Journal of Communication Disorders 87 (2020) 106033

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis

Factor analysis of signs of childhood apraxia of speech

Karen V. Chenausky[a,b,f,*], Amanda Brignell[c,d], Angela Morgan[c,d], Danielle Gagné[a], Andrea Norton[e], Helen Tager-Flusberg[f], Gottfried Schlaug[b,e], Aaron Shield[g], Jordan R. Green[a]

[a] Speech and Feeding Disorders Lab, Massachusetts General Hospital Institute of Health Professions, Boston, USA
[b] Department of Neurology, Harvard Medical School, Boston, USA
[c] Speech and Language, Murdoch Children's Research Institute, Melbourne, Australia
[d] Department of Audiology and Speech Pathology, University of Melbourne, Melbourne, Australia
[e] Brain Repair and Neurorestoration Center, Baystate Medical Center, Springfield Massachusetts, USA
[f] Center for Autism Research Excellence, Department of Pyschological and Brain Sciences, Boston University, Boston, USA
[g] Department of Speech Pathology & Audiology, Miami University, Oxford, Ohio, USA

ARTICLE INFO

Keywords:
Childhood apraxia of speech
Autism spectrum disorder
Minimally verbal

ABSTRACT

Purpose: To investigate the latent factors underlying signs of childhood apraxia of speech (CAS) in a group of 57 children with CAS.
Method: The speech of 57 children with CAS (aged 3;5–17;0) was coded for signs of CAS. All participants showed at least five signs of CAS and were judged to have CAS by speech pathologists experienced in pediatric speech disorders. Participants were selected to represent a range of severity of CAS: 30 children were verbal and 27 were minimally verbal with comorbid autism. Participants' scores for each sign (the number of times that sign appeared during a child's speech sample) were converted to z-scores, then entered as variables into an exploratory factor analysis. Models were compared using the Akaike Information Criterion (AIC).
Results: The three-factor model had the lowest AIC and best fit the data. After oblique rotation, syllable segmentation, slow rate, and stress errors loaded most highly on Factor 1. Groping, addition of phonemes other than schwa, and difficulty with coarticulation loaded most highly on Factor 2. Variable errors loaded most highly on Factor 3. Thus, factors were interpreted as being associated with (1) prosody, (2) coarticulation, and (3) inconsistency.
Conclusions: Results are consistent with the three consensus criteria for CAS from the American Speech-Language-Hearing Association: Inappropriate prosody, disrupted coarticulatory transitions, and inconsistent errors on repeated tokens. High loading of the syllable segmentation sign on the inappropriate prosody factor also supports the use of a pause-related biomarker for CAS.

1. Introduction

Childhood apraxia of speech (CAS) is a complex neurodevelopmental disorder in which the ability to plan and sequence speech

* Corresponding author at: 36 1st Avenue, Boston, MA, 02129, USA.
E-mail addresses: kvchenausky@mghihp.edu, kcvchenausky@mghihp.edu (K.V. Chenausky), amanda.brignell@mcri.edu.au (A. Brignell), angela.morgan@mcri.edu.au (A. Morgan), dcgagne@mghihp.edu (D. Gagné), acnorton@bidmc.harvard.edu (A. Norton), htagerf@bu.edu (H. Tager-Flusberg), gschlaug@bidmc.harvard.edu (G. Schlaug), shielda@miamioh.edu (A. Shield), jgreen2@mghihp.edu (J.R. Green).

https://doi.org/10.1016/j.jcomdis.2020.106033
Received 6 November 2019; Received in revised form 9 July 2020; Accepted 9 July 2020
Available online 18 August 2020

movements is impaired, thereby decreasing the precision, consistency, and intelligibility of speech. Pure apraxia (i.e., no other diagnoses) has been estimated to affect 1–2 children per thousand (Shriberg, Aram, & Kwiatkowski, 1997); however, recent research has identified it as a significant comorbidity in children with other neurodevelopmental disorders, such as galactosemia (Shriberg, Potter, & Strand, 2011), 16p11.2 deletion (Fedorenko et al., 2016), *FOXP2* mutations (Morgan, Fisher, Scheffer, & Hildebrand, 2016), and minimally verbal children with autism spectrum disorder (ASD; Chenausky, Brignell, Morgan, & Tager-Flusberg, 2019). As the field has moved from characterizing “CAS pure” (CAS as a sole diagnosis) to “CAS plus” (CAS with comorbidities), the question of what unifies the definition of the disorder across these different populations arises. The answer to this question is complicated by existing diagnostic challenges.

In this study, we address the need for an evidence-based pathognomonic marker or cluster of markers for CAS by examining the factors underlying observable signs of CAS in a large group of children who meet criteria for the disorder. A marker, in this context, is an observable sign that indicates a disorder. Optimally, the marker should be present in every individual with the disorder and not present in any individuals without the disorder. In the case of CAS, we argue that it is vital to include children whose symptoms range from mild to very severe, as well as children with CAS pure and children with CAS plus, in order to present and consider the most comprehensive view of the disorder possible. Omitting one group of children or the other might mean failing to account for a manifestation of the condition that clinicians nevertheless frequently encounter. This in turn could contribute to the misdiagnosis of children with certain comorbidities and lead to delays in finding the right treatment (or, worse, to providing the wrong treatment).

1.1. *Challenges in diagnosing CAS*

CAS is a clinical diagnosis made by speech-language pathologists with experience in pediatric motor speech disorders (Strand, McCauley, & Weigand, 2013). Diagnosis is based on the presence of multiple abnormal speech characteristics including consonant or vowel distortions, difficulty achieving the oral postures required to initiate speech, or pausing between syllables of a word. Several challenges present themselves during this process.

