AJSLP

Research Article

A Modeling-Guided Case Study of Disordered Speech in Minimally Verbal Children With Autism Spectrum Disorder

Karen V. Chenausky,[a,b] Amanda Brignell,[c,d] Angela T. Morgan,[c] Andrea C. Norton,[e] Helen B. Tager-Flusberg,[f] Gottfried Schlaug,[b,e,g] and Frank H. Guenther[h]

Purpose: Understanding what limits speech development in minimally verbal (MV) children with autism spectrum disorder (ASD) is important for providing highly effective targeted therapies. This preliminary investigation explores the extent to which developmental speech deficits predicted by Directions Into Velocities of Articulators (DIVA), a computational model of speech production, exemplify real phenotypes.
Method: Implementing a motor speech disorder in DIVA predicted that speech would become highly variable within and between tokens, while implementing a motor speech plus an auditory processing disorder predicted that DIVA's speech would become highly centralized (schwa-like). Acoustic analyses of DIVA's output predicted that acoustically measured phoneme distortion would be similar between the two cases, but that in the former case, speech would show more within- and between-token variability than in the latter case. We tested these predictions quantitatively on the speech of children with MV ASD. In Study 1, we tested the qualitative predictions using perceptual analysis methods. Speech pathologists blinded to the purpose of the study tallied the signs of childhood apraxia of speech that appeared in the speech of 38 MV children with ASD. K-means clustering was used to create two clusters from the group of 38, and analysis of variance was used to determine whether the clusters differed according to perceptual features corresponding to within- and between-token variability. In Study 2, we employed acoustic analyses on the speech of the child from each cluster who produced the largest number of analyzable tokens to test the predictions of differences in within-token variability, between-token variability, and vowel space area.
Results: Clusters produced by k-means analysis differed by perceptual features that corresponded to within-token variability. Nonsignificant differences between clusters were found for features corresponding to between-token variability. Subsequent acoustic analyses of the selected cases revealed that the speech of the child from the high-variability cluster showed significantly more quantitative within- and between-token variability than the speech of the child from the low-variability cluster. The vowel space of the child from the low-variability cluster was more centralized than that of typical children and that of the child from the high-variability cluster.
Conclusions: Results provide preliminary evidence that subphenotypes of children with MV ASD may exist, characterized by (a) comorbid motor speech disorder and (b) comorbid motor speech plus auditory processing disorder. The results motivate testable predictions about how these comorbidities affect speech.
Supplemental Material: https://doi.org/10.23641/asha.14384432

[a]MGH Institute of Health Professions, Boston, MA
[b]Harvard Medical School, Boston, MA
[c]Speech and Language, Murdoch Children's Research Institute, Melbourne, Victoria, Australia
[d]University of Melbourne, Victoria, Australia
[e]Music, Neuroimaging, and Stroke Recovery Laboratory, Beth Israel Deaconess Medical Center, Boston, MA
[f]Department of Brain and Psychological Sciences, Boston University, MA
[g]Brain Repair and Neurorestoration Center, Baystate Medical Center, Springfield, MA
[h]Sargent College, Boston University, MA

Correspondence to Karen V. Chenausky: kvchenausky@mghihp.edu

Editor-in-Chief: Jessica E. Huber
Editor: Caroline Niziolek

Received May 11, 2020
Revision received September 14, 2020
Accepted January 8, 2021
https://doi.org/10.1044/2021_AJSLP-20-00121
Publisher Note: This article is part of the Special Issue: Select Papers From the 2020 Conference on Motor Speech.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by impaired social communication and the presence of repetitive behaviors and restricted interests (American Psychiatric Association, 2013). Language ability in ASD ranges from above average to completely absent, but we do not yet understand why up to 30% of children with ASD fail to develop spoken language by age 5 years (Tager-Flusberg & Kasari, 2013). These minimally verbal children use only a very small number of single words or fixed phrases (fewer than 20 or 30) to communicate, and do so for a very limited number of functions, such as requesting favorite objects from familiar adults or protesting nonpreferred activities (Kasari et al., 2013).

Communication is both a basic need and the right of all human beings (Brady et al., 2016), and better expressive communication is associated with fewer maladaptive behaviors in ASD (Baghdadli et al., 2003; Dominick et al., 2007; Hartley et al., 2008). Developing effective communication treatments for minimally verbal children is therefore an area of high clinical importance. Yet, in order to do this, we must first understand what limits their spoken language development. In this model-driven case study, we sought to understand some of the factors that may impede the ability of minimally verbal children with ASD to develop intelligible speech.

