5-2017

Speech Inconsistency in Children With Childhood Apraxia of Speech, Language Impairment, and Speech Delay: Depends on the Stimuli

Jenya Iuzzini-Seigel
Marquette University, jenya.iuzzini-seigel@marquette.edu

Tiffany P. Hogan
MGH Institute of Health Professions

Jordan R. Green
MGH Institute of Health Professions

Journal of Speech, Language, and Hearing Research, Vol. 60, No. 5 (May 2017) : 1194-1210. DOI.

Speech Inconsistency in children with Childhood Apraxia of Speech, Language Impairment, and Speech Delay: Depends on the Stimuli

Jenya Iuzzini-Seigel
Marquette University, Milwaukee, WI
Tiffany P. Hogan
MGH Institute of Health Professions, Boston, MA
Jordan R. Green
MGH Institute of Health Professions, Boston, MA

Abstract

Purpose.

The current research sought to determine 1-if speech inconsistency is a core feature of CAS or if it is driven by comorbid language impairment that affects a large subset of children with CAS, and 2-if speech inconsistency is a sensitive and specific diagnostic marker that can differentiate between CAS and speech delay.

Method.

Participants included 48 children ranging between 4;7 to 17;8 (years; months) with CAS ($n = 10$), CAS+language impairment ($n = 10$), speech delay ($n = 10$), language impairment ($n = 9$) and typical development ($n = 9$). Speech inconsistency was assessed at phonemic and token-to-token levels using a variety of stimuli.

Results.

Children with CAS and CAS+language impairment performed equivalently on all inconsistency assessments. Children with language impairment evidenced high levels of speech inconsistency on the phrase "buy Bobby a puppy". Token-to-token inconsistency of monosyllabic words and the phrase "buy Bobby a puppy" was sensitive and specific in differentiating children with CAS and speech delay whereas inconsistency calculated on other stimuli (e.g., multisyllabic words) was less efficacious in differentiating between these disorders.

Conclusions.

Speech inconsistency is a core feature of CAS and is efficacious in differentiating between children with CAS and speech delay; however, sensitivity and specificity are stimuli-dependent.

Key words:

childhood apraxia of speech, inconsistency, diagnosis, diagnostic features, language impairment, speech delay

Childhood apraxia of speech (CAS) is a speech disorder that is arguably caused by difficulty programming the motor commands that activate speech musculature (Shriberg, Lohmeier, Strand, & Jakielski, 2012). The American Speech-Language-Hearing Association (ASHA; 2007) defines CAS as "a neurological childhood (pediatric) speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits" (ASHA, 2007, pp. 3-4). CAS may be idiopathic, acquired, or have a neurogenetic or metabolic basis such as galactosemia (e.g., Shriberg, Potter, & Strand, 2011) and is estimated to affect 1-2 children per thousand (Shriberg, Aram, & Kwiatkowski, 1997). Although CAS is most commonly considered a motor speech disorder, others have argued that it could be a more general linguistic-based disorder (Hodge, 1994) that impacts the temporal ordering of language (Rosenthal, 1994), that is, the sequencing of phonological units. This theory is supported by the high rate of co-occurring language deficits that are evident among children with CAS (Aram & Glasson, 1979; Iuzzini, 2012; Lewis, Freebairn, Hansen, Iyengar, & Taylor, 2004; Thoonen, Maassen, Gabreels, Schreuder, & de Swart, 1997).

Although a validated list of pathognomonic features does not exist (ASHA, 2007; Forrest, 2003), but symptoms can include inconsistent speech sound errors (ASHA, 2007; Forrest, 2003; Iuzzini & Forrest, 2010; Marquardt, Jacks, & Davis, 2004), decreased vowel contrasts (Iuzzini & Forrest, 2008; Nijland et al., 2002) and voicing contrasts (Iuzzini, 2012; Lewis et al., 2004), prosodic disturbances (ASHA, 2007; Davis, Jakielski, & Marquardt, 1998; Shriberg et al., 1997; Shriberg, Green, Campbell, McSweeny, & Scheer, 2003), and difficulty making coarticulatory transitions between sounds and syllables (ASHA, 2007; Maassen, Nijland, & van der Meulen, 2001; Nijland et al., 2003). Symptoms are known to change over time and may not fully remediate following treatment (e.g., Lewis et al., 2004). The disorder is difficult to differentially diagnose from other speech

disorders such as speech delay because the CAS speech symptoms vary significantly across individuals and throughout development.

Among the symptoms listed above, speech inconsistency (i.e., variable production of phonemes, words or utterances across multiple opportunities) is the feature most frequently reported by clinicians and researchers (Forrest, 2003; Malemenholt, Lohmander, & McAllister, 2016; Millspaugh & Weiss, 2006; c.f. Murray, McCabe, Heard, & Ballard, 2015) as being necessary to contribute to the differential diagnosis of CAS and other speech sound disorders, although speech inconsistency is not considered sufficient to warrant a CAS diagnosis on its own (ASHA, 2007). Although the development of motor skills is marked by variable performance (Bernstein, 1967; Cohen & Sternad, 2009; Green, Moore, & Reilly, 2002; Grigos, 2009; Smith & Zelaznik, 2004), the *persistently* variable (a.k.a. inconsistent) production of speech sounds observed in children with CAS (Lewis et al., 2004) is maladaptive and associated with poor learning of speech sound targets in treatment (Forrest, Dinnsen, & Elbert, 1997; Forrest & Elbert, 2001; Forrest, Elbert, & Dinnsen, 2000; Gierut, Elbert, & Dinnsen, 1987; Yao-Tresguerres, Iuzzini, & Forrest, 2009). The accurate detection of inconsistent speech patterns is, therefore, essential for identifying children at risk for CAS (ASHA, 2007), although the psychometrics of variability in speech are poorly understood and guidelines for best assessment practices are lacking. Moreover, the extant reports on speech inconsistency as a defining feature of CAS are mixed where some studies show higher speech inconsistency among children with CAS relative to those with speech delay (e.g., Iuzzini, 2012; Marquardt et al., 2004; Schumacher, McNeil, Vetter, & Yoder, 1986) and others show equivalent levels of speech inconsistency among these groups (e.g., Betz & Stoel-Gammon, 2005).

Possible explanations for equivocal findings include differences across studies in methods used to index speech inconsistency or in sample populations. Prior work on CAS has primarily measured speech inconsistency at two levels: token-to-token and phonemic. Token-to-token inconsistency indexes variations in speech that occur when a talker is asked to repeat the same syllable, word, or utterance multiple times (e.g., Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002). Items are typically scored as the “same” or “different” across the trials. Consequently, the trials may differ by one phoneme or by multiple phonemes, but both will receive a rating of “different.” For instance, for the target “elephant” one child could produce /elofant/ and /edofant/ while another child produces /epedam/ and /edofan/. For this word, both of these children would receive a score of “different” but the second child produced many more segments inconsistently compared to the first child. This level of analysis may too coarse for differentiating normal and disordered talkers in certain populations such as preschool-aged children who are known to exhibit high levels of variability during typical development (Iuzzini, 2012). In contrast, phonemic inconsistency reflects variability of phonemes within and across words and word positions, such that a child produces a variety of substitutions and/or distortions for a phonemic target (Forrest, Dinnsen, & Elbert, 1997; Iuzzini & Forrest, 2010; Tyler, Lewis, & Welch, 2003). For instance, a child may produce a target /t/ as [d, z, s, p]. Previous research shows that, in comparison to token-to-token estimates, this level of measurement captures inconsistency at a finer grain level, and can be more efficacious at differentiating preschool-aged children with typical and disordered speech (Iuzzini, 2012).

The extant research that compares token-to-token and phonemic speech inconsistency in the same sample of children is limited (e.g., Iuzzini, 2012; Macrae, Tyler, & Lewis, 2014). Macrae and colleagues (Macrae et al., 2014) compared “word variability” (i.e., token-to-token inconsistency) and “speech error variability” (i.e., phonemic inconsistency) in preschool-aged children with speech sound disorders. They found a weak correlation between these measures and that high “speech error variability” was associated with low language performance and nonword repetition scores, suggesting a link between phonemic inconsistency and language ability. Iuzzini (2012) found that although phonemic and token-to-token inconsistency tended to be highly correlated in children with typically developing and disordered speech, 18% of participants who evidenced high token-to-token inconsistency had low phonemic inconsistency, again demonstrating that different measures can yield

divergent results. Results also showed that the majority of participants (over 80%) with high speech inconsistency also had morphological language deficits (e.g., difficulty with pronouns, derivation of nouns, and the contractible copula) on the Word Structure subtest of the Clinical Evaluation of Language Fundamental-Preschool 2nd Edition (Wiig, Secord, & Semel, 2004). Although high speech inconsistency has often been attributed to motor speech deficits, the aforementioned results raise the question of whether speech inconsistency could instead be driven by a higher-order deficit that results in difficulty sequencing phonological morphological rules. Taken together, these studies suggest that different inconsistency measures yielddiffering results, and that high speech inconsistency and poor language often co-occur.

Comorbid language disorder is frequently reported among children with CAS (Aram & Glasson, 1979; Ekelman & Aram, 1984; Iuzzini, 2012; Lewis et al., 2004; Stackhouse, 1992; Thoonen et al., 1997;Velleman & Strand, 1994); however, it is not agreed upon whether language impairment should be considered part of the CAS symptom profile. Some suggest that the language deficit resolves once speech errors are remediated (Hayden & Square, 1994). For example, a child who does not produce /s/ and /z/ would not produce plurals or possessives accurately, and would therefore show a shift in language ability when /s/ and /z/ are acquired. In contrast, others have demonstrated that language errors in children with CAS are not restricted to those that hinge on accurate production of individual phonemes (e.g., Ekelman & Aram, 1984). Ekelman and Aram (1984) report that children with CAS evidence a breadth of language errors such as incorrect choice of pronouns and verbs that cannot be attributed to speech sound errors.