One challenge is that there are no universally agreed-upon diagnostic criteria for CAS. The 2007 Technical Report on CAS from the American Speech-Language-Hearing Association presented three main features that were arrived at by consensus, while noting that they are neither necessary nor sufficient for diagnosis (American Speech-Language-Hearing Association, 2007). Those features are (1) inconsistent errors over repeated productions of the same target; (2) lengthened and disrupted coarticulatory transitions between speech sounds; and (3) inappropriate prosody, especially word- or sentence-level stress. More comprehensive lists of signs of CAS have been presented by others. For example, Iuzzini-Seigel, Hogan., and Guarino (2015) employed a list of 11 signs and included

Table 1
Lists of signs of Childhood Apraxia of Speech.

ASHA Consensus Signs (ASHA, 2007)	Iuzzini-Seigel et al. (2015) Signs	Shriberg et al. (2017a, 2017b) Signs	Fedorenko et al. (2016) Signs
Lengthened and disrupted coarticulatory transitions	Groping	Groping	Silent groping
			Difficulty sequencing phonemes and syllables
	Voicing errors	Voicing errors	Voicing errors
	Increased difficulty with multisyllabic words	Increased difficulty with multisyllabic words	Increased errors with increased word length/complexity
	Difficulty achieving initial or transitionary movement gestures	Difficulty achieving initial or transitionary movement gestures	Difficulty achieving initial articulatory configurations
	Intrusive schwa	Intrusive schwa	Epenthesis/syllable repetition
			Phoneme/syllable repetition
			Difficulty maintaining syllable integrity
			Addition errors
			Metathesis
			Frequent phoneme omissions
			Prolongations
	Vowel error	Vowel distortions	Nonphonemic/distorted substitutions
	Consonant distortion	Distorted substitutions	Slowed/disrupted DDK sequences
	Nasality disturbance	Distorted substitutions	
		Slow speech/DDK[1] rate	
	Slow rate		
	Stress errors	Equal stress/lexical stress errors	Syllable segregation
Inappropriate prosody	Syllable segregation	Syllable segregation	Equal stress/lexical stress errors
			Altered suprasegmental features
Inconsistent errors			Same word/syllable different on repetition
			Same C/V different across different words

[1] DDK: Diadochokinetic rate (syllables per unit time during the repetition of a syllable such as “pa”).

operational definitions of each one to facilitate study replication. Fedorenko et al. (2016) used a rubric including 21 signs, grouped into three categories corresponding to the ASHA consensus features. Shriberg et al. (2017a) proposed a list of ten signs, divided into the segmental domain (i.e., affecting phonemes) and the prosodic domain (encompassing phrasing, stress, and rate). Table 1 summarizes the signs employed in each of these studies, grouped into rough correspondences.

Similarly, no sign or set of signs has been validated as differentially diagnosing CAS from related conditions such as phonological disorder. Instead, many signs used to diagnose CAS are common to other disorders or may even occur as part of typical development. Consonant distortions, for example, characterize the majority of speech disorders almost by definition. Bowen (2015) lists six characteristics common to both CAS and phonological disorder, including omission of phonemes that are already in a child's repertoire, increased errors associated with increased utterance length or complexity, and atypical prosody. Phoneme distortions such as dentalized /s/ are also characteristic of the speech of young, typically developing children, occurring with >15 % frequency before age 7 (Bowen, 2015). In one study, typical 3-yearolds substituted [w] for word-initial /r/ in almost 60 % of responses (Rvachew & Brousseau-Lapré, 2012). Distortions such as these can also signal the presence of phonological disorder and are often present in the speech of children with CAS.

Further complicating the situation, signs of CAS can differ between children and within a child at different points in time (American Speech-Language-Hearing Association, 2007). For example, CAS is generally agreed to be a disorder that can range from mild to severe. The finding of McCabe, Rosenthal, and McLeod (1998) that CAS severity (as measured by percentage of consonants correct) negatively correlates with age (r = -0.58, p = 0.001) illustrates one way in which CAS manifests differently in younger and older children. Specifically, greater percent consonants correct was associated with older age, suggesting that severity may have diminished as children acquired better speech.

In short, while CAS is commonly viewed as a symptom complex rather than a unitary disorder, its signs can differ depending on severity, can vary within and between children, and can change over time; and there is as yet no agreement on the number of categories into which signs fall (Ozanne, 2005). Because signs of CAS differ according to the factors listed above, studying children who represent a variety of ages, severities, and comorbidities is necessary to generate the most comprehensive view of CAS.

1.2. Attempts to address diagnostic challenges

In spite, or perhaps because, of the issues discussed above, debate about the pathognomonic features of CAS is ongoing, and several groups have attempted to identify clusters of signs of CAS in order to classify them into theoretically- or clinically-based categories or to establish biomarkers that aid in CAS diagnosis.

To identify a specific speech characteristic that could serve as a single diagnostic marker of CAS, Ozanne (1995) performed a cluster analysis on 18 behaviors thought to indicate a disorder of speech motor planning, in a group of 100 children (ages 3;0–5;6) who were referred for speech treatment. She identified four clusters. The first cluster described children who experienced difficulty maintaining accurate production as words increased in length, made inconsistent errors, and made vowel errors. Children in the second cluster had slow production rate and difficulty sequencing sounds, as well as poor (nonspeech) oromotor ability. The third cluster included children who showed groping, consonant deletion, and difficulty imitating speech. Finally, the fourth cluster included children with prosodic disturbances and a history of no babbling. In a later rubric (Ozanne, 2005), children with CAS could show deficits in any of the first three clusters but must have had a deficit in the second or third in order to be diagnosed with CAS.