Comorbidities in ASD and Their Relation to Spoken Language Development

While ASD is defined in terms of the two domains mentioned above, it is in fact a highly heterogeneous disorder in which comorbidities are common. In a sample of 2,568 eight-year-old children with ASD, Levy et al. (2010) documented that 82.7% had at least one additional developmental diagnosis. These (non–mutually exclusive) diagnoses included language impairment (63.4%), attention-deficit/hyperactivity disorder (21.3%), intellectual disability (18.3%), sensory integration disorder (15.7%), and specific learning disorders (6.3%). Problems and delays in both gross and fine motor development are also highly prevalent in ASD, affecting some 26% of children with ASD (Lloyd et al., 2011; Soke et al., 2018).

Sensory processing differences have also been found in ASD; for recent reviews, see Marco et al. (2011) and Robertson and Baron-Cohen (2017). More specifically to the current study, differences in the categorical perception of phonemes have been identified in children and adults with ASD. You et al. (2017) found that children with ASD showed less categorical precision in vowel and consonant identification than age-matched peers, due to better within-category perception of phonetic differences. In a similar study, Stewart et al. (2018) found adults with ASD to have both better within-category and poorer between-category discrimination of a /g /-/k / continuum than typical adults matched on nonverbal IQ, indicating less categorical phoneme perception. A measure of the level of categorical perception was significantly correlated with scores on tests of reading, spelling, and nonword judgment in these participants.

Motor delays are common in ASD and have also been shown to relate to language ability. Mody et al. (2017) examined 1,781 children with ASD, ranging in age from 2 to 17 years, and found a significant positive association between fine motor skills and both expressive and receptive language. Likewise, Choi et al. (2018) showed that infants who later developed ASD experienced slower growth in fine motor skills in the first 2 years of life than those who did not develop ASD, and that fine motor skills at 6 months of age significantly predicted expressive language at 36 months.

Comorbidities in Minimally Verbal ASD and Their Relation to Spoken Language Development

Minimally verbal children with ASD have been much less thoroughly investigated than verbal children, adolescents, and adults with ASD because of the challenges that minimally verbal individuals experience in participating in standardized testing and the limited utility of most standardized tests for describing their skills. For example, though auditory perceptual abilities are generally understood to influence language development, testing for auditory processing disorder is not part of the standard of care for minimally verbal children with ASD. Most behavioral tests of auditory processing are not considered reliable in children with mental ages below 7 years because of the challenging nature of auditory discrimination tasks (American Speech-Language-Hearing Association [ASHA], 2005).

In terms of motor speech skills, however, a persistent clinical hypothesis has been that comorbid childhood apraxia of speech (CAS) may limit spoken language development in at least some minimally verbal children with ASD (Prizant, 1996). CAS, which involves difficulty in planning or sequencing speech movements, results in effortful speech that is imprecise, inconsistent, and often unintelligible (ASHA, 2007). Shriberg et al. (2011) proposed that CAS may be "a sufficient cause of lack of speech development in at least some children classified as nonverbal ASD" (p. 405), while interpreting differences in speech, prosody, and voice production in verbal children with ASD as not generally being characteristic of CAS.

There is evidence that some low- and minimally verbal children with ASD have speech motor and nonspeech oromotor production difficulties, though CAS has not always been identified (Adams, 1998; Rapin et al., 2009; Wolk & Giesen, 2000). Gernsbacher et al. (2008) reported on a group of 111 children with ASD aged from 28 to 227 months that was divided into three groups that they termed *highly fluent*, *moderately fluent*, and *minimally fluent*—roughly equivalent to "verbal," "moderately verbal," and "minimally verbal." Early oromotor skills (e.g., sticking one's tongue out on request or blowing bubbles with a straw) differentiated groups, with the minimally fluent group showing the most severely impaired oromotor skills. More recently, Belmonte et al. (2013) reported on a group of

31 participants with ASD, ranging in age from 22 to 65 months. Eleven of these participants showed significantly lower expressive than receptive language scores, and this discrepancy was associated with the presence of oromotor impairments. Again, the children with the greatest oromotor impairments were also the ones with the lowest expressive language scores.