The interaction between the linguistic and motor systems has been demonstrated repeatedly (Goffman, 1999, 2004, 2010; Iuzzini-Seigel, Hogan, Rong, & Green, 2015a; Walsh, Smith, & Weber-Fox, 2006; Zelaznik & Goffman, 2010). Typically developing children simultaneously acquire organized language structure and coordinative mastery of the articulators (e.g., Nip, Green, & Marx, 2011). Because the neural substrates of both systems are highly interconnected (Arbib, 2006; Jancke, Siegenthaler, Preis, & Steinmetz, 2007; Kent, 2004; Ojemann, 1984), it is not surprising that kinematic analysis of children with language deficits reveal evidence of decreased oromotor coordination (Alcock et al., 2000; Goffman, 1999; Goffman, 2004). In addition, behavioral assessment of motor skills reveal poorer fine and gross motor performance among children with language disorder relative to typically developing peers (Powell & Bishop, 1992; Zelaznik & Goffman, 2010). Ojemann (1984) posits that sequential movement and language share a common substrate in the lateral perisylvian cortex of the dominant hemisphere; consequently, a disturbance or underdevelopment of this region could yield cooccurring speech and language deficits. Likewise, the cerebellum has been implicated as a key region serving movement and cognitive and linguistic performance (Leiner, Leiner, & Dow, 1991), where a cerebellar deficit can cause generalized impairments beyond the motor deficits we would typically attribute to the cerebellum (Bracke-Tolkmitt et al., 1989).

The link between the movement and linguistic systems is further supported by articulatory kinematic studies of children with specific language impairment. Specific language impairment is a developmental language disorder characterized by impaired syntax, discourse, and semantics (Paul, 2007) in the absence of cognitive deficits (Catts, Adlof, Hogan, & Weismer, 2005; Leonard, Miller, & Gerber, 1999). Children with specific language impairment evidence decreased articulatory stability and oromotor coordination, and often show impaired fine and gross motor ability (Cermak, Ward, & Ward, 1986; Hill, Bishop, & Nimmo-Smith, 1998; Johnston, Stark, Mellits, & Tallal, 1981; Rintala, Pienimaki, Ahonen, & Cantell, 1998). The co-occurrence of language and movement deficits in children with specific language impairment does not suggest that language deficits underlie movement deficits. Rather, the high rate of comorbidity may suggest that there is a third, higher-order mechanism at play such as difficulty integrating sensory information (e.g., Tallal, Miller, & Fitch, 1993), poor procedural learning ability (e.g., Nicolson & Fawcett, 2007) or reduced information-processing capacity (e.g., Miller, Kail, Leonard, & Tomblin, 2001) that impacts language and motor abilities. This relation, grounded in the

extant literature on specific language impairment, provides motivation for the current study that investigates speech inconsistency in children with CAS (with and without comorbid language impairment), language impairment, and speech delay. If speech inconsistency is due to a lower-order motor programming/planning deficit, then we expect to see high speech inconsistency in all participants with CAS, independent of language ability. If instead, speech inconsistency is driven by higher-order cognitive-linguistic deficits that impair sound sequencing, we expect to see less speech inconsistency in children with CAS only relative to children with CAS and comorbid language impairment or children with language impairment only.

Assessing speech inconsistency has been a formidable research and clinical challenge because the appropriate level of analysis (e.g., token-to-token vs phonemic) has not been identified and also because speech performance is dependent on the elicitation task (e.g., Macrae et al., 2014; Vuolo & Goffman, 2014) and other stimuli characteristics (e.g., short words versus longer words or phrases, real words versus nonwords); unfortunately, there is a paucity of empirical research comparing these factors in the same children. Speech inconsistency has been reported for a wide range of speech stimuli including mono and multisyllabic exemplars (e.g., Holm, Crosbie, & Dodd, 2007) and even words elicited in continuous speech (Macrae et al., 2014; Shriberg et al., 1997). Because children with CAS are often reported to produce more errors as the number of syllables increases (Davis et al., 1998), token-to-token inconsistency of multisyllabic words or phrases may be informative and useful in contributing to the differential diagnosis of children with CAS and speech delay. There are, however, equivocal reports of whether inconsistency measures at any level are efficacious in differentiating school-age children with CAS and speech delay (e.g., ASHA, 2007; Marquardt et al., 2004; Murray et al., 2015). Consequently, it is unknown if certain stimuli will be superior in eliciting speech inconsistency in talkers with CAS.

It is essential that different inconsistency metrics and stimuli are evaluated on the same group of children with CAS so that sensitivity estimates (i.e., true positive rate) can be calculated; this will provide information about the ability of different metrics to accurately identify CAS in children who actually have this disorder. Previous research (Plante & Vance, 1994) suggests that a minimum of 80% sensitivity is recommended to ensure that children with an unknown disorder status are properly identified and can therefore receive appropriate care. Likewise, we must also test these inconsistency metrics and stimuli on control groups of children with speech delay and typical development to calculate specificity estimates (i.e., true negative rate), which will tell us the proportion of children without CAS who test negative for it based on the different inconsistency metrics. Specificity of at least 80% (Plante & Vance, 1995) is suggested to ensure that children with typical development or a different disorder type are not misdiagnosed or referred for the wrong type of treatment.

Purpose of the Current Research

The aims of this study are to (1) determine if speech inconsistency is a feature of CAS, or if it is associated with the linguistic deficits that co-occur in a large subset of children in this population and (2) determine if speech inconsistency is a sensitive and specific diagnostic marker for differentiating between children with CAS and those with speech delay. To test our first aim, we investigated speech inconsistency in the following five participant groups: (1) children with CAS who have speech symptoms only (CAS), (2) children with CAS who have a comorbid language impairment (CAS+LI), and (3) children with language impairment only (LI), (4) children with speech delay, and (5) children with normal speech and language. If speech inconsistency is a core feature of CAS, we expect that subgroups of children with CAS (i.e., CAS and CAS+LI) will be equivalent on speech inconsistency measures; alternatively, if speech inconsistency is driven by higher-order cognitive-linguistic deficits that impact phonological sequencing and morphological rules, we expect children with CAS+LI and LI to be more inconsistent than those with CAS who have speech symptoms only. We also posit that speech inconsistency will differentiate

children with CAS and speech delay, but that some types of stimuli (e.g., longer stimuli items) will be more sensitive and specific than others (e.g., shorter stimuli items).

Methods

Group Assignment

Forty-eight children ranging in age between 4;7 to 17;8 (years; months) participated as part of a larger investigation on the biological pathways that underlie childhood speech and language disorders. All participants passed a bilateral pure-tone hearing screening (ASHA, 1997) and completed a series of assessments that included the Goldman Fristoe Test of Articulation-2nd Edition (GFTA-2; Goldman & Fristoe, 2000), the Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF; Semel, Wiig, & Secord, 2003) or Clinical Evaluation of Language Fundamentals-Preschool 2nd Edition (CELF-P2; Wiig, Secord, & Semel, 2004), and the Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003). To prevent bias against participants with speech errors during language testing, children received credit for responses in which they marked an appropriate grammatical inflection, even if they did not pronounce it correctly. For instance, if a child used a vocalic syllable to mark the present participle "ing" or if a child used a substitution to mark a plural or possessive /s/, the child was given credit for these responses.

Participants were assigned to the CAS group (n = 20) based on confirmation by two blindedraters with prior experience assessing children with CAS using the criteria below. Of the 20 participants in the CAS group, half were referred with a history of CAS diagnosis and treatment by an expert clinician. Participants in the CAS group were required to have a GFTA-2 percentile < 16 and evidence ≥ 5/11 CAS features using a well-established protocol in our laboratories (Centanni, Sanmann, Green, Iuzzini-Seigel, Bartlett, Sanger, & Hogan, 2015; Iuzzini-Seigel, Hogan, Guarino, & Green, 2015b; Zuk, Iuzzini-Seigel, Green, Cabbage, & Hogan, in revision). Features were assessed during production of the GFTA-2. For operational definitions used to assess these features see Supplemental Table 1. Of these 20 participants, 10 exhibited CAS without comorbid language impairment, and 10 exhibited CAS with comorbid language impairment (CAS+LI) based on a CELF or CELF-P core language standard score of 85 or below. See Supplemental Table 2 for features that were observed in each participant and Supplemental Table 3 for individual speech, language, and cognition scores for each participant.

Inclusion in the speech delayed group ($n = 10$) was based on a GFTA-2 percentile < 16, < 5/11 CAS features, and no previous CAS diagnosis or history of treatment for CAS; the last criterion was specified to prevent inclusion of children in the speech delay group who had resolved CAS symptoms. All children in the speech delayed group had normal language per a CELF or CELF-P core language score of 86 or higher.

Participants were assigned to the language impaired group ($n = 9$) based on normal performance on the articulation test (GFTA-2 percentile ≥ 16), fewer than 5 CAS features, and evidence of language impairment based on a CELF or CELF-P core language standard score of 85 or below.

Participants were assigned to the TD group ($n = 9$) based on a GFTA-2 percentile ≥ 16, normal language, and no history of speech or language treatment. The last criterion was to exclude children who had remediated speech or language deficits.

All participants had normal nonverbal IQs (RIAS standard scores ≥ 75) and no evidence of oral weakness or report of concomitant syndromes. Although the cutoff for CAS diagnosis was presence of five or more CAS features on the GFTA-2, children in the CAS group evidenced an average of 8 features ($SD = 2$) and those in the CAS+LI group displayed an average of 7 features ($SD = 1$). In contrast, participants with language impairment averaged 1 feature ($SD = 1$), speech delay averaged 3 features ($SD = 1$), and TD averaged 1 feature ($SD = 1$). All groups were matched on age and nonverbal IQ.