Still later, Peter (2006) examined a broad array of measures of speech, language, and manual clapping to distinguish subgroups of children with different kinds of speech sound disorders and typical development. Three clusters were found: one with low articulation/phonology scores but at least average language and clapping accuracy scores; a second with low articulation/phonology scores, average language scores, and intermediate clapping accuracy scores; and a third with low scores in all domains.

Murray, McCabe, Heard, and Ballard (2015) used a discriminant analysis to determine whether any combination of 24 variables predicted a diagnosis of CAS in a group of children with various speech disorders. Stress errors and syllable segregation (an inappropriate pause between syllables of a word) differentiated groups and accounted for 82 % of the variance in CAS diagnosis in their sample. A combination of stress matches, syllable segregation, percent phonemes correct, and accuracy on a diadochokinetic task differentiated children with CAS from children with phonological impairment and/or dysarthria with 91 % accuracy.

By contrast, Shriberg et al. (2017a) used a diagnostic checklist containing 10 signs of CAS, organized into two domains. The segmental domain included vowel distortions, voicing errors, distorted substitutions, difficulty with coarticulatory transitions, groping, intrusive schwa, and increased difficulty with multisyllabic words. The prosodic domain included syllable segregation, slow speech rate, and equal stress or lexical stress errors. Children had to meet criteria for at least four of the signs on at least three of 17 speech tasks from a detailed assessment protocol in order to receive a diagnosis of CAS.

These criteria were then used in Shriberg et al. (2017b) to diagnose a group of 41 speakers with idiopathic CAS and a group of 19 more speakers with CAS in the context of an identified neurogenetic disorder. Conversational speech samples from these speakers were then coded for the presence of between-word pauses at least 150 ms long that were inappropriate linguistically or associated with an inappropriate articulatory, prosodic, or vocalic feature. The presence of these four types of inappropriate pauses was then compiled to create a pause marker that showed 86.8 % sensitivity and 72.7 % specificity in classifying the 60 participants as having CAS or not, compared to clinical judgement. Specificity for just the 41 speakers with idiopathic CAS was 100 %.

Instead of identifying symptom clusters to diagnose CAS, other researchers have surveyed clinicians to understand what symptoms they use in differential diagnosis of CAS. Malmenholt, Lohmander, and McAllister (2017) surveyed 178 Swedish speech-language pathologists, asking them to select the signs of CAS from a list of 17 that they felt were commonly used. Signs selected more than 50 % of the time were inconsistent production (85 %), motor-programming deficits (82 %), sequencing difficulties (71 %), oro-motor deficits (63 %), vowel errors (62 %), voicing errors (61 %), consonant cluster deletion (54 %), and prosodic disturbance (53 %). This ranking closely corresponded to similar surveys of English-speaking speech pathologists (e.g., Forrest, 2003; Joffe & Pring, 2008).

1.3. Research question

In this paper, we expand on the above-mentioned work using an exploratory factor analysis (EFA). EFA is a statistical technique that seeks to reduce the dimensionality of a large set of measured variables by identifying underlying factors (latent or unobserved variables) that potentially account for correlation patterns between the measured variables (Watson, 2017). Determining the internal structure of a list of signs of CAS advances the state of our knowledge both by providing information about how well the collected signs represent the construct of CAS and by supporting the validity of the specific signs used to diagnose it. If observable signs of CAS group together into one or more common categories, it suggests that those signs arise from the same underlying process or deficit. This information can then be used to refine diagnostic criteria and guide the search for biomarkers of CAS.

Our specific research questions were: (a) how many factors does the best-fitting model include, and (b) what signs of CAS load most heavily onto those factors?

2. Methods

2.1. Participants

Data from 57 children with CAS were obtained from three different research centers: University of Nebraska, Miami University in Ohio, and Beth Israel-Deaconess Medical Center in Boston. These data are retrospective in nature and were not collected with the present study in mind. To ensure a representative sample of children with CAS, we included participants with different ages, severities, and comorbidities. Children fell into two main groups. There were 30 verbal children with CAS (ages 3;5−17;0), 20 of whom were recruited as part of a study at University of Nebraska and 10 as part of a study at Miami University. A total of 12 of the verbal children (10 from University of Nebraska, two from Miami University) carried community diagnoses of language disorder. Another 27 were minimally verbal children with ASD and CAS (ages 3;5–10;8), recruited as part of a study at Beth Israel-Deaconess Medical Center. While none of the minimally verbal children carried explicit diagnoses of language disorder, none used phrase speech despite being over three years of age.

All 57 children met criteria for CAS on the speech samples used for this analysis: they showed at least five unique signs of CAS from the list in Iuzzini-Seigel et al. (2015) during testing and were judged by speech pathologists experienced in pediatric motor speech disorders (first and second authors) to present with CAS. Because different stimuli and standardized tests were used for verbal and minimally verbal ASD children, the groups are described separately below and in Table 2.

Verbal Group: The 30 children with CAS in this group (3 female) were administered the Goldman-Fristoe Test of Articulation,

Table 2
Demographic information. Figures are listed as μ ± SD, [min-max]. Ages are years;months.

