptual measures (n = 59) were excluded at this point. If a feature was described in a way that it could be measured either nonperceptually or perceptually (e.g., vowel duration), the feature was considered to be perceptual and not removed from the set of features at this time.

The remaining 135 perceptual candidate features were screened to identify the features that could be derived from the study's narrow phonetic transcription conventions during the production of multisyllabic words. A total of 65 perceptual features that could not be recovered from the corpus were removed, such as diadochokinetic rate tasks. This resulted in 70 features that could be derived from narrow phonetic transcription of multisyllabic words.

The remaining 70 features were coded into seven subgroups in order to identify and consolidate nonexact redundancies between features. These seven subgroups represented both broad categories of speech errors and errors typical of individuals with CAS, including phoneme accuracy, prosodic errors, lengthened/disrupted transitions, respiration/voice/resonance errors, specific segmental errors, specific structural errors (e.g., migrations, syllable additions), and "other" (e.g., timing errors, intelligibility index). Nonidentical, conceptually redundant features (n = 52) were resolved into 18 unique perceptual candidate features; for example, although not exact duplicates of each other, conceptually redundant features such as "more atypical lexical stress," "percentage of lexical stress matches," and "percentage of appropriate stress" were consolidated into "percent stress correct." From these unique perceptual candidate features, 15 were operationalized for analysis. The three items that were not operationalized were removed for being overly broad (e.g., timing errors) or because the items could not be queried reliably in the speech task and corpus tools used in this study (e.g., percent glides correct, distortions).[2] The second author refereed all data coding decisions.

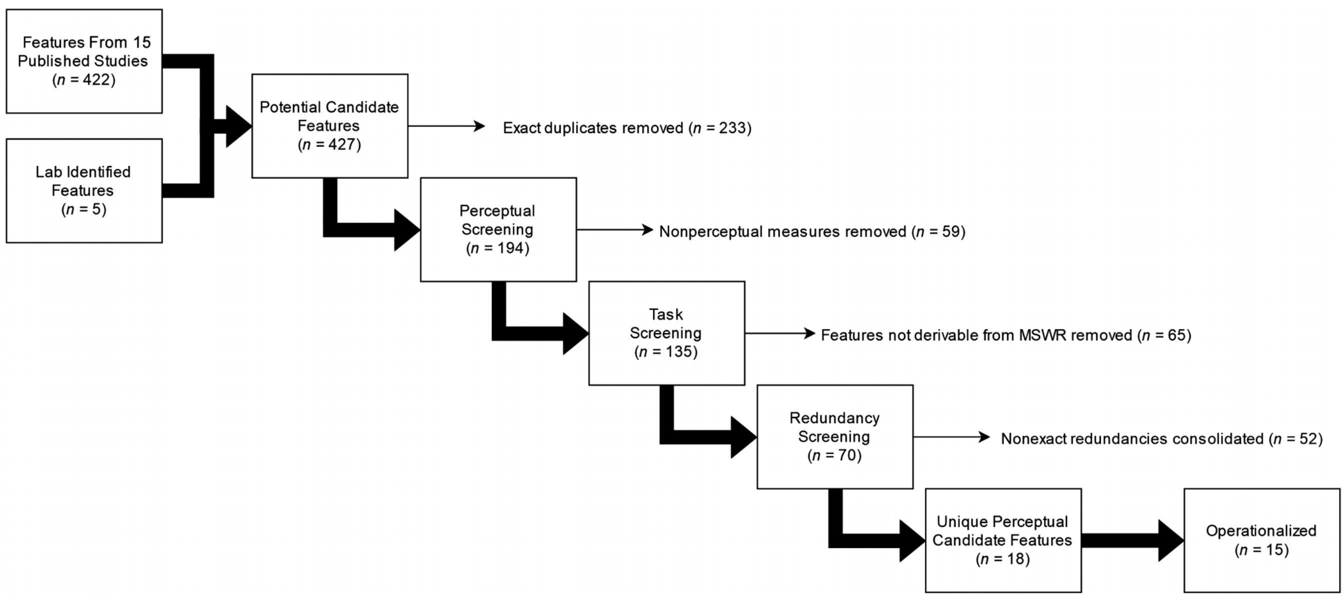
Results: Research Question 1

The 15 unique perceptual candidate features that resulted from the narrative review of the literature included lengthened vowels, syllable segregations, percent stress correct, percent consonants correct, percent vowels correct, percent phonemes correct, nasal emissions, metathesis, migrations, lenition, voicing changes, epenthesis, percentage of structurally correct words, syllable additions, and syllable deletions.[3] These features are reported in studies that conducted group comparisons of CAS versus SSD (Davis et al., 1998; Grigos et al., 2015; Munson et al., 2003; Murray et al., 2015; Shriberg et al., 1997a, 1997b, 2003, 2011; Velleman & Shriberg, 1999), CAS/CAS+ versus SSD (Murray et al., 2015), CAS or CAS with language impairment versus SSD (Iuzzini-Seigel et al., 2017, 2015), CAS versus SSD with language impairment (Lewis et al., 2004), CAS with comorbid galactosemia versus SSD with comorbid galactosemia versus SSD uncomplicated by galactosemia (Shriberg et al., 2011), and CAS

[2]As of Phon Version 3.0.6-beta.4, distortion errors derived from the WAP were not properly computed.

[3]Lenitions are defined as a "failure to reach articulatory target" (Bauer, 2008) and are perhaps a transcribable perceptual analog that complement articulatory kinematic findings in individuals with CAS, which were not included in this review.

Figure 1. Process for identifying speech features reported in the literature contrasting childhood apraxia of speech and other speech sound disorders. MSWR = multisyllabic word repetition.



marked with inappropriate stress versus CAS marked with appropriate stress (Velleman & Shriberg, 1999).[4]

Interim Discussion

Fifteen features were identified that could be operationalized and applied to phonetic transcriptions of multisyllabic words. Features were included without regard to their qualitative impact or statistical significance in each study. Disagreement between studies was therefore noted in the statistical conclusions reached with respect to some features. “Epenthesis” (as intrusive schwa) and “percent phonemes correct” were found to significantly distinguish between groups with CAS and SSD in the respective parent study (Murray et al., 2015), with no contrary evidence found in other studies during the review. However, it was found that CAS was not reliably distinguished from other types of SSD on the frequency of “syllable deletions” (Velleman & Shriberg, 1999), “lengthened vowels” and “nasal emissions” (Shriberg et al., 2011), and “voicing changes” (Murray et al., 2015). Several features had equivocal findings across multiple studies, including “percent consonants correct” (significant in Murray et al., 2015; Shriberg et al., 2011; not significant in Shriberg et al., 1997a), “percentage of structurally correct words” (Shriberg et al., 2011), vowel errors—including “percent vowels correct”—(significant in Murray et al., 2015; not significant in Shriberg et al., 1997a, or Velleman & Shriberg, 1999), “syllable segregations” (significant in Murray et al., 2015; not significant in Shriberg et al., 2011), and “percent stress correct" (significant in Murray et al., 2015; Shriberg et al., 1997b, 2003; not significant in Shriberg et al., 2011). Further complicating the above is that there are at times conflicting results for speech features within the same study, based on the number and identity of subgroups being tested (e.g., Shriberg et al., 2011). The remaining four perceptual features (i.e., syllable additions, metathesis, migrations, and lenitions) are exploratory features that were not explicitly tested in the studies included in the review, although some authors have described these or related features as relevant to CAS diagnosis (e.g., Case & Grigos, 2016; Lewis et al., 2004).