Chenausky et al. (2019) explicitly examined the speech of 54 low- or minimally verbal children with ASD for signs of CAS. Their participants fell into four groups: a group of 12 whose speech was within normal limits for their age, a group of 16 whose speech was disordered but was not judged to be consistent with CAS, a group of 13 whose speech met criteria for and was consistent with CAS, and a final group of 13 who did not produce enough speech to be able to rate. Scores on the Kaufman Speech Praxis Test (Kaufman, 1995) predicted the number of different words children produced during a language sample for the CAS and insufficient-speech groups. These findings suggest that comorbid CAS may affect spoken language production in a subgroup of minimally verbal children with ASD.

Using Computational Models to Understand Developmental Speech Disorders

A powerful method for investigating speech is to employ predictions from computational models of speech production such as DIVA (Directions Into Velocities of Articulators; Guenther, 1994; Guenther et al., 2006). DIVA is a biologically plausible neural network that models the feedforward and feedback control loops involved in both developing and mature speech production and thus may provide an explanatory account of neurodevelopmental disorders affecting speech.

DIVA's developmental model includes three stages. In the first stage, which approximates infant babbling, semirandom motor commands enable the model to learn the mapping between oral movements and their acoustic and somatosensory consequences. The model also acquires the mappings that translate sensory errors into corrective motor commands. These mappings are the core of DIVA's auditory and somatosensory feedback control systems. The second stage corresponds to imitation learning, during which DIVA is provided with auditory targets containing phonemes from its "native" language. DIVA's initial attempts to produce the phonemes rely heavily on auditory feedback control. With each attempt, however, the motor commands are updated and stored in synaptic weights that encode feedforward "motor programs" for the phonemes, improving production on the next try. After several attempts, DIVA can produce phonemes with minimal contribution from the feedback system because the learned feedforward commands are sufficient for error-free production. In this third, mature stage, DIVA relies on the feedforward commands it has refined in order to consistently produce correct utterances.

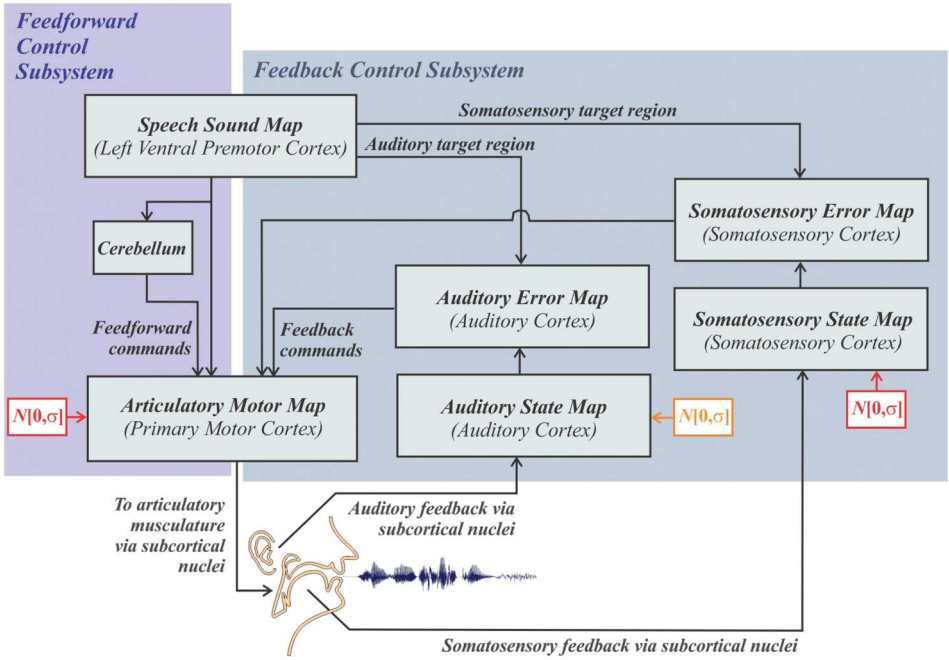
In addition to modeling typical speech development, DIVA has also been used to model disordered developing speech. Terband et al. (2014) used the DIVA model[1] to investigate possible origins of developmental speech deficits in children and, in particular, to investigate the question of whether there may be two distinguishable subgroups of children with speech deficits: those whose deficit is limited to the sensorimotor system and those who have an additional deficit in auditory processing. This was studied by "impairing" different DIVA model components during the developmental phases (babbling and imitation) and comparing the resulting model behavior to that of children with developmental speech disorders. Two versions of the model were implemented (see Figure 1). In the first, impaired neural processing was limited to the motor system (termed a *motor processing disorder*, or MPD, by Terband et al., 2014). This was implemented by applying stationary additive white Gaussian noise to DIVA's articulatory motor map (typically denoted by the variable M in the DIVA equations; e.g., Guenther et al., 2006) and somatosensory state map (denoted by S). For example, to calculate the overall motor command to the articulators at time t, denoted by $M(t)$, the following equation was used in the articulatory motor map:

$$M(t) = M(0) + \propto_{ff} \int_0^t \dot{M}_{ff}(t) + \propto_{fb} \int_0^t \dot{M}_{fb}(t) + N[0, \sigma](t), \quad (1)$$

where $M(0)$ is the starting configuration of the vocal tract; α_{ff} and α_{fb} are feedforward and feedback control gain parameters, respectively; $\dot{M}_{ff}(t)$ is the articulator velocity command generated by the feedforward controller; $\dot{M}_{fb}(t)$ is the articulator velocity command generated by the feedback controller; and the noise term $N[0, \sigma](t)$ is a random number chosen at each time point from a zero-mean unit normal distribution with standard deviation σ. Noise samples were uncorrelated across time points. This noise term approximates the aggregated effects of impaired neurophysiological function on neural activity in primary motor cortex, which is the key cortical region for generating articulator movements for speech, and can represent abnormalities from many sources, such as synaptic dynamics, neurotransmitter levels, receptor binding, axonal structure, and so forth. A similar noise term was also added to the equation for the somatosensory state map[2] to represent impaired function in somatosensory cortex, which is heavily interconnected with and functionally linked to primary motor cortex. To model varying degrees of impairment, the authors systematically varied the standard deviation of the noise sources across simulations from 5% (representing mild impairment) to 25% (representing severe impairment) of the range of activations of the model cells (which were normalized to vary between 0 and 1).

[1]The DIVA model equations utilized by Terband et al. (2014) are the same as those in the most recent version of the DIVA software, available at http://sites.bu.edu/guentherlab/software/diva-source-code/.
[2]See Guenther et al. (2006) for a complete treatment of the DIVA model equations.

Figure 1. Simplified block diagram of the Directions Into Velocities of Articulators model with the addition of zero-mean Gaussian noise sources (indicated by $N[0, \sigma]$, where σ indicates standard deviation) to simulate two hypothetical subtypes of children with speech disorders: those with a motor processing deficit limited to sensorimotor cortex (red) and those who additionally have a processing deficit in auditory cortex (orange). Adapted from Terband et al. (2014), Copyright © 2014, with permission from Elsevier.



In the second version of the model simulated by Terband et al. (2014), impairment was not limited to just motor and sensorimotor cortex but also involved an auditory processing disorder (MPD + APD). In this case, noise was added to the model's auditory state map (A) as well as to M and S to represent a subset of children who have impaired auditory processing in addition to impaired motor processing. Again, different levels of impairment were implemented in different simulations by varying the standard deviation of the noise across simulations from 5% to 25% of the range of activation of the model cells.

The authors then acoustically analyzed DIVA's speech in the mature state after each disorder profile had been implemented during development at each of the different noise levels. Noise deteriorated the model's speech in both the MPD and MPD + APD conditions, but in different ways. In the MPD condition (noise only in the motor and somatosensory state representations), the model in effect did not know the position of its articulators, so the phonemes it produced were distorted. However, because the auditory state map was intact, the model was able to "self-monitor" and use auditory feedback to home in on the correct target. This resulted in "searching articulation," which refers to articulator movement during phonation in an attempt to achieve the correct position for the desired target; it leads to a high degree of within-token variability. A related phenomenon, "groping" (searching articulator movement without phonation), was not implemented in Terband et al. (2014). Finally, because in the MPD-only condition repeated attempts at the same target came out differently each time, there was also a high degree of token-to-token (between-token) variability.

In the MPD + APD condition, noise in the auditory state map (as well as in the motor and somatosensory maps) meant that distinctions between phonemes were blurred, making them sound more similar to each other. Reduced or no auditory differences between target and output meant that searching articulation was not triggered. The model was in effect unable to use auditory feedback to detect acoustic errors; therefore, in this case, there were lower levels of within-token variability. Phoneme distortion levels were high because all phonemes, regardless of target, were produced as schwa at the highest noise levels, but between-token variability was low.