Verbal Participants (n = 30, 2 F)	
Age	8;7 ± 3;1 [3;5−17;0]
CELF[1] SS	81.7 ± 24.6 [44−133]
GFTA-2[2] SS	60.0 ± 15.2 [40−85]
GFTA-2 Pct.	2.9 ± 3.1 [1−15]
RIAS[3] SS (n = 10)	111.9 ± 9.0 [99−131]
Minimally Verbal Participants (n = 27, 3 F)	
Age	6;7 ± 1;5 [4;0−9;8]
MSEL[4] RL Raw (n = 13)	20.4 ± 8.2 [11−36]
MSEL EL Raw (n = 12)	11.2 ± 1.7 [8−14]
CARS[5] (n = 17)	38.4 ± 5.7 [30−47]
ADOS[6] (n = 15)	19.7 ± 3.3 [15−26]
MSEL VR Raw (n = 16)	29.6 ± 9.0 [19−48]

[1] CELF: Clinical Examination of Language Fundamentals. Standard score (SS) reported.
[2] GFTA: Goldman-Fristoe Test of Articulation. Standard score (SS) and percentile (pct.) reported.
[3] RIAS: Reynolds Intellectual Assessment Scales. Standard score (SS) reported.
[4] MSEL: Mullen Scales of Early Learning. Raw scores for Receptive Language (RL), Expressive Language (EL), and Visual Reception (VR) reported.
[5] CARS: Childhood Autism Rating Scale, 2nd Edition. Score ≥ 30 for autism.
[6] ADOS: Autism Diagnostic Observation Schedule. Score ≥ 12 for autism.

Second Edition (GFTA-2; Goldman & Fristoe, 2000) and the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). All scored below the 16th percentile on the GFTA for their age. Seventeen children in this group scored below 89 on the CELF (10 from University of Nebraska, seven from Miami University) and were considered to show comorbid language impairment as well as CAS. Nonverbal IQ for the 20 University of Nebraska participants was assessed using the Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003).

Minimally Verbal Group: This group was comprised of 27 minimally verbal children with CAS and ASD (3 female) who had participated in various treatment studies at Beth Israel-Deaconess Medical Center (Chenausky, Norton, Tager-Flusberg, & Schlaug, 2016; Chenausky, Norton, & Schlaug, 2017; Chenausky, Kernbach, Norton, & Schlaug, 2017; Wan et al., 2011). Diagnosis of ASD was confirmed by a score on the Childhood Autism Rating Scale (CARS; Schopler, Van Bourgondien, & Wellman, 2010) of at least 30 or a score on Module 1 of the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999) of at least 12. Minimally verbal status was defined by an expressive vocabulary of no more than 20 words, confirmed by parent report and by observation that the child produced fewer than 20 words during a 20-minute language sample. Language level was assessed for some children in this group using the receptive and expressive language subsections of the Mullen Scales of Early Learning (MSEL; Mullen, 1995), and nonverbal IQ using the Visual Reception subtest of the MSEL.

The studies from which these data come were approved by the Institutional Review Boards at the University of Nebraska, Miami University, and Beth Israel-Deacones Medical Center, respectively. Parents of all children gave written informed consent before their children were enrolled.

2.2. Stimuli

As mentioned above, separate stimulus sets were used for the verbal and minimally verbal groups because the groups differed on overall severity of involvement and on language skills. For the verbal children, stimuli consisted of children's responses on the GFTA-2 (53 words in total). For the minimally verbal children with ASD, stimuli were five bisyllabic words ("baby", "cookie", "daddy", "mommy", and "puppy"), repeated five times each during a baseline speech imitation task. All children in the MV cohort responded to all five prompts of each word, ensuring a consistent data set for this cohort.

2.3. Identification of signs of CAS

For both cohorts, video of children's performance was coded offline for signs of CAS. Operational definitions were taken from Iuzzini-Seigel et al. (2015). "Increased difficulty with multisyllabic words" was not used for either group since all of the stimuli for the minimally verbal children were bisyllabic. Two additional signs from Fedorenko et al. (2016) were used. "Additions" included insertion of phonemes or syllables (other than schwa) not in the target. "Variable errors" was coded once per target, when a child made a different error on one production of a word than on a previous production. For "variable errors" to be coded, a child needed to produce at least two tokens with different errors in them. Though the GFTA-2 does not require repetition of stimuli, children in the verbal cohort sometimes did repeat a stimulus more than once. The mean rate of repeated stimuli on the GFTA-2 was 6.7 % (SD 8.7 %, range 0-28%). "Variable errors" was scored for the verbal children when they repeated one of the GFTA-2 words and for the minimally verbal children on their repetitions of the stimulus words. Except for "variable errors", all signs were coded per attempt.

In Supplemental File 1, we include the definitions used by Iuzzini-Seigel et al. (2015), along with changes specific to the present study, such as definitions of the two additional signs. Further information on the Iuzzini-Seigel signs appears in Iuzzini-Seigel et al. (2015) and in a tutorial by Iuzzini-Seigel and Murray (2017).

Example: if a child produced [abʌndi] and [bʌndun] on two attempts for "baby", "variable errors" would be coded because the target was produced differently each time. Each attempt would also be coded "addition" for insertions of [n] and [a], as well as "consonant distortion" for the substitution of [d] for /b/. "Vowel error" would be coded for the incorrect vowels (once in the first token, twice in the second). Table 3 details mean raw scores for each sign, by group.

Table 3
Scores for each sign, by group. Figures are listed as μ ± SD.