Procedures

The GFTA-2, a picture naming task, was administered following standard practice. Participants also completed a customized speech battery that contained 46 real and nonwords, which sampled a range of word-lengths (2-4 syllables) as well as early, mid and late acquired phonemes (Shriberg, 1993). The extant literature reports that children with CAS have increased errors with increased number of syllables (e.g., Davis et al., 1998; Forrest, 2003), providing motivation for including stimuli with a range of wordlengths in this study. The real words for this subtest were selected from Shriberg, Jakielski, and Strand's (2010) Multisyllabic Words Tasks 1 and 2 and the Challenging Words Task. Nonword stimuli were included because children with language impairment (Archibald & Gathercole, 2007) and those with CAS (e.g., Bridgeman & Snowling, 1988) tend to have difficulty with nonword repetition. Consequently, we posited that inconsistency of nonword stimuli might differentiate between subgroups of children with CAS (CAS vs. CAS+LI) as well as between children with CAS (CAS and CAS+LI) and those with speech delay. We anticipated that children with language impairment would be equivalent to those with CAS+LI on this task. Phonotactic probability of all real and nonwords was calculated using the Child Mental Lexicon (Storkel & Hoover, 2010), an online database that determines the likelihood of occurrence for a phoneme or phoneme sequence in a particular language. Nonword and real word lists were equivalent on phonotactic probability for segmental, $t(16) = -.230$, $p = .821$, and biphone frequencies, $t(16) = 2.020$, $p = .06$ and were matched on stress and number of syllables. Although real and nonwords were not matched at the phoneme level, nonwords were designed to be phonetically balanced to real words as well as possible (e.g., /sainutæm/ was matched to 'vitamin'); consequently, the number of phonemes from each manner (e.g., fricatives, nasals) are generally equivalent across lists.

The customized speech battery stimuli were audio-recorded by a male talker and presented via sound-field as a real word and nonword repetition task. See Supplemental Tables 4 and 5 for the specific items included on this speech battery. Participant responses were recorded on digital video with the audio channel recorded at 44.1k, 16 bit for the purposes of transcription, feature-assessment, and reliability. Responses on all speech assessments were broadly-transcribed with distortions noted, consistent with the transcription procedure specified in the GFTA-2 (Goldman & Fristoe, 2000); consequently, distortions were considered errors on the GFTA-2 and were considered an error variant on the speech inconsistency assessments (further detail on calculating speech inconsistency is provided below). Speech inconsistency was intentionally not used as a criterion to determine speech disorder group assignment (CAS vs. speech delay). ASHA (2007) states that the features (e.g., speech inconsistency) described in their technical report on CAS are not intended to be considered "necessary and sufficient signs of CAS" (ASHA, 2007, p.5); therefore, we based group assignment on 11 other features that are associated with CAS in order to avoid a circularity confound with our experimental variable of interest.

Speech Inconsistency Assessments

Token-to-token inconsistency:

Word level. A list of monosyllabic ($n = 4$) and multisyllabic ($n = 8$) was repeated twice and used as the basis of the token-to-token inconsistency assessment. Monosyllabic words were produced in a picture naming task. Multisyllabic words were produced in direct imitation of a recorded model presented via sound-field. Items were presented in random order as part of the larger speech stimuli corpus. Each item received a binary score with "0" indicating that the responses were the same across both trials (e.g., "parrot" and "parrot), and "1" indicating that responses were "different" across trials (e.g., "parrot" and "pawit"). To maintain orthogonality between speech inconsistency and prosodic features, transcribers were instructed to ignore differences in stress and prosody across productions. For a trial to be considered inconsistent, there needed to be at least one phoneme level (consonant or vowel) variant (substitution or distortion). Token-to-token inconsistency was

calculated based on a procedure adapted from the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2002) using the following calculation:

Word Inconsistency Score =
(number of words produced differently across two trials/
the total number of words produced across two trials)
X 100

For example, a child produced a list of 10 words twice. For 8/10 words, the child produced the words the same across both trials. For 2/10 words, the child produced different variants across the two trials (e.g., for "snake" the child said /dek/ and /tek/ and for "cup" the child said /t^p/ and /t^d/). This child's Word Inconsistency Score would be 20%, calculated as (2/10) × 100.

The DEAP specifies that items that are inconsistent due to alternations between developmental errors and accurate productions should be removed and the test should be rescored. The DEAP scoring manual specifies numerous phonological error patterns that affect children who are 5;5 or younger. Because the current study included only two children who were 5;5 or younger, we did not assess for developmental phonological error patterns. See Supplemental Table 4 for stimuli items that contributed to assessment of token-to-token inconsistency.

Token-to-token inconsistency: Sentence level.

The sentence "buy Bobby a puppy" was produced five times in response to a verbal model that was provided at the beginning of the task. This short phrase contains simple words, with early-occurring bilabial voiced and voiceless consonants and various vowels, and is commonly used as the basis for calculation of the spatiotemporal index (e.g., Smith & Goffman, 1998), a measurement of articulatory kinematic variability that was collected but will not be reported for this study. Because children with CAS often evidence vowel distortions and difficulty producing the voicing contrast, we posit that this phrase may be useful in eliciting distinct responses between children with CAS and speech delay. Token-to-token inconsistency was calculated across the transcribed responses from the first two trials, the first three trials, and across all five trials. We opted to compare token-to-token inconsistency across two, three, and five trials to determine which number of trials yields the highest levels of sensitivity and specificity in differentiating children with CAS and speech delay.

Sentence Inconsistency Score =
number of different ways sentence was produced /
the total number of trials x 100

For example, a child produced the phrase "buy Bobby a puppy" correctly on the 1st, 4th and 5th trials. On the second trial he said "buy Boppy a puppy" and on the third trial he said "buy Bobby a pobby". Sentence Inconsistency Scores would be calculated across the first 2 trials, the first 3 trials, and then across all 5 trials. Consequently, his Sentence Inconsistency Score for the first 2 trials would be 50% calculated as (1/2) × 100 and his Sentence Inconsistency Score across all 5 trials would be 40% calculated as (2/5) × 100.

Phonemic inconsistency.

Phonemic inconsistency was calculated using an adaptation of a procedure previously used to characterize preschoolers with speech sound disorders (Iuzzini, 2012; Iuzzini & Forrest, 2010). The Inconsistency Severity Percentage (Iuzzini & Forrest, 2010) was calculated as

Inconsistency Severity Percentage =
Σ((number of different error types −1; for each phoneme)/
Σ(total number of target opportunities)) × 100

For example, a child produces a list of words that provides 100 consonant opportunities. The child produces 5 different error types for /s/ (e.g., /t, z, p, d/, and lateralized s), 3 different substitutes for /l/ and 4 different substitutes for /k/. Next, "1" is subtracted from the number of error types for each phoneme to allow for one consistent substitution for each errored phoneme. In this example, subtracting "1" results in 4, 2, and 3 error types for /s/, /l/, and /k/ respectively. After this adjustment, the number of error types is then summed across phonemes (i.e., 4+2+3 = 9) and divided by the total number of target opportunities on the assessment (i.e., 100). This child's Inconsistency Severity Percentage would therefore be 9%, calculated as (9/100) × 100.

Previous research on preschool-aged talkers showed that the Inconsistency Severity Percentage (ISP) calculated on responses from the GFTA-2 was highly correlated with this measure calculated on a larger corpus of words from a customized speech battery; consequently, the current study calculated the ISP on responses from the GFTA-2 as well as on a customized speech battery of real and nonwords. See Supplemental Table 5 for stimuli items that contributed to assessment of phonemic inconsistency.

Analyses

Prior to conducting statistical analyses, all speech inconsistency percentage scores were arcsin transformed. Linear mixed models were applied to determine the main effects of independent variables (i.e., group, word length, lexicality, number of trials) on speech inconsistency percentage scores (i.e., phonemic inconsistency, token-to-token word inconsistency, token-to-token inconsistency of "buy Bobby a puppy"), while controlling for age. Although the groups were equivalent for age, there was a broad age range within and across our groups. Consequently, we controlled for age in our mixed model analyses to ensure that significant group or condition effects were not impacted by age.

Because it can be challenging to differentiate between children with CAS (with and without language impairment) and those with speech delay, we calculated sensitivity and specificity estimates to determine how well each speech inconsistency measure discriminated children with CAS from those with speech delay, as well as relative to those with typically developing speech for control purposes. Sensitivity was calculated as the percentage of children with CAS who were positively identified by the measure of interest (e.g., phonemic inconsistency of multisyllabic nonwords) using a specified cutpoint. Specificity was calculated as the percentage of children with speech delay or typically developing speech who were not falsely identified as having CAS by the measure of interest. Apparent error rates were calculated as the percentage of children in each group who were misdiagnosed by each measure. For instance, a sensitivity level of 70% would correspond to an apparent error rate of 30% for children with CAS. Likewise, a specificity level of 80% for the typically developing group would correspond to an apparent error rate of 20%, indicating that 20% of children with typical development were misidentified as having CAS by the given measure. To date, there are no commonly agreed upon sensitivity and specificity criteria levels but Plante and Vance (1994, 1995) suggest 80% sensitivity and specificity to ensure that children with a specific disorder are properly diagnosed and provided appropriate treatment and that children without that disorder are not misdiagnosed and given an inappropriate treatment. Cutoff scores that captured the largest number of children with CAS and the fewest number of children with speech delay or typical development were determined. We did not include children from the language impaired group in these sensitivity and specificity estimates as children without an apparent speech deficit are not likely at risk for being misdiagnosed with CAS or speech delay.

Reliability

CAS Feature Ratings.

Inter-rater reliability for perceptual feature ratings was calculated on GFTA-2 responses from all participants. The intra-class correlation coefficient (ICC) with absolute error in parenthesis was .93 (.6 features), showing a

high level of agreement for perceptual feature rating using the operational definitions included in Supplemental Table 1.

Transcription.

A second rater broadly-transcribed responses on the customized speech battery for 20% of participants with speech sound disorder ($n = 6$). Inter-rater reliability was calculated using the following formula: Agreements/(Agreements+Disagreements) × 100. The mean level of agreement was 89% showing a high level of agreement between raters for transcription of vowels and consonants in disordered speech.