These qualitative differences between the two cases were reflected in acoustic differences. An acoustic index of phoneme distortion was similar between the MPD-only and MPD + APD conditions, but indices of searching articulation (within-token variability) and token-to-token (between-token) variability were larger in the MPD-only case than in the MPD + APD case.

Testing DIVA Predictions

The results of Terband et al. (2014) suggest that there may be groups of children whose speech resembles that of DIVA with an MPD alone and that of DIVA with an MPD + APD. Specifically, one prediction is that the group

corresponding to DIVA with an MPD alone will show higher levels of within- and between-token variability than the group corresponding to DIVA with an MPD + APD, and that these forms of variability should be evident both perceptually and acoustically. A second prediction is that the two groups should show similar levels of phoneme distortion, but that it will arise from different sources. In the case of the high-variability group, phoneme distortion will be a consequence of within- and between-token variability. In the case of the low-variability group, phoneme distortion will be a consequence of the targets being highly centralized and, in the extreme case, being all realized as schwa. Relatedly, the vowel space area of children in the low-variability group will be smaller than that of children in the high-variability group. We report here on the results of two studies that quantitatively test these perceptual and acoustic predictions in minimally verbal children with ASD.

Overview and Hypotheses

Because the predictions from Terband et al. (2014) are expressed in both perceptual and acoustic terms, our investigation employed a two-pronged approach. In Study 1, we tested the predictions perceptually. Trained listeners identified signs of CAS in the speech of a group of minimally verbal children with ASD. We hypothesized that, if we divided participants into two subgroups, one would show higher scores for perceptual signs related to within- and between-token variability than the other, but that the subgroups would show similar scores for perceptual signs related to phoneme distortion.

Study 2 took the form of a case study in which we performed acoustic analyses on the speech of a child from each subgroup. We hypothesized that the speech of the child from the high-variability subgroup would show higher values for acoustic indices of searching articulation and token-to-token variability (within- and between-token variability, respectively) than the speech of the child from the low-variability subgroup. Children's speech would show similar values for an acoustic index of phoneme distortion. Finally, the child from the low-variability subgroup would show a smaller vowel space area than the first child. Details justifying the case study approach are given in the Methods section for Study 2 below.

The perceptual and acoustic analyses used in this work complement each other. Auditory-perceptual evidence is ecologically valid and clinically relevant, but there are well-known limitations to perceptual analyses (Kent, 1996). Acoustic evidence, on the other hand, is objective but can reflect imperceptible differences and is often difficult to relate to how speakers actually sound. By combining both types of evidence, we hoped not only to use each type of analysis to its best advantage but also, crucially, to link the two. If perceptual and acoustic evidence support each other, this gives us confidence in our findings. It also provides evidence of the magnitude of the effect, that is, that any differences are both present acoustically and sufficiently large to be reliably perceived.

Study 1

Participants: Study 1

Participants for Study 1 were a convenience sample of 38 minimally verbal children with ASD (five female) who had been recruited for a series of studies testing the efficacy of an intonation-based therapy (Chenausky, Kernbach, et al., 2017; Chenausky et al., 2016, 2018; Chenausky, Norton, & Schlaug, 2017; Wan et al., 2011). Mean age of the group was 6;5 (years;months), *SD* = 1;7, range: 3;5–10;8. All children met criteria for ASD on either the Autism Diagnostic Observation Scale, Module 1 (Lord et al., 2002), administered by a research-reliable examiner, or on the Childhood Autism Rating Scale–Second Edition (Schopler & van Bourgondien, 2010). Minimally verbal status, defined by a spontaneous expressive vocabulary of fewer than 20 words, was ascertained by parent report and documentation of fewer than 20 words produced during a language sample. Vocalizations were considered words/word approximations if they were recognizable to an unfamiliar conversational partner (the examiner) during the language sample at baseline. Based on parent report and examination of medical records, none of the children had additional diagnoses of other neurodevelopmental conditions (e.g., Down's syndrome, hearing impairment) that might have affected their verbal status. Demographic information for these 38 participants appears in Table 1.

Stimuli: Study 1

Children had been administered a series of baseline assessments during which they attempted to repeat a set of

Table 1. Demographic information for Study 1 participants.