Sign	Verbal	Minimally Verbal
Consonant Distortion*	46.4 ± 34.7	14.9 ± 5.4
Voicing Error	9.2 ± 5.8	11.5 ± 5.7
Nasality Error	1.9 ± 4.5	5.3 ± 8.2
Vowel Error*	10.5 ± 7.3	21.4 ± 4.5
Intrusive Schwa	1.3 ± 1.5	2.4 ± 2.9
Syllable Segmentation*	2.2 ± 3.4	8.2 ± 7.2
Stress Error	4.0 ± 1.6	4.4 ± 9.3
Slow Rate*	2.2 ± 4.8	7.8 ± 7.2
Difficulty with Coarticulation*	1.6 ± 1.9	4.3 ± 7.3
Groping*	0.3 ± 0.6	4.3 ± 4.1
Variable Errors	2.7 ± 2.8	2.6 ± 2.2
Additions*	0.7 ± 1.5	8.0 ± 5.7

* significantly different between groups, $p < 0.05$.

2.4. Reliability

Reliability was assessed separately for the verbal and minimally verbal groups using intra-class coefficients (ICCs). For the verbal group, 10 % of each GFTA-2 was independently coded for the purposes of assessing reliability. A two-way mixed ICC for single measures and consistency over all signs yielded a value of 0.901 (excellent), $p < 0.0005$. Mean ICC for individual signs was 0.856, all $p < 0.05$.

For the minimally verbal group, a consensus coding procedure on all 25 of each child's tokens was used. To assess reliability, the number of signs agreed on by the judges after consensus was compared to the total number of signs proposed by the judges before consensus, similar to the process for percent agreement (# agreements/total # of judgements). A two-way mixed ICC for single measures and consistency over all signs yielded a value of 0.952 (excellent), $p < 0.0005$.

2.5. Analytic strategy

The total number of times each sign appeared per child across their entire speech sample was tallied. To account for the use of different stimulus tests, raw scores for each sign were converted to z-scores, separately for each group. The z-scores for each sign were then entered as variables into the factor analysis. Following Watson (2017), we employed the following five steps in the EFA analysis: (1) evaluate the factorability of the data, (2) determine the number of factors to extract, (3) determine the number of factors to retain, (4) determine the appropriate factor rotation method, and (5) interpret the factor structure and name the factors.

Factorabliity was assessed using the Kaiser-Meyer-Olkin (KMO) test, which is a measure of the proportion of variance among the measured variables that is common variance; and Bartlett's test of sphericity, which estimates the degree to which the correlation matrix is an identity matrix. Factor analysis is appropriate for data sets where KMO > 0.6 and where Bartlett's test is significant ($p < 0.05$). Because we wished to compare models with different numbers of factors, we extracted the maximum number of factors, which is the same as the number of measured variables. Then, to determine how many factors to retain, we compared models with different numbers of factors using the Akaike Information Criterion (AIC; Akaike, 1987). The AIC, a model-selection statistic based on goodness of fit (log-likelihood), imposes a penalty for addition of parameters, selecting the most parsimonious model that fits the data well. Lower AIC values indicate a better model. Loadings for the best-fit model were rotated using an oblique rotation (direct oblimin with Kaiser normalization). Oblique rotation allows for some correlation between factors, which is appropriate for biological or behavioral systems.

A post-hoc interpolation analysis was performed to determine which factor loadings to interpret, consistent with de Winter, Dodou, and Wieringa (2009). Analyses were performed in Stata v14 (StataCorp, 2015) and SPSS Statistics 25 (IBM Corp., 2017).

3. Results

The KMO test yielded a value of 0.633 and Bartlett's test was statistically significant ($\chi^2(66) = 271.3$, $p < 0.0005$), indicating that there were correlations between the variables and thus that that the data set was factorizable.

Twelve models were created, with AIC values ranging from 102.5 (3-factor model) to 152.6 (1-factor model). The 3-factor model had the lowest AIC value and was thus the best fit to the data. It accounted for a total of 57.0 % of the variance in the observed variables (27.3 % for Factor 1, 17.5 % for Factor 2, and 12.2 % for Factor 3).

To select which loadings could be confidently recovered (which ones to interpret), we referred to the simulation study in de Winter et al. (2009). They showed that, for models with 12 variables and 3 factors, satisfactory factor recovery can be achieved for loadings greater than $|0.8|$ with an N of 17. An N of 67 is required for satisfactory recovery of factor loadings greater than $|0.6|$. Using interpolation, we determined that our N of 57 allowed confident recovery of factor loadings greater than $|0.64|$. Lower factor loadings were not used to interpret the factors. Table 4 displays the pattern matrix, in which cell entries represent regression coefficients

Table 4
Pattern matrix for the 3- factor model. Loadings greater than |0.64] are **bolded and underlined**.

	Factor		
	1	2	3
Syllable Segmentation	**.939**	−.068	−.058
Slow Rate	**.921**	−.057	.057
Stress Error	**.656**	.029	−.226
Vowel Error	.513	.339	.483
Nasality Error	.359	.268	.135
Groping	−.106	**.800**	−.190
Additions	.173	**.785**	−.117
Difficulty with Coarticulation	.066	**.697**	−.067
Intrusive Schwa	−.115	.570	.304
Variable Errors	−.192	.089	**.717**
Consonant Distortion	.010	−.316	.639
Voicing Error	.374	−.073	.474
Interpretation of Factor:	***Prosody***	***Coarticulation***	***Inconsistency***

relating the measured variables to the factors, and from which factors were interpreted. Table 5 displays the structure matrix, which shows the correlations between the measured variables and the factors.

The variables "syllable segmentation" (0.939), "slow rate" (0.921), and "stress error" (0.656) had the highest rotated loadings on Factor 1. "Groping" (0.800), "additions" (0.785), and "difficulty with coarticulation" (0.697) had the highest loadings on Factor 2. "Variable errors" (0.717) had the highest loading on Factor 3. Correlations (Pearson's *r*) between Factors 1 and 2 was 0.163, between Factors 1 and 3 was 0.105, and between Factors 2 and 3 was 0.018. These low values indicate relative independence of the factors.