Results

Participant Demographics

Table 1 includes participant characteristics by group. Kolmogorov-Smirnov tests were used to test the normality assumption for each group, for each of our variables (i.e., age, GFTA-2 percentile score, RIAS standard score, CELF/CELF-P Core Language Standard Score, and number of CAS features). All groups had normal distributions for all variables except for the GFTA-2 for which the CAS group did not evidence a normal distribution. The Levene Statistic revealed that none of the variables met the assumption of homogeneity of variance. Consequently, we used nonparametric Kruskal-Wallis tests to determine group differences for these measures. Results revealed that groups were equivalent in age ($p = .134$), and nonverbal intelligence ($p = .062$) based on standard scores on the RIAS. As expected, groups differed on speech production accuracy ($p < .001$) based on percentile scores on the GFTA-2, and on language abilities ($p < .001$) based on standard scores on the CELF or CELF-P. Post hoc Mann-Whitney U-tests, adjusted for multiple comparisons, revealed that the CAS, CAS+LI and speech delayed groups were equivalent ($p > .05$) for speech severity based on GFTA-2 percentiles and that these disordered groups evidenced significantly lower percentiles than the TD and LI groups. The CAS, speech delayed, and TD groups were equivalent for language ability, and all were significantly higher than the CAS+LI and language impaired (LI) groups. Lastly, the CAS and CAS+LI groups were equivalent for number of CAS features. As expected, these groups evidenced more features than all other groups. In addition, the speech delay group evidenced more features than the TD group. See Table 2 for token-to-token and phonemic inconsistency data for each stimuli type, presented by group.

Token-to-token Inconsistency

Word level.

Token-to-token inconsistency was calculated on repeated productions of monosyllabic and multisyllabic real words. For real words, findings showed main effects of group ($p < .001$) and length ($p < .001$). See Figure 1 for token-to-token inconsistency data presented by group. The main effects will not be discussed further as they were qualified by a significant group by length interaction ($p = .002$). Due to non-normal distributions for these variables, Mann-Whitney U-tests—corrected for multiple comparisons (i.e., p-value of .05/10 comparisons = .005)—were used to test pairwise comparisons while preserving the family-wise error rate. No group differences were found for monosyllabic words. For multisyllabic words, all disordered groups (CAS: $M = 60\%$; $SE = 4$; CAS+LI: $M = 67\%$; $SE = 12$; LI: $M = 28\%$; $SE = 3$; SD: $M = 53\%$; $SE = 5$) were more inconsistent ($p < .001$) than the TD group ($M = 10\%$; $SE = 3$). In addition, children with CAS and CAS+LI were more inconsistent ($p \leq .001$) than children with LI. Children with SD were also more inconsistent ($p = .001$) relative to those with LI. No differences were found among children with CAS, CAS+LI, and SD.

Sensitivity and specificity were calculated for mono- and multisyllabic words to determine if either would be superior in correctly identifying children with CAS while not misclassifying children with speech delay or TD. See Table 3 for sensitivity and specificity estimates for phonemic and token-to-token inconsistency in children with CAS, speech delay, and TD. For monosyllabic real words, a token-to-token inconsistency cut point of 0% was

used to differentiate groups. That is, if a child evidenced any difference in production across the two trials of the monosyllabic words (> 0%), they were assigned to the CAS group. This cutoff accurately identified the majority of children with CAS (Sensitivity = 75%) and misidentified only 30% of children with speech delay (Specificity = 70%). See Table 4 for a cross-table showing sensitivity and specificity estimates for token-to-token inconsistency of monosyllabic real words in identifying children with CAS and speech delay.

It should be noted that two out of nine children with TD were falsely assigned to the CAS group by the 0% cutoff on token-to-token inconsistency of monosyllabic real words. (Specificity = 78%). In addition, repeated production of multisyllabic real words resulted in high rates of token-to-token speech inconsistency among both the CAS and speech delayed groups. A cut point of 50% corresponded to 80% sensitivity but only 30% specificity for speech delayed talkers. None of the children with TD were incorrectly identified as CAS by this cutoff (Specificity = 100%).

Sentence level.

Children were asked to produce the utterance "buy Bobby a puppy" five times. The percentage of inconsistent productions was calculated across two, three, and five trials in order to determine the smallest number of trials that could be used to sensitively and specifically differentiate groups. Mixed model analyses revealed a significant effect of group ($p = .002$) and number of trials ($p < .05$), and an interaction between group and number of trials ($p < .05$). See Figure 2 for group level token-to-token inconsistency data across two, three and five productions. Post hoc Mann Whitney U-tests were used to test the interaction. Due to a small sample size and large intragroup variation, the significance levels for these post hocs did not stand up to a correction for multiple comparisons; consequently, we are reporting the findings from uncorrected significant U-tests ($p < .05$) and will interpret these findings with caution. Tests revealed that for two productions of "buy Bobby a puppy", the LI group ($M = 44\%; SE = 39$) was significantly more inconsistent than the TD ($M = 6\%; SE = 17$) and SD ($M = 0\%; SE = 0$) groups. All other groups were equivalent on inconsistency for two trials. For three trials, the LI group ($M = 41\%; SE = 32$) was more inconsistent than the TD ($M = 4\%; SE = 11$) and SD groups ($M = 3\%; SE = 10$) and the CAS+LI group ($M = 30\%; SE = 29$) was also more inconsistent than the SD group. For five repetitions, the CAS ($M = 18\%; SE = 15$), CAS+LI ($M = 30; SE = 24$) and LI groups ($M = 29\%; SE = 27$) were more inconsistent than the SD group ($M = 4\%; SE = 8$). The CAS+LI and LI groups were also more inconsistent than the TD group ($M = 4\%; SE = 9$). Nonparametric related-samples Friedman's Two-Way Analysis of Variance by Ranks test revealed that none of the groups showed an effect of number of productions on their inconsistency of the phrase "buy Bobby a puppy".

Sensitivity and specificity were then calculated for two, three, and five repetitions of the phrase 'buy Bobby a puppy.' Sensitivity of identifying children with CAS increased with number of trials such that 2 repetitions identified 50% of children with CAS, 3 repetitions identified 60% and 5 repetitions identified 70% of children with CAS. Specificity for children with speech delay decreased slightly with number of repetitions where two repetitions had 100% specificity, three had 90%, and five repetitions had 80% specificity. Specificity for typically developing children also decreased slightly with multiple repetitions where two and three repetitions yielded 89% specificity, and five repetitions yielded 78% specificity. Based on these findings, five repetitions of 'buy Bobby a puppy' yielded the best combination of sensitivity (70%) and specificity (80%) for correctly identifying children with CAS and not misclassifying children with speech delay or typical development. See Table 5 for a cross-table showing sensitivity and specificity estimates for token-to-token inconsistency of "buy Bobby a puppy" in identifying children with CAS and speech delay.

Phonemic Inconsistency

Phonemic Inconsistency of multisyllabic real and nonwords.

Phonemic inconsistency was calculated on the GFTA-2 responses, and on the multisyllabic real and nonwords. Mixed model analyses revealed a significant effect of group ($p < .001$), word length ($p < .001$), and lexicality ($p < .001$), which were qualified by a significant interaction between group and lexicality ($p < .001$). No interaction was observed between group and word length. See Figure 3 for phonemic inconsistency data presented by group. Due to non-normal distributions, nonparametric posthoc tests were used to determine within and between group differences for the group by lexicality interaction. All groups showed an effect (p < .008) of lexicality where nonwords were more inconsistent than real words. Tests were adjusted for multiple comparisons (i.e., significance level, .05, divided by the number of tests, $10 = .005$). The CAS group (RW: $M = 7\%$; $SE = 5$; NW: $M = 38\%$; $SE = 23$) was more inconsistent than the TD group on real ($M = 1\%$; $SE = 1$) and nonwords ($M = 7\%$; $SE = 6$) and more inconsistent than the LI group for real words only (M = 1%; SE = 1). The CAS+LI group (RW: $M = 8\%$; $SE = 4$; NW: $M = 49\%$; $SE = 13$) was more inconsistent than the LI (RW: $M = 1\%$; $SE = 1$; NW: $M = 18\%$; $SE = 8$) and TD groups for real and nonwords. The LI group was more inconsistent than the SD group for real words ($M = 5\%$; $SE = 4$) but not for nonwords ($M = 30\%$; $SE = 21$). Lastly, the SD group was more inconsistent than the TD group for both real and nonwords.

A cut point of > 0% was used to determine sensitivity and specificity for phonemic inconsistency of the GFTA-2 responses. There was a significant overlap between the inconsistency percentages of the CAS and speech delay groups. A cut point of > 0% inconsistency corresponded to 55% sensitivity and 30% specificity for the CAS and speech delayed groups. In contrast, none of the children with typical development were falsely grouped by this measure (Specificity = 100%).

For multisyllabic real words, we selected a cutoff of ≥ 5%, which corresponded to 75% sensitivity to identify CAS. Four out of ten children with speech delay were falsely assigned to the CAS group by this cutoff (Specificity = 60%); in contrast, none of the children in the TD group were identified as being inconsistent by this measure (Specificity = 100%). For phonemic inconsistency of multisyllabic nonwords, a cut point of 29% was selected. Sensitivity and specificity estimates were identical to those observed for multisyllabic real words.