	M (*SD*)	Range
Age (*n* = 38)	6;5 (1;7)	3;5–10;8
Phonetic Inventory[a]	8.0 (5.4)	0–21
CARS[b] **(*n* = 25)**	37.4 (5.3)	30–47
ADOS[c] **(*n* = 24)**	20.4 (3.5)	16–28
MSEL VR[d] **(*n* = 26)**	30.3 (9.8)	13–48
RL[e] **(*n* = 19)**	19.5 (8.4)	8–36
EL[f] **(*n* = 19)**	11.3 (2.7)	6–19
Sex (female)	18.4%	
Race		
Asian	31.6%	
African American	—	
White	63.2%	
Multiracial	5.3%	
Ethnicity		
Hispanic	15.8%	

[a]Phonetic inventory: number of English consonant and vowels repeated correctly (max = 36). [b]Childhood Autism Rating Scale. Score ≥ 30 for autism spectrum disorder. [c]Autism Diagnostic Observation Schedule. Score ≥ 16 for autism spectrum disorder. [d]Mullen Scales of Early Learning (Mullen, 1995), Visual Reception subtest. Raw score (max = 50). [e]MSEL, Receptive Language subtest. Raw score (max = 48). [f]MSEL, Expressive Language subtest. Raw score (max = 50).

30 high-frequency bisyllabic words or phrases relevant to their activities of daily living. Imitation accuracy ranged from approximately 1% to 62% at baseline. Video recordings of the participants repeating the words "baby," "cookie," "daddy," "mommy," and "puppy" 5 times each were subsequently coded for signs of CAS by speech-language pathologists (SLPs) with experience in childhood motor speech disorders (details below). These words were selected because they contained relatively early-developing consonants and because the majority of participants responded to all five prompts for each word.

Perceptual Coding and Reliability: Study 1

Videos were coded for the signs of CAS from Chenausky et al. (2020), adapted from Iuzzini-Seigel et al. (2015). Coding was performed for a different study and took place before the current one was initiated, so the judges were in effect blind to the purposes of this study. Only *increased difficulty with multisyllabic words* was not assessed, as all stimuli were bisyllabic. Table 2 lists the signs, and operational definitions appear in the Appendix.

Two SLPs experienced in pediatric motor speech disorders (K. C. and A. B.) consensus-coded the videos. Disagreements and questions were resolved through discussion with a senior SLP with expertise in CAS (A. M.). Subsequently, a set of eight randomly selected videos (additional baseline videos of a subset of participants, not previously viewed) was independently coded by the first two judges in order to establish interjudge reliability. Of the 38 participants, 27 showed at least five signs of CAS and were judged to present with difficulties in planning and sequencing speech movements.

Figure 2 shows an utterance where signs of difficulty in planning and sequencing speech movements are evident. It is a spectrogram of case study participant C1 repeating the word "baby," broadly transcribed [bwɪs.dɛ]. Features of CAS (i.e., those that suggested movement planning difficulty) were lengthened formant transitions from the initial [b] to the first vowel, giving rise to the percept of an inserted [w] (difficulty with coarticulation, addition error); incorrect vowel in the first syllable (vowel error); excessive vocal tract constriction after the first vowel, giving rise to the percept of an inserted [s] (addition error); a long interval between the syllables (syllable segregation, slow rate); error on the second [b] (consonant distortion); and incorrect vowel in the second syllable (vowel error).

A two-way mixed-effects intraclass coefficient for consistency and single measures was .95, $p < .001$ over all signs. Interjudge percent agreement on diagnosis of CAS was 87.5% (7/8), and Cohen's κ was .83 ("very good").

Analytic Strategy: Study 1

K-means clustering was used to divide participants into two subgroups. In this method, only the number of clusters is specified a priori. The algorithm then partitions the data into the selected number of clusters, assigning each data point to the cluster with the nearest mean. The number of times each child showed each of the 12 signs over 25 utterances (five repetitions of five words) was entered into the clustering analysis with k = 2. Note that our aim was not to test whether two clusters was the best fit to the data, but to determine whether, if the data were divided into only two clusters, the clusters would show the predicted characteristics.