4. Discussion

The aim of this study was to identify latent factors underlying observable signs of CAS in a heterogeneous group of children meeting criteria for the disorder. Here, we interpret the factors, discuss the findings in the context of potential circularity in the results, and relate the findings to previous research. We then address issues concerning stimulus selection and how they may have affected the results. Finally, we discuss limitations, future work, and clinical implications.

A three-factor model best fit the data. We interpreted the factors to be related to prosody (Factor 1), coarticulation (Factor 2), and inconsistency (Factor 3). These factors are highly similar to the three American Speech-Language-Hearing Association (2007) consensus criteria of inappropriate prosody, lengthened and disrupted coarticulatory transitions, and variable errors on phonemes in repeated productions. Our results thus support the use of the three ASHA (2007) consensus criteria both clinically and in studies like that of Fedorenko et al. (2016), in the sense of ensuring that children suspected to have CAS show signs in each of the three areas delineated by the ASHA consensus criteria in addition to showing a minimum number of signs from a particular list. Our results do not support a division of signs of CAS into just two factors, such as segmental and prosodic categories (as in Shriberg et al., 2017a), because the two-factor model was not the best fit to the data and because signs directly relating to segmental errors (e.g., consonant distortions, voicing errors, nasality errors, and vowel errors) did not load heavily onto the second factor in the two-factor model.

Note that the current results may give the impression of circular reasoning if it is assumed that the same signs used to verify diagnosis of CAS were also employed in the factor analysis itself, leading to the conclusion that participants were selected to fit certain criteria and were then found to show those same criteria. This is not the case, however: Participants were selected to show Iuzzini-Seigel's signs (minus "increased difficulty with multisyllabic words", plus "additions" and "variable errors"), not the three ASHA consensus criteria. Furthermore, the current study did not assume that Iuzzini-Seigel's signs conformed to the ASHA consensus criteria. Instead, it explored whether that might be the case. In principle, the factor analysis could have shown that the signs loaded onto any number of underlying factors. There was no guarantee that the best model would contain three factors, nor that those factors would load onto sign groups consistent with the ASHA consensus criteria. Whichever way the signs did group would have been a useful and informative result. The actual results support a set of more abstract criteria (the ASHA consensus criteria) that were arrived at via a process separate from that which gave rise to the observable signs described in Iuzzini-Seigel et al. (2015) and elsewhere. An example from the field of autism may also help clarify. A common diagnostic instrument for autism is the Autism Diagnostic Interview-Revised (ADI-R). It employs three scales of items corresponding to the three core symptom domains of autism that were laid out in the DSM-IV-TR (APA, 2000): deficits in social interaction, deficits in communication, and the presence of stereotyped behaviors. However, a factor analysis of the ADI-R by Frazier, Youngstrom, Kubu, Sinclair, and Rezai (2008) showed that the ADI-R items were best represented by a two-factor structure (deficits in social communication and the presence of stereotyped behaviors). In this case, the ADI-R items were derived from a set of three symptom domains, but factor analysis revealed an underlying set of just two domains. Our factor analysis could have found an underlying structure corresponding to a prosodic and a segmental factor, for example; but that was not the case.

Despite not being consistent with associating signs of CAS with only a prosodic and a segmental factor, as implied in Shriberg et al. (2017a), our findings do strongly support employing a pause-related measure as a potential biomarker for CAS, similar to the pause marker of Shriberg et al. (2017a, 2017b). First, the factor accounting for the largest amount of variance in the variables was associated with prosody. Second, and more importantly, the variable with the highest loading on the prosody factor was syllable segmentation

Table 5
Structure matrix for the 3- factor model.

	Factor		
	1	2	3
Syllable Segmentation	.922	.084	.039
Slow Rate	.917	.094	.152
Stress Error	.637	.132	−.157
Vowel Error	.619	.432	.543
Nasality Error	.417	.329	.177
Groping	.004	.779	−.187
Additions	.288	.811	−.085
Difficulty with Coarticulation	.172	.707	−.048
Intrusive Schwa	.009	.556	.302
Variable Errors	−.102	.070	.698
Consonant Distortion	.025	−.303	.634
Voicing Error	.411	−.003	.512

(inappropriate within-word pausing). Though Shriberg et al. (2017b) describe their pause marker as being independent of intellectual ability and language status, it is a measure of between-word pausing from a continuous speech sample that was collected only from verbal participants. Because minimally verbal children by definition do not use phrase speech and cannot produce a continuous speech sample, a between-word pause marker cannot be used with this group. However, inappropriate pauses between syllables of bisyllabic words, or even between phonemes within a syllable, may be a useful biomarker for more severe forms of CAS. Of note, Murray et al. (2015) also found that syllable segregation significantly discriminated between participants with and without CAS, explaining 79 % of the variability in their discriminant analysis, despite the fact that this sign occurred on a minority of opportunities (approximately 30 % of opportunities in their CAS-only group and approximately 8.5 % of opportunities in their CAS-plus group). Thus, syllable segregation and similar pauses may not occur very frequently but may be highly indicative of CAS.

Another potential concern with the current findings is that they may have been driven by the use of different test stimuli for the verbal and minimally verbal groups. The same stimuli cannot be used with both groups because words easy enough for minimally verbal participants would likely not sufficiently challenge verbal participants enough to reveal signs of CAS, but words that would appropriately challenge verbal participants would be too hard for minimally verbal speakers. It is possible that the use of different stimuli may have affected the results by imposing different limits on the number of times a particular sign could have appeared in a child's speech sample. That is, one might posit that the GFTA-2 would provide more opportunities for showing consonant distortions, for example, by virtue of including words like "feather" and "jumping" that contain more consonants (and more late-appearing ones) than words such as "mommy" or "baby" from the minimally verbal stimulus set.