Previous attempts to differentiate the speech of individuals with CAS from the speech of individuals with other SSDs are varied in the features identified for analysis, the diagnostic characteristics of participants, and the statistical results of the features examined for differentiating between groups. Some of the features that have been studied address timing of articulatory movements that impact segments (e.g., lengthened vowels, voicing errors), some capture structural errors that relate to sound or syllable additions (e.g., epenthesis) or deletions (e.g., syllable omissions), some are broad descriptions of overall segmental accuracy (e.g., percent phonemes correct and percent consonants correct), and some capture suprasegmental aspects of speech (e.g., percent stress correct and syllable segregations).

There remains neither consensus nor replication of the features that may differentiate CAS from other types of SSD, nor consistent reporting of effect sizes from which clinical significance can begin to be interpreted. Furthermore, this review only considered studies in which speech features were contrasted between groups of individuals with

[4]Studies included in the review are marked with an asterisk in the reference list.

CAS and another SSD. There are other features in the literature seeking to distinguish CAS but were not included in this review, as there was either no comparison group or no SSD comparison group (Case & Grigos, 2016; Peter & Stoel-Gammon, 2008). Also excluded from the review were unpublished works (e.g., Murray et al., 2018; Schumacher et al., 1986). A full, systematic review of all features previously investigated around CAS was beyond the scope of the present project. Future research should consider replicating this study within the guidelines of a systematic review.

Following from this literature review, the 15 resultant features were explored to determine which, if any, might discriminate among school-age children with CAS and a comparison group of children with non-CAS SSD. We sought to determine how well the 15 candidate perceptual features, extracted from multisyllabic word repetitions, discriminate between groups of children with CAS and another subtype of SSD in children and adolescents between the ages of 7 and 17 years.

Method: Research Question 2

Data came from children who were evaluated for several treatment studies of school-age children in Syracuse, New York (Preston & Leece, 2017; Preston et al., 2016, 2017; Sjolie et al., 2016; and an ongoing study of CAS). Data reported here were collected before any experimental treatment had commenced. Data collection procedures were approved by the Syracuse University Institutional Review Board under Protocols 14–117 and 17–177.

The participants were 61 children and adolescents aged 7–17 years who received a diagnosis of either CAS (CAS group, n = 20) or SSD characterized by non-CAS motor-based articulation errors (SSD group, n = 41). The overall sample included 22 girls and 39 boys, with the SSD group containing 18 girls and 23 boys, and the CAS group containing four girls and 16 boys. All children spoke English as their primary language and passed a hearing screening. Participants either participated in articulation testing with the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000) or the Goldman-Fristoe Test of Articulation–Third Edition (Goldman & Fristoe, 2015), depending on the study protocol utilized during their evaluation. Regardless of test edition, children within the sample were required to have scores below the 7th percentile for inclusion in their parent study. Across all parent studies, participants were required to have a receptive vocabulary score (Peabody Picture Vocabulary Test–Fourth Edition; Dunn & Dunn, 2007) and a nonverbal cognition score (Wechsler Abbreviated Scale of Intelligence–Second Edition Matrix Reasoning subtest; Wechsler, 2011) not less than one and one third standard deviations below the mean. The Phonological Awareness Composite from the Comprehensive Test of Phonological Processing–Second Edition (Wagner et al., 2013) was administered, but results were not exclusionary. Descriptive information for the participants in the current sample is shown in Table 1.

Diagnoses for members of the CAS group were established by at least one of three experienced clinicians (Preston, Vose, or Leece) on the basis of a combination of factors including evidence of at least two of the following: (a) a minimum of four addition errors on the Syllable Repetition Task (Shriberg et al., 2009, 2012), (b) slow and/or errored productions of the trisyllable /pɑtɑkɑ/ on the Maximum Performance Task (Rvachew et al., 2005; Thoonen et al., 1999, 1996), (c) lexical stress errors on a multisyllabic picture-naming task, and (d) evidence of syllable segregation on a multisyllabic picture-naming task (e.g., Murray et al., 2015). Therefore, multiple measures were used to inform the diagnosis.

Multisyllabic Word Repetition Task

As a part of a larger evaluation battery, children directly imitated recordings of 20 words ranging from three to five syllables spoken by a female. Productions were not utilized to make CAS diagnostic decisions. The stimuli were the same as Preston and Edwards (2007), which were modeled after Lewis and Freebairn (1992). Productions were recorded at 44.1 kHz using Praat software (Boersma & Weenink, 2019) for later phonetic transcription from the uncompressed audio file. The stimuli, used to assess imitation accuracy, are listed in Appendix A, and descriptions of the stimulus properties are shown in Table 2.

Transcription

All productions were independently transcribed by two listeners, blind to the diagnosis of the participants, using Phon software (Hedlund & Rose, 2019). The productions were narrowly transcribed using the conventions of the International Phonetic Alphabet (International Phonetic Association, 2015) and Extensions to the International Phonetic Alphabet (International Clinical Phonetics and Linguistics Association, 2015) with both segmental and suprasegmental features noted. Primary and secondary lexical stress was transcribed. Source characteristics were not transcribed, so that alterations of source (e.g., murmured voice) were not captured in segmental accuracy counts. Only the last, complete attempt at the word was transcribed in the case of false starts or repetitions. Following transcription, the segments appearing in both target and production transcriptions were manually aligned to each other, and all segments were aligned into syllables. Both transcribers had training in articulatory phonetics as well as SSDs. The second author contributed orthographic and broad phonemic transcriptions of the stimuli, accounting for linguistic variation.

Interrater reliability for the two transcribers was calculated for general categories of perceptual features in the transcriptions: segment accuracy, segregated syllables, and lexical stress deviations. Because the two transcribers were the same for all tokens, reliability was calculated with raters as fixed effects using the average of the two ratings. Intraclass correlation coefficients indicate that the interrater

Table 1. Participant characteristics.