Discussion

This study sought to determine if speech inconsistency is a core feature of CAS or if it is predominantly driven by comorbid language impairment, which affects a large subset of children diagnosed with CAS. We also aimed to determine the diagnostic efficacy of speech inconsistency for differentiating children with CAS from those with speech delay. Results varied dependent on stimuli items. Specifically, children with CAS and CAS+LI were more inconsistent than children with language impairment at the phonemic level and for token-to-token production of multisyllabic words. In contrast, for repeated productions of “buy Bobby a puppy”, children with language impairment evidenced an equivalent level of speech inconsistency relative to children with CAS and CAS+LI. These data suggest that for children with CAS (i.e., CAS and CAS+LI), inconsistency is evident across a breadth of stimuli whereas inconsistency in children with language impairment is heavily dependent on stimuli items, and in this study, tended to be associated with phrase level context. In addition, token-to-token inconsistency of the simple phrase ‘buy Bobby a puppy’ was fairly sensitive (70%) and specific (80%) in differentiating between the CAS and speech delayed groups, whereas other stimuli (e.g., token-to-token inconsistency of multisyllabic words) had similar sensitivity but poor specificity. These data support the diagnostic efficacy of speech inconsistency measures for differentiating school-aged talkers with CAS and speech delay while emphasizing the importance of stimuli selection.

Is Speech Inconsistency a Core Feature of CAS?

The current findings contribute to the body of literature that promotes speech inconsistency as a feature central to the CAS profile (e.g., Marquardt et al., 2004). The CAS and CAS+LI groups were statistically equivalent on all experimental measures, and both groups evidenced speech inconsistency across the full range of stimuli tested, including monosyllabic real words. Although high motor performance variability is often considered a limiting characteristic of many neurologically disordered populations such as developmental coordination disorder (Smits-Engelsman & Wilson, 2013; van Waelvelde, De Weerdt, De Cock, Peersman, & Smits-Engelsman, 2004) and cerebral palsy (Chen & Yang, 2007), it is also considered an essential and adaptive characteristic of normal motor development and skill acquisition (Forrest, Weismer, Elbert, & Dinnsen, 1994; Green & Nip, 2010; Thelen & Smith, 1994). Speech variability in young children, for example, increases during emergence of two-word utterances (Sosa & Stoel-Gammon, 2006). Transient spikes in performance variability are observed during development and may reflect a state of transitional knowledge (Green et al., 2002; Iuzzini-Seigel et al., 2015a) or flexibility of a system to adapt to changing environments or contexts (Thelen & Smith, 1994). In this study, however, the speech inconsistency observed in children with CAS and CAS+LI could not be solely attributed to developmental factors because it was greater than that of age-matched, typically developing peers. Different underlying mechanisms for speech inconsistency in children with CAS have been proposed compared to those suggested for typically developing children. For example, Terband and Maassen (2010) suggest that a somatosensory deficit of the tongue or palate, or neural noise could limit a speaker's ability to establish stable speech motor programs, which would result in inconsistent speech errors as a child with CAS tries to produce accurate speech.

We observed an interesting trend in the current study that suggests that for some types of stimuli, speech inconsistency tended to be higher among children with CAS+LI relative to those with CAS only, although this was not a statistically significant effect. Results showed that children with CAS+LI tended to be more inconsistent than those with CAS on production of multisyllabic nonwords, which is not surprising given that children with language impairment are known to often have difficulty producing nonword sequences (Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, & Jones, 2000). This potential interaction needs further investigation in a larger cohort of children as it may have important clinical implications for assessment and treatment in children with CAS. The Rapid Syllable Transition Treatment (ReST; Ballard, Robin, McCabe, & McDonald, 2010; Murray et al., 2015), for instance, is a well-tested treatment protocol for children with CAS that uses nonword treatment targets to promote speech accuracy, prosody, and improved coarticulation. Results of a recent randomized control study (Murray et al., 2015) showed that speech treatment using ReST resulted in gains to treated and untreated nonwords and generalization to real words in a group of 13 children with mild to severe CAS. These children all had normal receptive language and on average, evidenced normal expressive language as well. Consequently, it is unknown if this treatment would be similarly effective for children with CAS with comorbid language impairment, which represents a large percentage of children with CAS.

Children with Language Impairment Evidence Inconsistent Speech on the Phrase "Buy Bobby a Puppy"

Children with language impairment (LI) showed substantial difficulty in producing the short phrase "buy Bobby a puppy" such that their inconsistent performance was equivalent to that of the CAS groups. Examples of errors for the LI group included schwa insertion (e.g., /əbai babi ^ p^pi/), use of spoonerisms (e.g., /bai bapi ^ p^bi/), and assimilation across words (e.g., /bai bapi ^ b^pi). These were interesting findings for a group whose articulation accuracy on the GFTA-2 exceeded that of the typically developing speakers, and were well within the normal range of performance. These results may be explained, in part, by challenges resultant from the phonological similarity of the words "Bobby" and "puppy". That is, the extant research shows that words with high phonological similarity can be more difficult to repeat (e.g., Baddeley, 1966; Conrad, 1964) and that when

phonological similarity is increased, production errors increase as well (e.g., Coltheart, 1993; Conrad, 1965; Hintzman, 1965). Likewise, previous kinematic analysis of children with language disorder has revealed decreased oromotor coordination for this population (Goffman, 1999; Goffman, 2004). To our knowledge, the current study is the first to compare groups of children with CAS and language impairment and has revealed some interesting overlaps in the behavioral speech profiles between these populations. Consequently, future research should further compare these groups to learn more about similarities and differences in these populations and to learn more about the mechanisms underlying each disorder.

Can Speech Inconsistency Contribute to the Differential Diagnosis of CAS and SpeechDelay?

Our findings suggest that speech inconsistency can contribute to the differential diagnosis of school-aged children with CAS and speech delay but that sensitivity and specificity of speech inconsistency measures depend heavily on stimuli selection. For certain stimuli we observed nearly equivalent high levels of speech inconsistency among children with CAS as well as those with speech delay, whereas other stimuli were more efficacious in discriminating between these disorders. Because children with CAS and speech delay performed differently on certain stimuli, this may suggest that different mechanisms underlie inconsistent performance in these disorders.

One possible cause for inconsistency in children with speech delay could be due to poor speech perception (e.g., Edwards, Fox, & Rogers, 2002; Nijland, 2009; Rvachew, 1994; Zuk et al., in revision). Zuk et al. (in revision) found that children with speech delay had significantly poorer speech perception than those with CAS only or typical development. Zuk and colleagues examined speech perception in 47 children with CAS, speech delay, and TD using discrimination of synthesized speech syllables (/d/-/g/) where stimuli differed in F3 onset frequency. Results showed that children with CAS who had speech symptoms only (no comorbid language impairment) performed well and similar to TD controls; that is, they were able to accurately discriminate pairs even when the "different" stimuli were highly similar in F3 onset frequency. In contrast, those with speech delay performed significantly worse relative to children with CAS and TD. These findings suggest that poor speech perception could underlie speech errors and inconsistency in children with speech delay. Zuk et al.'s research (in revision) adds to the extant experimental (e.g., Edwards et al., 2002; Raaymakers & Crul, 1988) and modelling literature (e.g., Terband, Maassen, Guenther, & Brumberg, 2014) that reports a relation between speech perception and speech production (Edwards et al., 2002; Perkell et al., 2004; Terband et al., 2014) in typical and disordered populations.

Alternatively, the children with speech delay who evidenced high token-to-token inconsistency on the multisyllabic real words may have "inconsistent speech disorder" (Dodd & Iacono, 1989), a subtype characterized by high token-to-token inconsistency in the absence of other markers of CAS. Dodd and colleagues attribute inconsistent speech disorder to a deficit at the level of "constructing, storing, and/or retrieving a phonological output plan" (Dodd & Bradford, 2000, p. 190), a distinct mechanism from the transcoding deficits purported to affect children with CAS. Dodd and colleagues report that inconsistent speech disorder can impact consistency at the single word or utterance levels (Holm, Farrier, & Dodd, 2008).

Prior studies that investigated validity and reliability of diagnostic markers of CAS aimed for a sensitivity and specificity goal of 90% or higher, a level that is in line with goals for diagnostic markers in clinical medicine (e.g., Shriberg et al., 1997). Previous research that tested the use of inappropriate stress as a diagnostic marker of CAS revealed 58% sensitivity to accurately identify participants with CAS and 82% specificity to not misdiagnose the participants with speech delay (Shriberg et al., 1997). The majority of stimuli we tested in the current study elicited varied levels of speech inconsistency by both children with CAS and speech delay. Inconsistency of "buy Bobby a puppy" did not yield statistical differences between groups when corrections for multiple comparisons

were made; consequently we reported uncorrected statistical differences between groups. However, this phrase did yield our highest levels of sensitivity and specificity and elicited an inconsistent response in 70% of participants with CAS and in only 20% of participants with speech delay. Although we suggest interpreting our uncorrected statistical comparisons with caution, we believe that the sensitivity and specificity metrics suggest clinical utility for the use of this phrase in contributing to the differential diagnosis of school-aged children with CAS and speech delay.

The phrase “buy Bobby a puppy” is short and contains sequences of voiced and voiceless bilabials and vowels. Because children with CAS often produce errors in voicing (e.g., Iuzzini, 2012; Iuzzini-Seigel et al., 2015b) and vowel production (Iuzzini & Forrest, 2008; Iuzzini-Seigel et al., 2015b), this short phrase may tax areas of struggle for children with CAS, while playing to the strengths of children with speech delay. Bilabial stops and vowels are considered early occurring sounds, typically acquired prior to age 3 (Macken & Barton, 1980; Templin, 1957) but are considered challenging for children with CAS (Iuzzini, 2012; Iuzzini & Forrest, 2008; Lewis et al., 2004). Prior research in children with CAS and speech delay showed that children with CAS evidenced reduced vowel space area (Iuzzini & Forrest, 2008), were late to acquire the voicing contrast (Iuzzini, 2012), and often produced voiced cognates and voicing distortions for voiceless bilabial and alveolar targets (Iuzzini, 2012). The voicing contrast is believed to reflect proficiency in timing and coordination between the glottis and articulators (Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980) and voicing errors provide support of motor programming deficits in children with CAS (Iuzzini, 2012; Iuzzini-Seigel et al., 2015b). By comparison, children with speech delay show proficiency of the voicing contrast and vowel production (Iuzzini, 2012; Iuzzini-Seigel et al., 2015b). Other stimuli we tested were either so simple that talkers with CAS were minimally taxed (i.e., phonemic inconsistency of GFTA-2 responses) or so challenging that talkers with speech delay performed poorly (i.e., token-to-token inconsistency of multisyllabic nonwords). This finding is consistent with recent work (Murray et al., 2015) that showed that token-to-token inconsistency on the Inconsistency Subtest of the Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2002) was not efficacious in differentiating CAS from other speech sound disorders. A bivariate discriminant function analysis revealed only 30% accuracy in discriminating CAS from non-CAS participants when speech inconsistency was entered as a predictor (Murray et al., 2015). Because the Inconsistency Subtest contains a large number of challenging multisyllabic items, it is possible that this test elicited varying levels of speech inconsistency in all groups, whereas results may have differed if different stimuli had been used.