Two factors mitigate this concern, however. First, "consonant distortions" did not (quite) meet the cutoff for interpretability (loading = 0.639, less than 0.64). Therefore, even though consonant distortions were more common in the verbal children compared to the minimally verbal children (mean = 46.4 per verbal child vs. 14.9 per minimally verbal child), that did not cause this sign to load heavily on any factor. Second, the way that "variable errors" was coded was different from how other signs were coded and had a smaller maximum value than the other signs, yet it still loaded heavily onto Factor 3. As mentioned in the Methods section, while other signs of CAS could be rated as many times as they occurred in each of a child's tokens, "variable errors" was only rated once per target, when at least two different tokens were produced with different errors. Thus, the total number of times that this sign could appear in a child's speech sample was much lower than for a sign like consonant distortion. Still, "variable errors" met the criterion for interpretability, while "consonant distortion" did not. We ascribe this to the fact that it was the correlation between variables, not the score ranges, that determined the factor structure.

4.1. *Limitations and future work*

As with most studies involving relatively rare disorders, the small number of participants and stimuli are limitations. The present results should be replicated in a larger independent sample that is representative of age and the range of severity of CAS. In addition, children in that sample should also be assessed for other speech-related comorbidities common to CAS, such as childhood dysarthria and phonological disorder, in order to fully document their speech performance.

The nature of the convenience sample is also a limitation, in that including children with other comorbidities (ASD, language impairment) will have introduced additional variability into the sample. As mentioned earlier, however, we felt that including children with CAS plus would be more representative of how CAS appears clinically, given recent work identifying high rates of CAS comorbidity in children with genetic and neurodevelopmental disorders (Shriberg et al., 2011; Fedorenko et al., 2016; Morgan et al., 2016; Chenausky et al., 2019).

Our reliance on different stimulus sets is also a limitation. Future work could include stimuli that are common across groups to some extent – for example, by making the stimulus set for the verbal participants a superset of that for the minimally verbal participants. Equal numbers of opportunities per target to show the signs assessed (e.g., "variable errors", by having children repeat each stimulus three times) might also be important, as might a larger stimulus set.

Regarding the use of pausing as a biomarker for CAS across the range of severity, syllable segmentation and other types of intra-word pauses should be explored as a measure that may distinguish minimally verbal children with CAS from those without. This could be accomplished using simple bisyllabic words like the ones employed here. Similarly, a diadochokinetic task like that described in Thoonen, Maassen, Gabreëls, and Schreuder (1999) could be explored as a potential differentiator.

Finally, because we investigated what a heterogeneous group of children with CAS have in common, not what may distinguish them from children without CAS, an exploratory factor analysis like the one reported on here should be performed with the same signs, but on children with non-CAS disorders such as phonological disorder or ataxic dysarthria, as well as on children with typical speech. This would be an important study because children with phonological disorder or ataxia are expected to show many of the same signs from the list employed here, but those signs would not be expected to load on the same number or kind of underlying factors as in CAS.

4.2. *Clinical implications*

The current results support the use of the three ASHA consensus criteria, in addition to other published lists of features, to aid diagnosis of CAS. Future revisions of the diagnostic criteria might include the specification that, along with showing a certain criterion number of signs of CAS across a minimum number of speech tasks, a child should also show at least one sign in each of the three ASHA domains of inappropriate prosody, lengthened/disrupted coarticulatory transitions, and inconsistent errors on phonemes in repeated productions.

Funding sources

This work was supported by P50 DC 13027 to HTF (supporting HTF, GS, ACN, and KVC), P50 DC 18006 to HTF (supporting HTF, JRG, and KVC), the Nancy Lurie Marks Family Foundation and Autism Speaks (supporting GS), T32 DC 013017 to C. A. Moore (supporting KVC), K24 DC 016312 to JRG, the American Speech-Language-Hearing Foundation (supporting KVC), K99 DC 017490 (to KVC), a NHMRC Practitioner Fellowship #1105008 (to AM), a NHMRC Centre of Research Excellence in Speech and Language Neurobiology #1116976 (to AM), and NHMRC Project grant #1127144 (to AM).

CRediT authorship contribution statement

Karen V. Chenausky: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft, Funding acquisition. **Amanda Brignell:** Writing - review & editing, Investigation, Validation. **Angela Morgan:** Writing - review & editing, Investigation, Funding acquisition, Validation. **Danielle Gagné:** Validation. **Andrea Norton:** Writing - review & editing, Data curation, Investigation. **Helen Tager-Flusberg:** Writing - review & editing, Resources, Supervision, Funding acquisition. **Gottfried Schlaug:** Writing - review & editing, Resources, Supervision, Funding acquisition. **Aaron Shield:** Writing - review & editing, Resources. **Jordan R. Green:** Writing - review & editing, Resources, Supervision.

Acknowledgements

We thank the children who participated in the studies leading to this manuscript, and their families for allowing them to participate.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jcomdis.2020.106033.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317–332.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. APA. https://doi.org/10.1176/appi.books.9780890423349 (4th ed., text rev.).

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

Bowen, C. (2015). *Children's speech sound disorders*. Oxford, UK: Wiley Blackwell.

Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2016). Auditory Motor Mapping Training: Comparing the effects of a novel speech treatment to a control treatment for minimally verbal children with autism. *PLoS One, 11*(11), e0164930.