Characteristic	SSD group (*n* = 40) *M* (*SD*)	CAS group (*n* = 21) *M* (*SD*)
Age	11.06 (3.08)	11.92 (3.36)
PPVT-4 standard score	113.56 (14.69)	103.15 (10.73)
GFTA-2[a] standard score (*n* = 50)	72.20 (9.93)	53.50 (16.57)
GFTA-3[a] standard score (*n* = 10)	—	41.30 (4.11)
CTOPP-2 Phonological Awareness Composite score	101.44 (10.12)	85.55 (13.93)
Maximum Performance Task: Dysarthria score	0.05 (0.31)	0.26 (0.65)
Maximum Performance Task: Apraxia score	0.73 (0.923)	1.58 (0.84)
Syllable Repetition Task: Percent Consonants Correct	92% (7.24)	85% (11.97)
Syllable Repetition Task: Additions	0.75 (1.45)	3.37 (2.81)

Note. SSD = speech sound disorders; CAS = childhood apraxia of speech; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; GFTA-3 = Goldman-Fristoe Test of Articulation–Third Edition; CTOPP-2 = Comprehensive Test of Phonological Processing–Second Edition.

[a]Participants engaged in either the GFTA-2 (*n* = 50) or GFTA-3 (*n* = 10) based on the specific study protocol under which their evaluation took place. None of the 10 children who received the GFTA-3 were assigned to the SSD group; this is indicated with an em dash in the table above. One participant presented for evaluation with a recent GFTA-2 clinical score, and that data point was not recorded.

reliability for the two transcribers was .97 for segmental accuracy (percent phonemes correct; 95% CI [.95, .98]), .83 for transcription of syllable segregations (95% CI [.72, .90]), and .71 for transcription of percent stress correct (95% CI [.52, .83]).

Operationalization of Features and Phon Queries

The 15 unique perceptual features included in the study were operationalized during preparation for the analysis. Counts or percentages of each empirical feature were obtained by querying the transcriptions using Phon 3.0.6-beta-4 (Hedlund & Rose, 2019), a tool that has previously been used to analyze transcriptions for children with CAS (Barrett et al., 2019). Some queries, such as percent consonants correct, were run as included in the Phon software. Other features were queried using Phon's specialized analysis set for the Word-Level Analysis of Polysyllables (WAP), described by Masso et al. (2018). The WAP incorporates the systematic sampling of multisyllabic words representing a variety of word lengths, phonotactic syllable shapes, and lexical stress patterns; these recommendations can be applied to any word-list with a sufficient number of opportunities to examine the following errors: segmental substitutions, deletions of segments or syllables, segmental distortion, segmental addition, phonotactic alteration (changes to word length or word shape), timing alteration (lexical stress alterations or syllable segregation), and assimilation/alteration of sequence.

The remaining queries were encoded and tested using Phonex, a pattern-matching language similar to regular expression pattern matching, that is included within Phon.

Table 2. Most prevalent phonological features in multisyllable word repetition task.

Word-initial syllable type (2 types)	Syllable shapes (13 shapes)	Syllables with complex onsets (3 types)	Syllables with complex codas (1 type)	Syllable stress (3 levels)	Lexical stress pattern (11 patterns)	Consonant phonemes in task (20 consonants)	Vowel phonemes in task (12 vowels)
Consonantal onset (11)	CV (47)	CCV (3)	CVCC (4)	Unstressed (50)	U1U (4)	/s/ (20)	/ə/ (26)
Null onset (9)	CVC (11)	CCVC (1)		Secondary stress (11)	2U1UU (3)	/t/ (14)	/ɪ/ (22)
	CVCC (4)	CCCV (1)		Primary stress (20)	1UU (2)	/k/ (13)	/i/ (6)
	V (4)				2U1U (2)	/m/ (11)	/æ/ (6)
	VC (4)				U1UU (2)	/n/ (11)	/ɛ/ (5)

Note. Null onset syllables are those that begin with vowels. Numbers in parentheses represent frequency of occurrence of the exemplar listed in the cell. Where more than five different exemplars occurred for a given phonological feature, the five most frequently occurring patterns are listed in the column. C = consonant; V = vowel. For lexical stress pattern: 1 = syllable receiving primary stress; 2 = syllable receiving secondary stress; U = unstressed syllable.

Phonex allows for transcribed corpora to be queried based on phonetic features and natural classes (e.g., coronal, fricative, rhotic) as well as traditional regular expressions. The feature counts returned by Phon were averaged across the two transcribers to determine the average instance of each feature for each participant. The specific commands and expressions used to query the multisyllable word repetition (MSWR) corpus are defined in Appendix B.

Statistical Analysis

A one-tailed, independent-samples nonparametric test (Mann–Whitney *U*) was utilized to determine the probability that a member of the CAS group will have a larger incidence of speech features (or lower percent accuracy) than a member of the SSD group. Spearman's rho was also utilized to examine the correlations between the phonetic information captured by these speech features. Nonparametric statistical analyses were selected over less conservative parametric variants in order to offset the impact of small sample size, nonnormal distributions and outliers on the comparisons. In order to offset the inflated risk of Type I error when multiple comparisons are conducted, Holm's alpha correction (Holm, 1979) was applied, such that, beginning with an alpha level of .05, the threshold for statistical significance for each tested speech feature was reduced iteratively using the rank order of each calculated *p* value. That is, based on rank order, the first comparison required a *p* value of less than .05/15 = .0033 to reach significance, the next comparison required a *p* value of less than .05/14 = .0036 to reach significance, and so forth. Nonparametric effect sizes were calculated for each variable as described in Field (2005). The correlation of $\frac{|z|}{\sqrt{n}}$ was then converted to a derived Cohen's *d* using the method by Rosenthal (1994) and the formula provided by DeCoster (2012).

The pairwise comparisons were conducted within SPSS Version 24 (IBM Corp, 2016), while the correlation matrix and boxplots were generated with the Python Data Analysis Library (pandas; McKinney, 2010), SciPy Library, Seaborn Library, and Matplotlib Library (Hunter, 2007).

Results: Research Question 2

Statistically significant group differences were found between the CAS and the SSD groups for four variables: voicing changes, percentage of structurally correct words, percent stress correct, and syllable deletions. Broad categories of errors associated with CAS are represented in the significant features, including timing of articulatory movements that impact segments (e.g., voicing errors), errors that relate to phonological structure (e.g., structurally correct words, syllable deletions), and suprasegmental aspects of speech (e.g., percent stress correct). The data are summarized in Table 3 and Figure 2 (in both of which speech features are presented in order of derived effect size).

Among the features that were significantly different between the groups, derived effect sizes were in the moderate range (0.5–0.8; percent stress correct, percent full syllable deletion) to large range (> 0.9; voicing changes, percentage of structurally correct words). There were also several variables with moderate effect sizes that did not reach statistical significance when the correction for multiple comparisons was applied (including percent consonants correct, migrations, lengthened vowels, percent phonemes correct, and percent vowels correct). Although four variables were significantly different between the groups, as can be seen in Figure 2, there is a great deal of overlap in the distributions of the two groups, suggesting that no single measure would necessarily serve as a diagnostic marker in this sample.