While 70% sensitivity and 80% specificity for token-to-token inconsistency of “buy Bobby a puppy” are fair-good, there is room for improvement. Future research should focus on further optimizing stimuli, perhaps by lengthening the phrase (e.g., Kleinow & Smith, 2000). It is also necessary to determine how many repetitions of a word or phrase are optimal to maximize diagnostic sensitivity and specificity. For the purpose of differentiating talkers with CAS and speech delay, it is essential to develop stimuli that sample areas of weakness in talkers with CAS but which are sufficiently simple for children with speech delay. It is expected that core features, such as speech inconsistency, will be present across talkers with CAS, but that the stimuli and procedures used to elicit such features will vary across ages. For instance, previous work (Iuzzini, 2012) suggests that phonemic inconsistency may be more efficacious in differentiating preschool-aged children with CAS and speech delay, where the current findings suggest that token-to-token inconsistency of certain phrases is efficacious for differentiating school-aged children in these populations.

Conclusion

The current study investigated speech inconsistency in children with CAS, language impairment, speech delay, and typical development. The CAS group comprised two subgroups that included: (1) children with speech symptoms only and (2) children with CAS and comorbid language impairment. Two main questions were

addressed: (1) is speech inconsistency a core feature central to the CAS profile or is it attributed to comorbid language impairment? and (2) can speech inconsistency be used to differentiate children with CAS and speech delay? Results showed that the two CAS subgroups were equivalent on all speech inconsistency measures suggesting that speech inconsistency is a core feature of CAS and cannot be attributed to language impairment in this population. In addition, children with language impairment only were inconsistent in their production of the phrase "buy Bobby a puppy". Finally, data revealed that token-to-token inconsistent production of monosyllabic real words and the phrase "buy Bobby a puppy" were moderately sensitive and specific differential markers for children with CAS and speech delay. The short phrase "buy Bobby a puppy" appears to tax the voicing and vowel contrasts that are known to be challenging for children with CAS, while being sufficiently simple for children with speech delay to perform with ease. In contrast, speech inconsistency calculated on other stimuli such as multisyllabic real and nonwords elicited an inconsistent response in children with CAS and speech delay, making it less efficacious in differentiating between these disorders.

Although the LI, CAS, and CAS+LI groups all evidenced similar levels of speech inconsistency on the phrase "buy Bobby a puppy", performance does not necessarily suggest a common underlying mechanism for CAS and language impairment. For children with CAS, inconsistent productions could result from difficulty planning and programming the articulatory sequences needed to produce the phrase correctly. Here vowel errors could reflect difficulty grading articulatory movements and voicing errors could point to difficulty coordinating the articulatory and phonatory systems (Iuzzini, 2012; Iuzzini-Seigel et al., 2015b). In contrast, speech inconsistency among children with language impairment could reflect a higher order linguistic deficit that impacts their appropriate usage of grammar and their ability to sequence similar phonological units (Rosenthal, 1994).

Future research should continue to optimize assessment stimuli so that researchers and clinicians can confidently differentiate between children with CAS and speech delay. In addition, further investigation of this line of research will help us to better ascertain the underlying basis of speech inconsistency in children with speech and language disorders.

Acknowledgements

This research was supported by the University of Nebraska Health Research Consortium (Co-PIs: Hogan & Green), and the Childhood Apraxia of Speech Association of North America (PI: Iuzzini). The authors wish to thank the following individuals: Kimber Green, Sara Benham, Panying Rong, Dyann Rupp, Tacy Corson, Phoebe Chung, Natalie Vanderveen, Ashley Africa, and Kristin Schneller. Portions of these data were presented at the American Speech, Language and Hearing Association's annual convention (Iuzzini, Hogan, & Green, 2014).

References

Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and language*, 75(1), 17-33.

American Speech-Language-Hearing Association. (2007). Childhood Apraxia of Speech [Technical Report]. Available from www.asha.org/policy.

American Speech-Language-Hearing Association. (1997). Guidelines for audiologic screening [Technical Report]. Available from www.asha.org/policy.

Aram, D. M., & Glasson, C. (1979, November). Developmental apraxia of speech. *Annual Convention of the American Speech-Language-Hearing Association Meeting*, Los Angeles, CA.

Arbib, M. A. (2006). The mirror system hypothesis on the linkage of action and languages. *Action to language via the mirror neuron system*, 3-47.

Archibald, L. M., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, 14(5), 919-924.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly journal of experimental psychology*, 18(4), 302-309.

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53(5), 1227-1245.

Bernstein N. (1967) *The Coordination and Regulation of Movements*. Oxford: Pergamon.

Betz, S. K., & Stoel-Gammon, C. (2005). Measuring articulatory error consistency in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 19(1), 53-66.

Bishop, D. V. (1994). Grammatical errors in specific language impairment: Competence or performance limitations?. *Applied Psycholinguistics*, 15(04), 507-550.

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language*, 9(2), 153-170.

Bracke-Tolkmitt, R., Linden, A., Canavan, A. G. M., Rockstroh, B., Scholz, E., Wessel, K., & Diener, H. C. (1989). The cerebellum contributes to mental skills. *Behavioral Neuroscience*, 103(2), 442.

Bridgeman, E., & Snowling, M. (1988). The perception of phoneme sequence: A comparison of dyspraxic and normal children. *International Journal of Language & Communication Disorders*, 23(3), 245-252.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research*, 48(6), 1378-1396.

Centanni, T. M., Sanmann, J. N., Green, J. R., Iuzzini-Seigel, J., Bartlett, C., Sanger, W. G., & Hogan, T.P. (2015). The role of candidate-gene CNTNAP2 in childhood apraxia of speech and specific language impairment. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168(7), 536-543.

Cermak, S. A., Ward, E. A. and Ward, L. M. (1986). The relationship between articulation disorders and motor coordination in children. *American Journal of Occupational Therapy*, 40, 546–550.

Chen, Y. P., & Yang, T. F. (2007). Effect of task goals on the reaching patterns of children with cerebral palsy. *Journal of Motor Behavior*, 39(4), 317-325.

Cohen, R. G., & Sternad, D. (2009). Variability in motor learning: relocating, channeling and reducing noise. *Experimental Brain Research*, 193(1), 69-83.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on shortterm-memory recall of repeated and novel word lists. *Memory & Cognition*, 21(4), 539-545.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British journal of Psychology*, 55(1), 75-84.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4(3), 161-169.

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.

Dodd, B. & Bradford, A. (2000). A comparison of three therapy methods for children with different types of developmental phonological disorder. *International Journal of Language & Communication Disorders*, 35(2), 189-209.

Dodd, B., & Iacano, T. (1989). Phonological disorders in children: Changes in phonological process use during treatment. *International Journal of Language & Communication Disorders*, 24(3), 333-352.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology* (DEAP). Psychology Corporation.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. *Journal of Speech, Language, and Hearing Research*, 45(2), 231-242.

Ekelman, B. L., & Aram, D. M. (1984). Spoken syntax in children with developmental verbal apraxia. *Seminars in Speech and Language*, 5(2), 97-110).

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology*, 12(3), 376-380.

Forrest, K., Dinnsen, D. A., & Elbert, M. (1997). Impact of substitution patterns on phonological learning by misarticulating children. *Clinical Linguistics & Phonetics*, 11(1), 63-76.

Forrest, K., & Elbert, M. (2001). Treatment for phonologically disordered children with variable substitution patterns. *Clinical Linguistics & Phonetics*, 15(1-2), 41-45.

Forrest, K., Elbert, M., & Dinnsen, D. A. (2000). The effect of substitution patterns on phonological treatment outcomes. *Clinical Linguistics & Phonetics*, 14(7), 519-531.

Forrest, K., Weismer, G., Elbert, M., & Dinnsen, D. A. (1994). Spectral analysis of targetappropriate/t/and/k/produced by phonologically disordered and normally articulating children. *Clinical Linguistics & Phonetics*, 8(4), 267-281.

Gierut, J., Elbert M., & Dinnsen, D. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech, Language, and Hearing Research*, 30(4), 462-479.

Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research*, 42(6), 1499-1517.

Goffman, L. (2004). Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research*, 47(5), 1088-1102.

Goffman, L. (2010). Dynamic interaction of motor and language factors in development. B. Maassen, P. H. H. M. van Lieshout, R. Kent, & W. Hulstijn (Eds.), Sp*eech motor control: New developments in applied research*. Oxford: Oxford University Press.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation*, 2nd Edition. Circle Pines, MN: Pearson.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, 39(1), 1-50.

Green, J. R., & Nip, I. S. B. (2010). Organization principles in the development of early speech. In B. Maaseen & P. H. H. M. van Lieshout, eds., *Speech motor control: New developments in basic and applied research* (pp. 171–188). Oxford, UK: Oxford University Press.

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45(1), 66-79.

Grigos, M. I. (2009). Changes in articulator movement variability during phonemic development: a longitudinal study. *Journal of Speech, Language, and Hearing Research*, 52(1), 164-177.