Chenausky, K., Norton, A., & Schlaug, G. (2017). Auditory-motor mapping training in a more verbal child with autism. *Frontiers in Human Neuroscience, 11*. https://doi.org/10.3389/fnhum.2017.00426. Article 426.

Chenausky, K., Kernbach, J., Norton, A., & Schlaug, G. (2017). White matter integrity and treatment-based change in speech performance in minimally verbal children with autism spectrum disorder. *Frontiers in Human Neuroscience, 11*. https://doi.org/10.3389/fnhum.2017.00175.

Chenausky, K., Brignell, A., Morgan, A., & Tager-Flusberg, H. (2019). Motor speech impairment predicts expressive language in minimally verbal, but not low verbal, individuals with autism spectrum disorder. *Autism and Developmental Language Impairment, 4*, 1–12.

de Winter, J., Dodou, D., & Wieringa, P. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44*(2), 147–181.

Fedorenko, E., Morgan, A., Murray, E., Cardinaux, A., Mei, C., Tager-Flusberg, H., et al. (2016). A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2. *European Journal of Human Genetics, 24*(2), 302–306.

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

Frazier, T., Youngstrom, E., Kubu, C., Sinclair, L., & Rezai, A. (2008). Exploratory and confirmatory factor analysis of the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 38*, 474–480.

Goldman, R., & Fristoe, M. (2000). *Goldman–Fristoe test of articulation* (2nd ed.). Circle Pines, MN: Pearson.

IBM Corp. (2017). *IBM SPSS statistics for windows, version 25.0*. Armonk, NY: IBM Corp.

Iuzzini-Seigel, J., & Murray, E. (2017). Speech assessment in children with childhood apraxia of speech. *Perspectives of the ASHA Special Interest Groups, 2*(August), 47–60.

Iuzzini-Seigel, J., Hogan, T., Guarino, A., et al. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54*, 32–42.

Joffe, V., & Pring, T. (2008). Children with phonological problems: A survey of clinical practice. *International Journal of Language and Communication Disorders, 43*(2), 154–164.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism diagnostic observation schedule*. Los Angeles, CA: Western Psychological Services.

Malmenholt, A., Lohmander, A., & McAllister, A. (2017). Childhood apraxia of speech: A survey of praxis and typical speech characteristics. *Logopedics, Phoniatrics, Vocology, 42*(2), 84–92.

McCabe, P., Rosenthal, J., & McLeod, S. (1998). Features of developmental dyspraxia in the general speech-impaired population? *Clinical Linguistics and Phonetics, 12*(2), 105–126.

Morgan, A., Fisher, S., Scheffer, I., Hildebrand, M., et al. (2016). FOXP2-related speech and language disorders. In M. Adam (Ed.), *GeneReviews [internet]*. Seattle, WA: University of Washington, Seattle.

Mullen, E. M. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service.

Murray, E., McCabe, P., Heard, R., & Ballard, K. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech Language and Hearing Research, 58*, 43–60.

Ozanne, Anne (1995). The search for developmental verbal dyspraxia. In Barbara Dodd (Ed.), *The differential diagnosis and treatment of children with speech disorder* (pp. 91–109). Wiley.

Ozanne, Anne (2005). Childhood apraxia of speech. In Barbara Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (2nd ed., pp. 71–82). London, UK: Whurr Publishers.

Peter, B. (2006). *Multivariate characteristics and data-based disorder classification in children with speech disorders of unknown origin. Doctoral dissertation*. University of Washington. ProQuest Dissertations and Theses Global.

Reynolds, C., & Kamphaus, R. (2003). *Reynolds intellectual assessment scales*. Lutz, FL: Psychological Assessment Resources.

Rvachew, S., & Brousseau-Lapré. (2012). *Developmental phonological disorders: Foundations of clinical practice*. San Diego, CA: Plural Publishing.

Schopler, E., Van Bourgondien, M., Wellman, J., et al. (2010). *Childhood autism rating scale-second edition (CARS2)*. Los Angeles, CA: Western Psychological Services.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamental* (4th ed.). San Antonio, TX: Pearson.

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

Shriberg, L., Potter, N., & Strand, E. (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech Language and Hearing Research, 54*, 487–519.

Shriberg, L., Strand, E., Fourakis, M., Jakielski, K., Hall, S., Karlsson, H., et al. (2017a). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker. *Journal of Speech, Language, and Hearing Sciences. Supplement: A Diagnostic Marker to Discriminate Childhood Apraxia of Speech from Speech Delay*, 1096–1117.

Shriberg, L., Strand, E., Fourakis, M., Jakielski, K., Hall, S., Karlsson, H., et al. (2017b). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: II. Validity studies of the pause marker. *Journal of Speech, Language, and Hearing Research. Supplement: A Diagnostic Marker to Discriminate Childhood Apraxia of Speech from Speech Delay*, 1118–1134.

StataCorp. (2015). *Stata statistical software: Release 14*. College Station, TX: StataCorp LP.

Strand, E., McCauley, R., Weigand, S., et al. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech Language and Hearing Research, 56*, 505–520.

Thoonen, G., Maassen, B., Gabreëls, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics and Phonetics, 13*(1), 1–23.

Wan, C., Bazen, L., Baars, R., Libenson, A., Zipse, L., et al. (2011). Auditory-Motor Mapping Training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLoS One*, e2550. https://doi.org/10.1371/journal.pone.0025505.

Watson, J. (2017). Establishing evidence for internal structure using exploratory factor analysis. *Measurement and Evaluation in Counseling and Development, 50*(4), 232–238.