Many of the speech features entered into the analysis were significantly correlated with each other even when setting a conservative threshold for Type I error (α = .001). Some of these correlations are expectedly high because there is overlap in the nature of the measurement; for example, percent phonemes correct is influenced by percent consonants correct and percent vowels correct, and therefore, there are high correlations among these variables. Similarly, percentage of structurally correct words is impacted by sound additions (epenthesis, syllable additions) and omissions (reflected in variables such as syllable deletions and in percent consonants correct, percent vowels correct, and percent phonemes correct; see Figure 3).

Discussion

At the broadest level, there has been general agreement about the underlying characteristics of CAS (i.e., a disorder that includes, but may not be limited to, motor planning/programming; ASHA, 2007). However, there is no theoretical or empirical consensus yet regarding the speech features that are associated with the overt realization of the impairment. This study sought to replicate and extend prior work on speech features that may be most prevalent in children with CAS, focusing on perceptual features recoverable from multisyllabic word repetitions in school-age children and adolescents.

From a review of the literature, 15 features were identified as being applicable to phonetic transcription on a multisyllabic word repetition task. Only four speech features —with medium to large effect sizes—significantly differed between the groups in the present transcribed multisyllabic word corpus, which may be due in part to the sample size and the conservative statistical adjustment. The speech features in this study that were found to differ between children with CAS from other SSDs during multisyllabic word repetition (voicing changes, percentage of structurally correct words, correct lexical stress, and syllable deletions) were partially aligned with findings from existing literature. These features capture a range of segmental, structural, and prosodic aspects of speech that are consistent with the theorized underlying impairment in planning and programming speech movements.

Some variables that have been shown to identify CAS in prior studies failed to replicate here; for example, epenthesis and percent phonemes correct (Murray et al., 2015) did not significantly differentiate the groups in this study.

Table 3. Descriptive statistics, inferential statistics, and effect sizes for group comparisons of investigated speech features.

Speech Feature	SSD group *M* (*SD*)	SSD group *Mdn* (range)	CAS group *M* (*SD*)	CAS group *Mdn* (range)	*U*	Holm's α	*p*	Derived Cohen's *d*
Voicing change	1.6 (1.5)	1 (0–8.5)	3.4 (2.9)	3 (0–13)	194	.0033	.0004*	0.950
Percent structurally correct words	91.2 (6.2)	92.6 (75.1–99.4)	81.4 (13.3)	82.4 (39–96.3)	200	.0036	.0006*	0.907
Percent stress correct	93.4 (4.3)	94.4 (81.6–99.4)	89 (6.3)	90.1 (72.1–96.9)	208	.0038	.0010*	0.865
Percent full syllable deletion	0.6 (0.8)	0 (0–3.7)	2.7 (3.7)	1.2 (12.2)	240	.0042	.0034*	0.740
Percent consonants correct	82.8 (9.3)	85.6 (60.2–94.4)	74.8 (15)	77.9 (28–93.8)	251	.0045	.0073	0.659
Migrations	0.9 (1.1)	0.5 (0–5)	2.7 (4.3)	1.3 (0–19)	260.5	.005	.0093	0.631
Lengthened vowels	0.3 (0.5)	0 (0–2.5)	0.6 (0.8)	0.5 (0–3)	278.5	.0056	.0098	0.627
Percent phonemes correct	85.6 (6.8)	86.7 (69.4–94)	79.3 (11.5)	82 (43.1–94.2)	260	.0063	.0106	0.618
Percent vowels correct	89.3 (4.3)	90.1 (79.5–97)	85.1 (8.1)	84.5 (61.7–97.1)	277.5	.0071	.0209	0.541
Epenthesis	2.7 (2.6)	2.5 (0–13)	3.6 (2.8)	3.5 (0–12.5)	313	.0083	.0674	0.389
Nasal changes	1.3 (1.5)	1 (0–6)	1.9 (1.7)	1.5 (0–5.5)	320	.01	.0801	0.366
Percent full syllable addition	1.2 (1.7)	0.61 (0–7.9)	1.6 (1.6)	1.2 (0–5.8)	327.5	.0125	.0985	0.335
Lenitions	1.6 (1.4)	1.5 (0–6)	2.9 (3.8)	1.5 (0–16)	339	.0167	.1350	0.285
Metathesis	0.7 (0.9)	0.5 (0–3.5)	0.8 (1.1)	0.3 (0–3.5)	406.5	.025	.4770	0.014
Syllable segregation	2 (2.7)	1 (0–12)	1.9 (1.9)	1.3 (0–5.5)	408.5	.05	.4907	0.006

Note. SSD = speech sound disorders; CAS = childhood apraxia of speech.
*$p < \alpha$ (statistical significance).

Conversely, features that had not been found to significantly distinguish CAS from other subtypes of SSD in previous samples—including syllable omissions (Velleman & Shriberg, 1999) and voicing errors (Murray et al., 2015)—significantly differentiated CAS from SSD in the present sample. In fact, voicing errors had the largest effect size in the current sample. There were also features that were supported in our analysis that were associated with equivocal results in previous studies, including percentage of structurally correct words (not significant for one comparison within Shriberg et al., 2011) and appropriate lexical stress (significant in Murray et al., 2015; Shriberg et al., 1997b, 2003; not significant in one comparison of Shriberg et al., 2011). Previously equivocal features that were not supported in our analysis include vowel errors/percent vowels correct (significant in Murray et al., 2015; not significant in younger children in Shriberg et al., 1997a, or Velleman & Shriberg, 1999), percent consonants correct (significant in Murray et al., 2015, and one comparison in Shriberg et al., 2011; not significant in Shriberg et al., 1997b), and syllable segregation (significant in Murray et al., 2015; not significant in Shriberg et al., 2011). Of the four exploratory features—syllable additions, migrations, metathesis, and lenitions—none were found to reach our statistical threshold for differentiating the two groups.