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

Hill, E. L., Bishop, D. V. M., & Nimmo-Smith, I. (1998). Representational gestures in developmental coordination disorder and specific language impairment: Error-types and the reliability of ratings. *Human Movement Science*, 17(4), 655-678.

Hintzman, D. L. (1965). Classification and aural coding in short-term memory. *Psychonomic Science*, 3(1-12), 161-162.

Hodge, M. (1994). Assessment of developmental apraxia of speech: A rationale. *Clinics in Communication Disorders*, 4, 91–101.

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: normative data. *International Journal of Language & Communication Disorders*, 42(4), 467-486.

Holm, A., Farrier, F., & Dodd, B. (2008). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language & Communication Disorders*, 43(3), 300-322.

Iuzzini, J. (2012). *Inconsistency of speech in children with childhood apraxia of speech, phonological disorders, and typical speech*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, Indiana.

Iuzzini, J. & Forrest, K. (2008, March). Acoustic vowel area in phonologically disordered, CAS, and normal speech. Poster presented at the Conference for Motor Speech, Monterey, CA.

Iuzzini, J., & Forrest, K. (2010). Evaluation of a combined treatment approach for Childhood Apraxia of Speech. *Clinical Linguistics & Phonetics*, 24(4-5), 335-345.

Iuzzini, J., Hogan, T. P., & Green, J. R. (2014, November). Inconsistency is task dependent in children with CAS and speech delay. Technical session at the annual convention of the American Speech-Language, and Hearing Association, Orlando, Florida.

Iuzzini-Seigel, J., Hogan, T.P., Rong, P. & Green, J.R. (2015a). Longitudinal development of speech motor control: Motor and linguistic factors. *Journal of Motor Learning and Development*, 3(1), 53-68.

Iuzzini-Seigel, J., Hogan, T.P., Guarino, A. & Green, J.R. (2015b). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders*, 54, 32-42.

Jäncke, L., Siegenthaler, T., Preis, S., & Steinmetz, H. (2007). Decreased white-matter density in a left sided fronto-temporal network in children with developmental language disorder: evidence for anatomical anomalies in a motor-language network. *Brain and language*, 102(1), 91-98.

Johnston, R. B., Stark, R. E., Mellits, E. D. and Tallal, P. (1981). Neurological status of language impaired and normal children. *Archives of Neurology*, 10, 159–163.

Kent, R. D. (2004). The uniqueness of speech among motor systems. *Clinical Linguistics & Phonetics*, 18(6-8), 495-505.

Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 43(2), 548-559.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1991). The human cerebro-cerebellar system: its computing, cognitive, and language skills. *Behavioural brain research*, 44(2), 113-128.

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42(3), 678-689.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Iyengar, S. K., & Taylor, H. G. (2004). School-age followup of children with childhood apraxia of speech. *Language, Speech, and Hearing Services in Schools*, 35(2), 122-140.

Maassen, B., Nijland, L., & Van Der Meulen, S. (2001). Coarticulation within and between syllables by children with developmental apraxia of speech. *Clinical linguistics & phonetics*, 15(1-2), 145-150.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7(03), 433-458.

Macrae, T., Tyler, A. A., & Lewis, K. E. (2014). Lexical and phonological variability in preschool children with speech sound disorder. *American Journal of Speech-Language Pathology*, 23(1), 27-35.

Malemenholt, A., Lohmander, A., & McAllister, A. (2016). Childhood apraxia of speech: A survey of praxis and typical speech characteristics. Logopedics Phoniatrics Vocology, 1-9.

Maner, K. J., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, 43(2), 560-573.

Marquardt, T. P., Jacks, A., & Davis, B. L. (2004). Token-to-token variability in developmental apraxia of speech: three longitudinal case studies. *Clinical Linguistics & Phonetics*, 18(2), 127-144.

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. Journal of Speech, *Language, and Hearing Research*, 44(2), 416-433.

Millspaugh, S. & Weiss, D. (2006, November). Diagnosing childhood apraxia of speech: A national survey of speech-language-pathologists. Poster presentation at the annual convention of the American Speech, Language, and Hearing Association, Miami Beach, Florida.

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.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders?. *TRENDS in Neurosciences*, 30(4), 135-141.

Nijland, L. (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics*, 23(3), 222-239.

Nijland, L., Maassen, B., Meulen, S. V. D., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16(6), 461-483.

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

Nip, I. S., Green, J. R., & Marx, D. B. (2011). The co-emergence of cognition, language, and speech motor control in early development: A longitudinal correlation study. *Journal of Communication Disorders*, 44(2), 149-160.

Ojemann, G. A. (1984). Common cortical and thalamic mechanisms for language and motor functions. American *Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 246(6), R901-R903.

Paul, R. (2007). Language disorders from infancy through adolescence: Assessment & intervention. Elsevier Health Sciences.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4), 2338-2344.

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests A Data-Based Approach. *Language, Speech, and Hearing Services in Schools*,25(1), 15-24.

Plante, E., & Vance, R. (1995). Diagnostic accuracy of two tests of preschool language. *American Journal of Speech-Language Pathology*, 4(2), 70-76.

Powell, R. P., & Bishop, D. V. (1992). Clumsiness and perceptual problems in children with specific language impairment. *Developmental Medicine & Child Neurology*, 34(9), 755-765.

Raaymakers, E. M., & Crul, T. A. (1988). Perception and production of the final/s-ts/contrast in Dutch by misarticulating children. *Journal of speech and Hearing Disorders*, 53(3), 262-270.

Reynolds, C.R., & Kamphaus, R.W. (2003). Reynolds Intellectual Assessment Scales. Lutz, FL: PAR Inc.

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech, Language, and Hearing Research*, 38(4), 850-863.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1239-1257.

Rintala, P., Pienimaki, K., Ahonen, T. and Cantell, M. (1998). Effects of a psychomotor training programme on motor skill development in children with developmental language disorders. *Human Movement Science*, 17, 721–737.

Rosenthal, J. B. (1994). Rate control therapy for developmental apraxia of speech. *Clinics in communication disorders*, 4(3), 190.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research,* 37(2), 347-357.

Schumacher, J. G., McNeil, M. R., Vetter, D. K., & Yoder, D. E. (1986, November). Articulatory consistency and variability in apraxic and non-apraxic children. In *Annual Convention of the American Speech-Language-Hearing Association*, Detroit, MI.

Semel, E., Wiig, E.H., Secord, W. A. (2003). Clinical Evaluation of Language Fundamental-4th Edition. San Antonio, TX: Pearson.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, 36(1), 105-140.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research*, 40(2), 273-285.

Shriberg, L. D., Green, J. R., Campbell, T. F., Mcsweeny, J. L., & Scheer, A. R. (2003). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Clinical Linguistics & Phonetics*, 17(7), 575-595.

Shriberg, L.D., Jakielski, J.J., & Strand, E.A. (2010, November). Diagnostic markers of childhood apraxia of speech. Presentation at the American Speech-Language-Hearing Association Convention, Philadelphia, PA.

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.

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, 487-519.

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, 41(1), 18-30.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental psychobiology*, 45(1), 22-33.

Smits-Engelsman, B., & Wilson, P. H. (2013). Noise, variability, and motor performance in developmental coordination disorder. *Developmental Medicine & Child Neurology*, 55(s4), 69-72.

Sosa, A. V., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*, 33(1), 31-50.

Stackhouse, J. (1992). Developmental verbal dyspraxia I: A review and critique. *International Journal of Language & Communication Disorders*,27(1), 19-34.

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, 42(2), 497-506.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences*, 682(1), 27-47.

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. Minneapolis, MN: University of Minnesota Press.

Terband, H., & Maassen, B. (2010). Speech motor development in childhood apraxia of speech: generating testable hypotheses by neurocomputational modeling. *Folia Phoniatrica et Logopaedica*, 62(3), 134-142.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2014). Auditory–motor interactions in pediatric motor speech disorders: Neurocomputational modeling of disordered development. *Journal of communication disorders*,47, 17-33.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thoonen, G., Maassen, B., Gabreels, F., Schreuder, R., & Swart, B. (1997). Towards a standardized assessment procedure for developmental apraxia of speech. *International Journal of Language & Communication Disorders*, 32(1), 37-60.

Tyler, A. A., Lewis, K. E., & Welch, C. M. (2003). Predictors of phonological change following intervention. *American Journal of Speech-Language Pathology*, 12(3), 289-298.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399-433.

Van Waelvelde, H., De Weerdt, W., De Cock, P., Peersman, W., & Smits-Engelsman, B. C. M. (2004). Ball catching performance in children with developmental coordination disorder. *Adapted Physical Activity Quarterly*, 21(4), 348-363.

Velleman, S. L., & Strand, K. (1994). Developmental verbal dyspraxia. Child phonology: Characteristics, assessment, and intervention with special populations, 110-139.

Vuolo, J. & Goffman, L. (November, 2014). Phonetic and speech motor variability in children with childhood apraxia of speech and phonological Disorders. Technical session at the American Speech-Language Hearing Association Convention, Orlando, FL.

Walsh, B., Smith, A., & Weber-Fox, C. (2006). Short-term plasticity in children's speech motor systems. *Developmental psychobiology*, 48(8), 660-674.

Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43(4), 865-878.

Wiig, E.H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamental-Preschool*-2nd Edition. Bloomington, MN: Pearson.

Zelaznik, H. N., & Goffman, L. (2010). Generalized motor abilities and timing behavior in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53(2), 383-393.

Yao-Tresguerres, S., Iuzzini, J., & Forrest, K. (2009, November). The effect of therapy on children's use of inconsistent substitutes. Poster presented at the annual convention of the American Speech, Language, and Hearing Association, New Orleans, Louisiana.

Zuk, J., Iuzzini-Seigel, J., Green, J., Cabbage, K. & Hogan, T. (In revision). Auditory perception in children with childhood apraxia of speech: language matters.