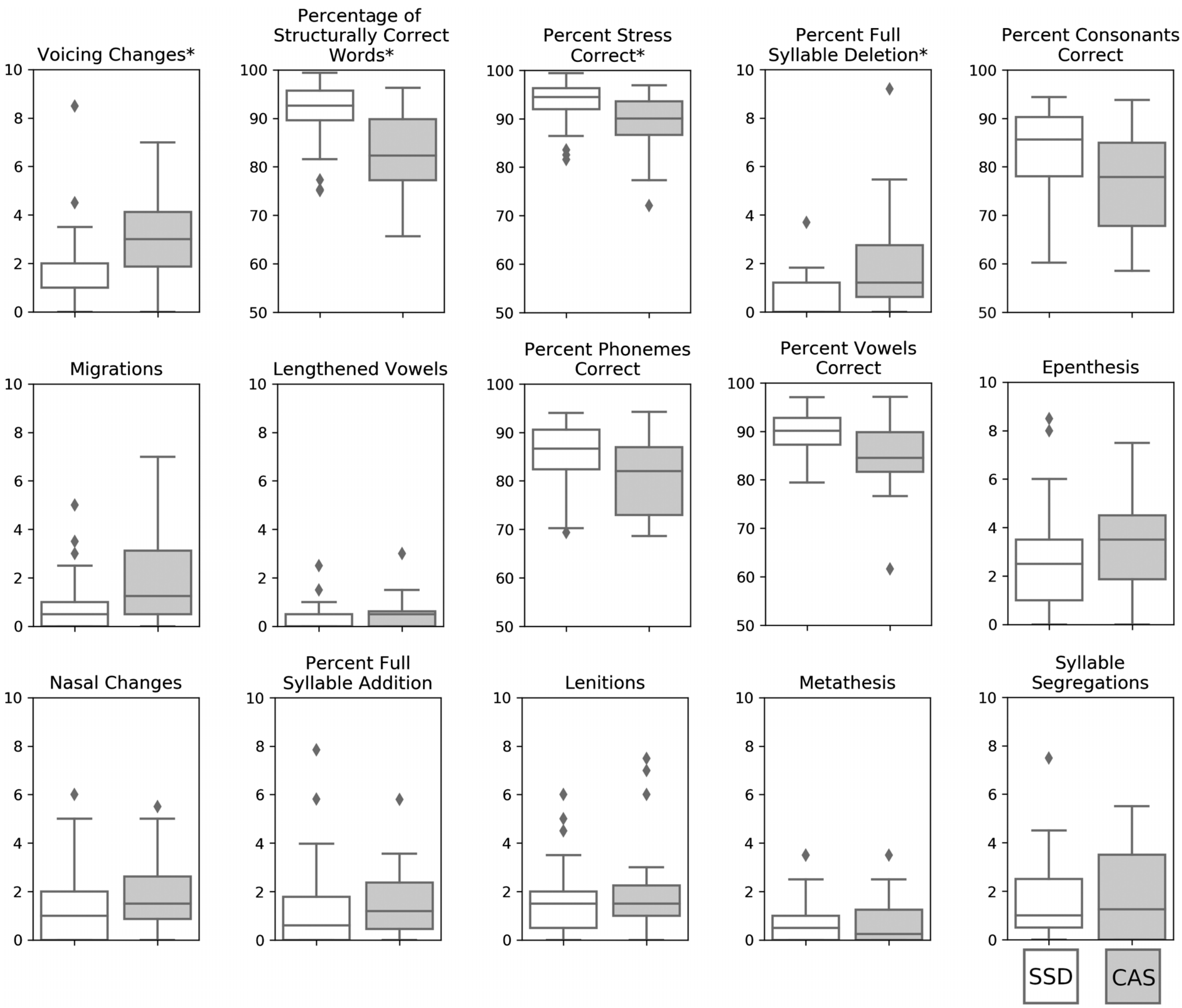
The speech features examined herein are not orthogonal. That is, the perceptual features selected for this study coexist within the same system and are, to a certain extent, colinear as a function of their operational definitions. This was seen in the demonstrated correlations between the variables. For example, instances of epenthesis are expected to correlate with the number of syllable additions, lexical stress pattern matches, and percentages of structurally correct words, which was found within this data set. The extent to which the perceptual features influence each other makes it difficult to distill the salient speech features of CAS down to rudimentary, distinct, clinically useful diagnostic markers. Although appropriate analyses of colinear features can be completed at the group level using certain statistical techniques (such as principle component analysis), such techniques are not applicable to identify the specific features observed in a specific child and are thus limited in clinical utility. This may point, however, to the utility of featural checklists to identify combinations of variables to diagnose CAS, as has been proposed elsewhere (e.g., Shriberg et al., 2011).

Clinical Implications

Relatively little is understood about presentations of CAS in older children and adolescents. The participants in this study were somewhat older than participants in previous studies—including Shriberg et al. (2011) and Murray et al. (2015)—which, as a whole, primarily address CAS in preschool and young school-age children (i.e., approximately ages 3–8 years). Because the presentation of CAS can change from the preschool to the school-age years (Lewis et al., 2004), it is possible that some overt speech features taken from the existing literature become less salient as children with CAS mature. It is also possible that these features may have already been remediated in our sample during periods of speech therapy. This punctuates two related needs: the need for clinical guidelines that address CAS as both a spectrum of difficulty and as a condition whose manifestation changes over the course of development as well as the need for clinical service delivery that addresses the unique presentation of a given child rather than those that adhere to the canonical feature set of a given diagnosis.

The current study highlights that the differential diagnosis of school-age children and adolescents with CAS

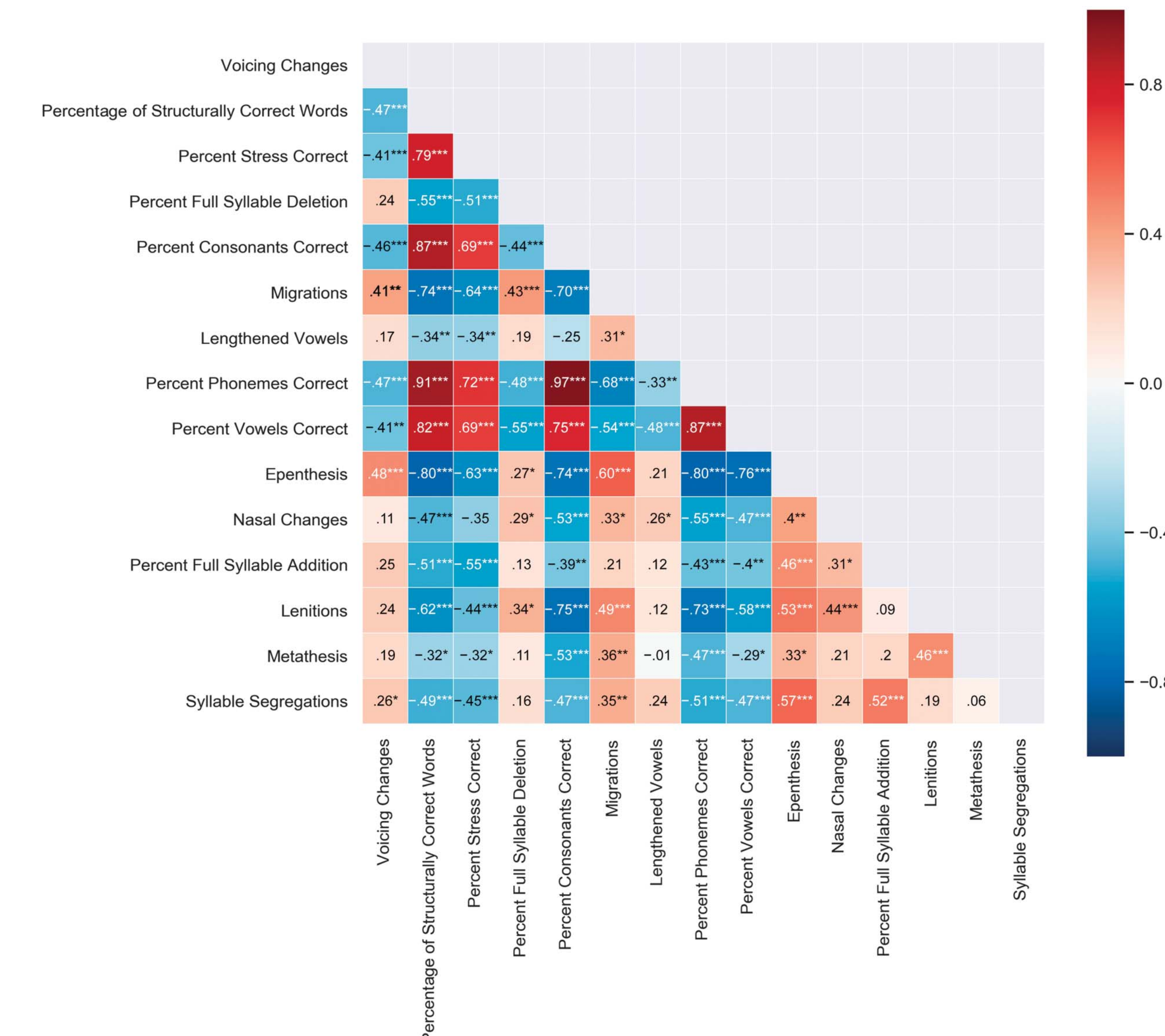
Figure 2. Comparisons of childhood apraxia of speech (CAS) and speech sound disorder (SSD) groups on 15 measures derived from multisyllabic word repetition. Asterisks following feature names indicate at the difference between SSD and CAS was statistically significant for that feature.



using multisyllabic words may be informed by analysis of lexical stress errors, word-level structural errors, and certain segmental errors, which occur at a rate different than peers with motor-based articulation disorders. These collectively point to multifaceted realization of CAS that affects prosody, oral–laryngeal timing, and faithfulness to phonological word shape.

Although this study focused on the differential diagnosis of CAS from other subtypes of SSD, it also provides descriptive information about the multisyllabic word repetition skill of school-age individuals with SSD not characterized by CAS. These children presented with counts of certain structural and segmental features that did not differ from the rate seen in children with CAS; noteworthy among these were syllable segregations, metathesis, and epenthesis. Multisyllabic words are frequently represented in Tier II and Tier II vocabulary (e.g., Beck et al., 2013) and therefore have high ecological validity for educationally relevant speech and language goals. The results herein provide justification for the use of multisyllabic productions as an important component of speech assessment for children with SSD as well as CAS, with Phon (Hedlund & Rose, 2019) and the WAP (Masso et al., 2018) as tools that may facilitate the inventory of deletions, changes in phonotactics, and changes in timing in school-age children and adolescents regardless of CAS diagnosis. Such aspects of multisyllable

Figure 3. Nonparametric correlations between the 15 variables derived from multisyllabic word repetition. *Significant at $p < .05$. **Significant at $p < .01$. ***Significant at $p < .001$.



productions may be challenging even for children with typical speech development and non-CAS motor-based articulation disorders but may be particularly difficult for children with CAS to master. This and other questions could be elucidated by future research that longitudinally compares multisyllabic word production ability between typically developing children, children with CAS, and non-CAS SSD.

Finally, it is also worth noting that perceptual features were selected for this study to emphasize their utility in clinical case planning. An intervention plan for a child with CAS, however, must also address skills in domains beyond the scope of this project known to be impacted in individuals with CAS, including auditory perceptual tasks, phonological skills, and literacy (e.g., Lewis et al., 2004; McNeill et al., 2009).

Reliability of CAS Feature Identification and Study Limitations

Limitations will arise in any study of CAS due to a variety of factors, including data encoding (whether perceptual, acoustic, or kinematic). Narrow transcription of perceptual features, the data encoding method used in this study, is a processing-intensive skill that requires a transcriber to simultaneously perceive multiple auditory signatures and match salient features to a symbol set based on factors such as phoneme correctness, phoneme identity, phoneme distortion, phoneme timing, coarticulation, and relative syllable emphasis, among other things. Narrow transcription relies on an individual's speech sound perception, and it is probable that auditory speech sound perception is a continuous skill among transcribers, with some transcribers not capturing enough narrow phonetic detail to be able to differentiate between children with CAS from other subtypes of SSD and some transcribers capturing too much phonetic detail that overrides categorical perception and serves to obscure salient features that lead to parsimonious distinctions between subtypes of SSD. The weaker transcriptional reliability for prosodic features is especially impactful regarding SSDs with a prosodic component, such as CAS. Further complicating difficulties with phonetic transcription is a lack of concrete operationalization of the speech features that do appear in the literature under the same or similar names. Finally, the validity and reliability of measurement from phonetic transcriptions is a function of the speech sample used; thus, the characteristics and number of multisyllabic word stimuli used in this study may have hindered the identification of particular CAS features.

Because CAS presents along a spectrum of difficulty (e.g., Strand et al., 2013), features that are pooled from a body of literature that focuses on individual phenotypes of CAS (e.g., CAS resulting from galactosemia, as in Shriberg et al., 2011, or CAS with appropriate stress patterns, as in Velleman & Shriberg, 1999) may fail to be replicated in samples with different characteristics. Large and diverse samples are needed to address such questions, and the relatively small number of participants with CAS in this study may limit the conclusions that can be drawn. Future research could address these questions within a larger sample size, utilizing a community sample.

Future research on features of CAS will need to consider several factors. The relative agreement (or lack thereof) on the diagnosis is hindered by a lack of an operational definition. Whereas some specific tasks have been developed to identify CAS (Shriberg et al., 2009, 2012; Strand et al., 2013; Thoonen et al., 1999, 1996), it is unclear the extent to which any of these tasks identify the "same" individuals as having CAS. Furthermore, to the extent that CAS can be associated with a number of etiological factors, the relationship between the speech features and underlying biological variables (e.g., genetic and neurobiological differences) requires further exploration. Finally, future studies seeking to differentially diagnose CAS from other subtypes of SSD should explore longitudinal changes in CAS across the life span to understand developmental realization of the disorder.