Figure 1. Mean token-to-token inconsistency for mono- and multisyllabic real words, and repeated production of the phrase 'buy Bobby a puppy.' Error bars represent standard error of the mean. For monosyllabic words and the phrase level, the CAS group was more inconsistent than the speech delay and typically developing groups, who were equivalent. For multisyllabic words, the CAS and speech delay groups were equivalent, and more inconsistent than the typically developing group.

Figure 2. Mean token-to-token inconsistency for repeated productions of "buy Bobby a puppy" where the independent variable was number of productions (i.e., 2, 3, or 5 productions). Error bars represent standard error of the mean. None of the groups showed an effect of number of productions on inconsistency of this phrase. For 2, 3, and 5 productions of "buy Bobby a puppy", the language impaired group was more inconsistent than the typically developing and speech delayed groups. For 3 and 5 trials, the comorbid CAS+LI group was more inconsistent than the speech delayed group, and more inconsistent than the typically developing group for 5 trials only. The CAS group was also more consistent than the speech delayed group for 5 trials.

Figure 3. Mean phonemic inconsistency calculated on responses from the GFTA-2 and multisyllabic real and nonwords. Error bars represent standard error of the mean. For all stimuli the typically developing group was less inconsistent relative to the CAS and speech delayed groups. The speech delayed group was less inconsistent than the CAS group on multisyllabic nonwords only.

Supplemental Table 1. Operational definitions for CAS feature ratings.

Supplemental Table 2. CAS features by participant.

Supplemental Table 3. Speech, language, and cognition scores by participant.

Supplemental Table 4. Mono- and multisyllabic real word stimuli used for token-to-token inconsistency assessment.

Supplemental Table 5. Multisyllabic real and nonwords stimuli used for phonemic inconsistency analyses.

Table 1. Participant characteristics by group.

Measure	TD ($n = 9$)	CAS ($n = 10$)	CAS+LI ($n = 10$)	LI ($n = 9$)	SD ($n = 10$)
Age in months	140 (43)$_a$	119 (46)$_a$	120 (28)$_a$	122(12)$_a$	102 (14)$_a$
GFTA-2 Percentile	28 (13)$_a$	4 (5)$_b$	3 (2)$_b$	34 (6)$_a$	4 (3)$_b$
CELF Core SS	110 (9)$_a$	107 (14)$_a$	65 (17)$_b$	74(10)$_b$	106 (15)$_a$
RIAS SS	108 (16)$_a$	112 (9)$_a$	97 (17)$_a$	103(2)$_a$	113 (17)$_a$
# CAS features[a]	1 (1)$_a$	8 (2)$_b$	7 (1)$_b$	1 (1)$_a$	3 (1)$_c$

Note. Groups that share subscript are statistically equivalent for the measure of interest. Standard deviations are in parentheses. TD = typically developing; CAS = childhood apraxia of speech; CAS+LI = CAS+ language impairment; LI = language impairment; SD = speech delay; GFTA-2 = Goldman Fristoe Test of Articulation- 2nd Edition (Goldman & Fristoe, 2000); CELF = Clinical Evaluation of Language Fundamentals- 4th Edition (Semel et al., 2004) or Clinical Evaluation of Language Fundamentals-Preschool (Wiig et al, 2004) depending on participant's age; RIAS = Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003); SS = standard score; * = Feature list and assessment procedure adapted from Shriberg et al., 2011.

Table 2. Mean phonemic and token-to-token inconsistency percentages by group. Standard deviations in parentheses.

Measure	TD ($n = 9$)	CAS ($n = 10$)	CAS+LI ($n = 10$)	LI ($n = 9$)	Speech delay ($n = 10$)
Phonemic inconsistency					
GFTA-2 responses	0 (0)	1 (2)	1 (2)	0 (1)	1 (1)
Multisyllabic real words	1 (1)	7 (5)	8 (4)	1 (1)	5 (4)
Multisyllabic nonwords	7 (6)	38 (23)	49 (13)	18 (8)	30 (21)
Token-to-token inconsistency					
Monosyllabic real words	6 (11)	23 (25)	33 (31)	13 (19)	8 (12)
Multisyllabic real words	10 (8)	56 (12)	58 (24)	27 (7)	50 (13)
Utterance, two trials	6 (17)	30 (35)	30 (35)	44 (39)	0 (0)
Utterance, three trials	4 (11)	23 (27)	30 (29)	41 (32)	3 (10)
Utterance, five trials	4 (9)	18 (15)	30 (24)	29 (27)	4 (8)

Note: TD = typically developing; CAS = childhood apraxia of speech; CAS+LI = CAS+ language impairment; LI= language impairment; SD = speech delay.

Table 3. Sensitivity and specificity estimates for phonemic and token-to-token inconsistency measures.

Measure	Sensitivity (CAS)	Specificity (speech delay)	Specificity (TD)
Phonemic inconsistency			
GFTA-2 (cut point > 0)	55	30	100
Multisyllabic real words (cut point ≥ 5)	75	60	100
Multisyllabic nonwords (cut point ≥ 29)	75	60	100
Token-to-token inconsistency			
Monosyllabic real words (cut point > 0)	75	70	78
Multisyllabic words (cut point ≥ 50)	80	30	100
Utterance, two trials (cut point > 0)	50	100	89
Utterance, three trials (cut point > 0)	60	90	89
Utterance, five trials (cut point > 0)	70	80	78

Note. CAS = childhood apraxia of speech (includes children with speech symptoms only and those with CAS + language impairment); TD = typically developing; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000).

Table 4. Cross-table showing the number of children with a diagnosis of CAS and speech delay who were assigned to the CAS and speech-delayed groups based on their token-to-token inconsistency of monosyllabic real words.

	Clinical diagnosis	
Inconsistency score results	CAS	Speech delay
CAS	15 (true-positives)	3 (false-positives)
Speech delay	5 (false-negatives)	7 (true-negatives)

Note. Sensitivity = 75%; specificity = 70%. CAS = childhood apraxia of speech.

Table 5. Cross-table showing the number of children with a diagnosis of CAS and speech delay who were assigned to the CAS and speech delay groups based on their token-to-token inconsistency of "buy Bobby a puppy."

	Clinical diagnosis	
Inconsistency score results	CAS	Speech delay
CAS	14 (true-positives)	2 (false-positives)
Speech delay	6 (false-negatives)	8 (true-negatives)

Note. Sensitivity = 70%; specificity = 80%. CAS = childhood apraxia of speech.

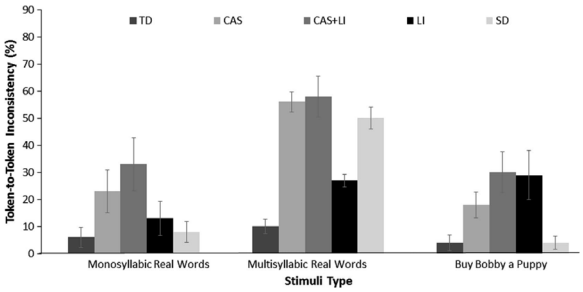

Figure 1. Mean token-to-token inconsistency for mono- and multisyllabic real words and repeated production of the phrase "buy Bobby a puppy." Error bars represent the standard error of the mean. For monosyllabic words and the phrase level, the childhood apraxia of speech (CAS) group was more inconsistent than the speech delay and typically developing (TD) groups, who were equivalent. For multisyllabic words, the CAS and speech delay groups were equivalent and more in consistent than the (TD) group. CAS+LI = CAS + language impairment; LI = language impairment; SD = speech-delayed.

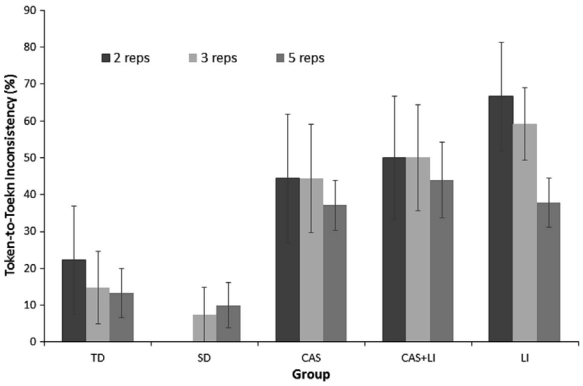

Figure 2. Mean token-to-token inconsistency for repeated productions of "buy Bobby a puppy" in which the independent variable was number of productions (i.e., two, three, or five productions). Error bars represent the standard error of the mean. None of the groups showed an effect of number of productions on inconsistency of this phrase. For two, three, and five productions of "buy Bobby a puppy," the language-impaired (LI) group was more inconsistent than the typically developing (TD) and speech-delayed (SD) groups. For three and five trials, the comorbid childhood apraxia of speech (CAS) + LI group was more inconsistent than the SD group and more inconsistent than the TD group for five trials only. The CAS group was also more consistent than the SD group for five trials.

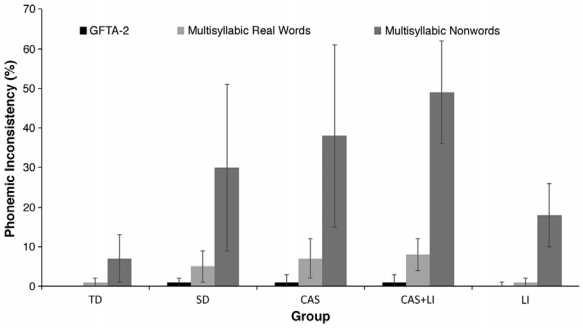

Figure 3. Mean phonemic inconsistency calculated on responses from the Goldman-Fristoe Test of Articulation–Second Edition and multisyllabic real and nonwords. Error bars represent the standard error of the mean. For all stimuli, the typically developing (TD) group was less inconsistent relative to the childhood apraxia of speech (CAS) and speech-delayed (SD) groups. The SD group was less inconsistent than the CAS group on multisyllabic nonwords only. GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); LI = language impairment.