Conclusion

A review of the literature yielded 190 perceptual and acoustic speech features previously investigated for their ability to distinguish CAS from another subtype of SSD. Of the 15 perceptual items selected for this investigation, voicing changes, percentage of structurally correct words, correct lexical stress, and syllable deletions were found to significantly differentiate CAS versus SSD within this sample of school-age children and adolescents. There is no consensus yet regarding the statistical or clinical significance of many of these features; the current study indicates that only some, but not all, features identified in previous literature were replicated in this sample of school-age children. The diagnosis of CAS should likely be informed by a number of speech features impacting prosody, oral–laryngeal timing, and faithfulness to phonological word shape.

Acknowledgments

Funding was provided by grants from the National Institutes of Health (R03DC012152 and R15DC016426; J. Preston, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to acknowledge Raya Campeas for her transcriptions and her assistance with earlier stages of this work, as well as Qiu Wang for helpful comments on a previous version of the statistical analysis and Yvan Rose for his insights related to Phon.

References

Studies included in the review are marked with an asterisk in the reference list.

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

Barrett, C., McCabe, P., Masso, S., & Preston, J. (2019). Protocol for the connected speech transcription of children with speech disorders: An example from childhood apraxia of speech. *Folia Phoniatrica et Logopaedica,* 1–15. Advance online publication. https://doi.org/10.1159/000500664

Bauer, L. (2008). Lenition revisited. *Journal of Linguistics, 44*(3), 605–624. https://doi.org/10.1017/S0022226708005331

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford.

Boersma, P., & Weenink, D. (2019). Praat (Version 6.0.43) [Computer software]. University of Amsterdam.

Case, J., & Grigos, M. I. (2016). Articulatory control in childhood apraxia of speech in a novel word-learning task. *Journal of Speech, Language, and Hearing Research, 59*(6), 1253–1268. https://doi.org/10.1044/2016_JSLHR-S-14-0261

*__Davis, B. L., Jakielski, K. J., & Marquardt, T. P.__ (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics & Phonetics, 12*(1), 25–45. https://doi.org/10.3109/02699209808985211

DeCoster, J. (2012). *Spreadsheet for converting effect size measures* (Version 2012-06-19). http://www.stat-help.com/spreadsheets/Converting%20effect%20sizes%202012-06-19.xls

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research, 36*(5), 1051–1054. https://doi.org/10.1044/jshr.3605.1051

Dunn, L. M., & Dunn, D. D. (2007). *Peabody Picture Vocabulary Test–Fourth Edition* [Database record]. Pearson. https://doi.org/10.1037/t15144-000

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Sage.

Goldman, R., & Fristoe, M. (1986). *Goldman-Fristoe Test of Articulation–Second Edition*. AGS.

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

Gozzard, H., Baker, E., & McCabe, P. (2008). Requests for clarification and children's speech responses: Changing "pasghetti" to "spaghetti." *Child Language Teaching and Therapy, 24*(3), 249–263. https://doi.org/10.1177/0265659008096292
*__Grigos, M. I., Moss, A., & Lu, Y.__ (2015). Oral articulatory control in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(4), 1103–1118. https://doi.org/10.1044/2015_JSLHR-S-13-0221
Hedlund, G., & Rose, Y. (2019). Phon (Version 3.0.6-beta.4) [Computer software]. https://phon.ca
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.
Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics, 72*(5), 1218–1227. https://doi.org/10.3758/APP.72.5.1218
Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
IBM Corp. (2016). IBM SPSS statistics for Windows (Version 24.0).
International Clinical Phonetics and Linguistics Association. (2015). *Extensions to the International Phonetic Alphabet*. https://www.internationalphoneticassociation.org/sites/default/files/extIPA_2016.pdf
International Phonetic Association. (2015). *Full IPA chart*. https://www.internationalphoneticassociation.org/content/full-ipa-chart
*__Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R.__ (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research, 60*(5), 1194–1210. https://doi.org/10.1044/2016_JSLHR-S-15-0184
*__Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R.__ (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54,* 32–42. https://doi.org/10.1016/j.jcomdis.2015.01.002
James, D. G., van Doorn, J., McLeod, S., & Esterman, A. (2008). Patterns of consonant deletion in typically developing children aged 3 to 7 years. *International Journal of Speech-Language Pathology, 10*(3), 179–192. https://doi.org/10.1080/17549500701849789
Klein, H. B. (1981). Productive strategies for the pronunciation of early polysyllabic lexical items. *Journal of Speech and Hearing Research, 24*(3), 389–405. https://doi.org/10.1044/jshr.2403.389
Leitãto, S., Hogben, J., & Fletcher, J. (1997). Phonological processing skills in speech and language impaired children. *International Journal of Language & Communication Disorders, 32*(2s), 91–111. https://doi.org/10.1111/j.1460-6984.1997.tb01626.x
Lewis, B. A., & Freebairn, L. (1992). Residual effects of preschool phonology disorders in grade school, adolescence, and adulthood. *Journal of Speech and Hearing Research, 35*(4), 819–831. https://doi.org/10.1044/jshr.3504.819
*__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)
Masso, S., McLeod, S., & Baker, E. (2018). Tutorial: Assessment and analysis of polysyllables in young children. *Language, Speech, and Hearing Services in Schools, 49*(1), 42–58. https://doi.org/10.1044/2017_LSHSS-16-0047
McKinney, W. (2010). *Data structures for statistical computing in Python*. Paper presented at the Proceedings of the 9th Python in Science Conference, Austin, TX, United States.
McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100
McNeill, B. C., Gillon, G. T., & Dodd, B. (2009). 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
*__Munson, B., Bjorum, E. M., & Windsor, J.__ (2003). Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder. *Journal of Speech, Language, and Hearing Research, 46*(1), 189–202. https://doi.org/10.1044/1092-4388(2003/015)
Murray, E., Iuzzini, J., Maas, E., Terband, H., & Ballard, K. (2018). *Diagnosis of childhood apraxia of speech compared to other speech sound disorders: A systematic review*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Boston, MA, United States.
*__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
Peter, B., & Stoel-Gammon, C. (2008). Central timing deficits in subtypes of primary speech disorders. *Clinical Linguistics & Phonetics, 22*(3), 171–198. https://doi.org/10.1080/02699200701799825
Preston, J. L., & Edwards, M. L. (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech, and Hearing Services in Schools, 38*(4), 297–308. https://doi.org/10.1044/0161-1461(2007/032)
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,* 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
Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (p. 239). Sage.
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.
Schumacher, J. G., McNeil, M. R., Vetter, D. K., & Yoder, D. E. (1986). *Articulatory consistency and variability in apraxic and nonapraxic children*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Detroit, MI, United States.
*__Shriberg, L. D., Aram, D. M., & Kwiatkowski, J.__ (1997a). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research, 40*(2), 286–312. https://doi.org/10.1044/jslhr.4002.286
*__Shriberg, L. D., Aram, D. M., & Kwiatkowski, J.__ (1997b). 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., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics, 17*(7), 549–574. https://doi.org/10.1080/0269920031000138123

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., & Wilson, D. L. (2010a). Perceptual and acoustic reliability estimates for the speech disorders classification system (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 825–846. https://doi.org/10.3109/02699206.2010.503007

*Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., & Wilson, D. L. (2010b). Extensions to the speech disorders classification system (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795–824. https://doi.org/10.3109/02699206.2010.503006

*Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., 1Green, J. R., & Moore, C. A. (2009). A nonword repetition task for speakers with misarticulations: The Syllable Repetition Task (SRT). *Journal of Speech, Language, and Hearing Research, 52*(5), 1189–1212. https://doi.org/10.1044/1092-4388(2009/08-0047)

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)

Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., Mabie, H. L., McSweeny, J. L., Tilkens, C., & Wilson, D. L. (2017a). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker. *Journal of Speech, Language, and Hearing Research, 60*(4), S1096–S1117. https://doi.org/10.1044/2016_JSLHR-S-15-0296

*Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., Mabie, H. L., McSweeny, J. L., Tilkens, C., & Wilson, D. L. (2017b). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: II. Validity studies of the pause marker. *Journal of Speech, Language, and Hearing Research, 60*(4), S1118–S1134. https://doi.org/10.1044/2016_JSLHR-S-15-0297

Shriberg, L. D., Strand, E. A., Fourakis, M., Jakielski, K. J., Hall, S. D., Karlsson, H. B., Mabie, H. L., McSweeny, J. L., Tilkens, C., & Wilson, D. L. (2017c). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: IV. The pause marker index. *Journal of Speech, Language, and Hearing Research, 60*(4), S1153–S1169. https://doi.org/10.1044/2016_JSLHR-S-16-0149

Sjolie, G. M., Leece, M. C., & Preston, J. L. (2016). Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders, 64,* 62–77. https://doi.org/10.1016/j.jcomdis.2016.10.003

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders, 39*(5), 331–349. https://doi.org/10.1016/j.jcomdis.2006.06.017

Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research, 56*(2), 505–520. https://doi.org/10.1044/1092-4388(2012/12-0094)

Terband, H., Namasivayam, A., Maas, E., van Brenk, F., Mailend, M. L., Diepeveen, S., van Lieshout, P., & Maassen, B. (2019). Assessment of childhood apraxia of speech: A review/tutorial of objective measurement techniques. *Journal of Speech, Language, and Hearing Research, 62*(8S), 2999–3032. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-19-0214

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

Thoonen, G., Maassen, B., Wit, J., Gabreëls, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10*(4), 311–336. https://doi.org/10.3109/02699209608985178

*Velleman, S. L., & Shriberg, L. D. (1999). Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research, 42*(6), 1444–1460. https://doi.org/10.1044/jslhr.4206.1444

Wagner, R., Torgesen, J., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition* [Database record]. APA PsycTests. https://doi.org/10.1037/t52630-000

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Iintelligence–Second Edition* [Database record]. APA PsycTests. https://doi.org/10.1037/t15171-000

Appendix A

Multisyllabic Word Repetition Task Stimuli

Abolitionist	Astronomical	Escargot	Practitioner
Abominable	Cavalry	Excavator	Sclerosis
Accessibility	Communicative	Linoleum	Specificity
Aluminum	Denominator	Logistics	Statistics
Amethyst	Emphysema	Mahogany	Synergistic

Note. Participants imitated these 20 words.

Appendix B (p. 1 of 2)

Definitions and Phonex Queries for Speech Features

	Speech feature	Operational definition	Phon command or Phonex expression
Broad segmental accuracy	Percent consonants correct	Percentage of consonant productions matching the target	Phon > Analysis > PPC > PCC ignore diacritics = off
	Percent vowels correct	Percentage of vowel productions matching the target	Phon > Analysis > PPC > PVC ignore diacritics = off
	Percent phonemes correct	Percentage of phoneme productions matching the target	Phon > Analysis > PPC > PPC ignore diacritics = off
Errors of manner and timing	Lengthened vowels	Count of lengthened vowels	Phon > Query > Phones IPA Target: \v IPA Actual: \v:
	Nasal changes	Sum of count nasal changes and of count nasal emission	Phon > Query > Phones IPA Target: {oral} IPA Actual: {nasal} Phon > Query > Phones IPA Target: {nasal} IPA Actual: {oral} Phon > Query > Phones IPA Target: {vowel} IPA Target: {nasal} Phon > Query > Data Tiers IPA Actual {narealfricative}
	Voicing changes	Sum of count [+VOICED] ➔ [–VOICED] and of count [–VOICED] ➔ [+VOICED]; includes prevoicing and partial voicing/devoicing	Phon > Analysis > Phonological Processes > Voicing IPA Target: {voiceless} IPA Actual: {voiced} ignore diacritics=off Phon > Analysis > Phonological Processes > Voicing IPA Target: {voiceless} IPA Actual: {voiced} ignore diacritics=off

Appendix B (p. 2 of 2)

Definitions and Phonex Queries for Speech Features

	Speech feature	Operational definition	Phon command or Phonex expression
	Lenitions	Count of lenitions and spirantized stops	Phon > Query > Phones IPA Target: {stop} IPA Actual: {affricate} Phon > Query > Phones IPA Target: {stop} IPA Actual: {fricative} Phon > Query > Phones IPA Target: {stop} IPA Actual: {approximant} Phon > Query > Phones IPA Target: {affricate} IPA Actual: {fricative} Phon > Query > Phones IPA Target: {affricate} IPA Actual: {approximant} Phon > Query > Phones IPA Target: {fricative} IPA Actual: {approximant} Phon > Query > Phones IPA Actual: {spirant} ignore diacritics = off
Structural errors	Epenthesis	Count of epenthesis	Phon > Analysis > Specialized > WAP
	Metathesis	Count of phones metathesized	Phon > Analysis > Specialized > WAP
	Migrations	Count of phones migrated	Phon > Analysis > Specialized > WAP
	Percent full syllable addition	Percentage of full syllable addition	Phon > Analysis > Specialized > WAP
	Percent full syllable deletion	Percentage of full syllable deletion	Phon > Analysis > Specialized > WAP
	Percentage of structurally correct words	Full syllable addition + Full syllable deletion + Partial syllable deletion (onset, nucleus, or coda) + Syllable expansion (onset or coda)	Phon > Analysis > Specialized > WAP
Suprasegmental errors	Percent stress correct	Percent primary stress, secondary stress, and unstressed syllables correct	Phon > Analysis > Specialized > WAP
	Syllable segregation	Count of intraword pauses [^]	Phon > Query > Data Tiers IPA Actual: \^

Note. PPC = percent consonants correct; PVC = percent vowels correct; IPA = International Phonetic Alphabet; WAP = World-Level Analysis of Polysyllables.