The perceptually identified signs of CAS were then associated with the constructs of "searching articulation," "phoneme distortion," and "token-to-token variability" from Terband et al. (2014) based on a factor analysis of signs of CAS (Chenausky et al., 2020) and the putative contribution of each sign to the construct. The factor analysis identified three underlying factors that give rise to the observable signs of CAS: disrupted coarticulatory transitions (onto which the signs *difficulty with coarticulation, groping,* and *addition error* loaded most highly), variable errors (onto which the sign *inconsistent errors* loaded most highly), and inappropriate prosody (onto which the signs *syllable segmentation, slow rate,* and *stress error* loaded most highly). The signs *consonant distortion, voicing error, nasality error,* and *vowel error* did not load highly onto any one factor. Based on these results, we associated *difficulty with coarticulation, groping,* and *addition error* with "searching articulation" from Terband et al. (2014). *Inconsistent errors* was associated with "token-to-token variability" (between-token variability). *Consonant distortion, voicing error, nasality error,* and *vowel error* were associated with "phoneme distortion." The remaining signs, relating to prosodic disturbances, were not associated with any construct from Terband et al. (2014) because that work did not consider prosody. The bottom row of Table 2 indicates the correspondence between the perceptual signs of CAS and the constructs from Terband et al. (2014). In all cases, an increase in the score for any perceptual sign would contribute to an increase in the construct with which it was associated. For example, higher scores for any of *difficulty with coarticulation, groping,* or *addition error* would lead to an increase in searching articulation.

After the k-means analysis, a one-way analysis of variance was performed to determine which signs differed between clusters. Means, standard deviations, and analysis of variance results by cluster for each sign used in the k-means analysis also appear in Table 2.

Results and Discussion: Study 1

The two clusters produced by k-means analysis contained $n = 27$ (Cluster 1) and $n = 11$ (Cluster 2) participants, respectively. Clusters did not differ by mean age ($p > .07$). Cluster 1 included all five female participants.

Cluster 1 had significantly higher mean scores than Cluster 2 for the signs *difficulty with coarticulation* (10.0 vs. 2.9, respectively, $p = .04$) and *groping* (5.6 vs. 1.6 respectively, $p = .01$), which were associated with searching articulation (within-token variability). Mean scores for *addition error,* the third sign associated with searching articulation, did not differ significantly between clusters (9.3 for Cluster 1

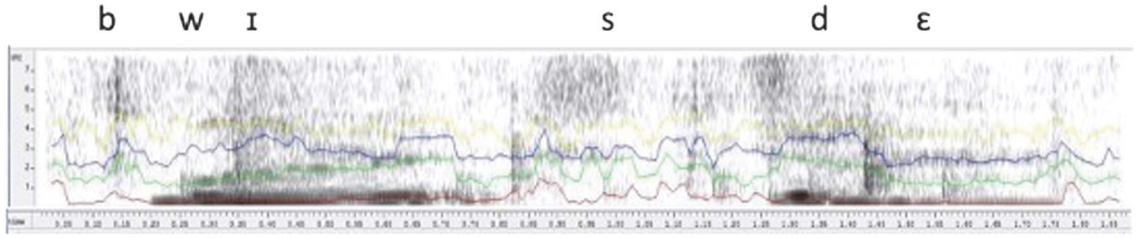
Table 2. Means and standard deviations for signs of childhood apraxia of speech in Study 1, by cluster.

Cluster		Age	Consonant distortion	Voicing error	Nasality error	Vowel error	Addition error	Difficulty w/ coarticulation	Silent groping	Inconsistent errors	Intrusive schwa	Syllable segmentation	Stress error	Slow rate
1 (*n* = 27, 5 F)	**μ**	6;4	18.0	10.4	5.6	24.4	9.3	10.0	5.6	3.3	3.5	3.2	2.6	3.1
	SD	1;7	9.7	5.5	8.5	9.7	7.0	10.8	4.9	3.7	4.3	3.6	6.9	3.7
2 (*n* = 11, 0 F)	**μ**	6;10	14.8	11.2	4.9	25.1	5.9	2.9	1.6	1.8	1.0	17.5	11.2	16.6
	SD	1;10	6.8	8.0	8.6	6.0	4.1	3.6	2.0	1.9	1.4	10.0	13.8	9.1
***p* value**		n.s.[a]	n.s.	n.s.	n.s.	n.s.	n.s.	.04	.01	n.s.	n.s.	< .001	.01	< .001
			Signs relating to phoneme distortion					Signs relating to searching articulation (within-token variability)		Sign relating to token-to-token variability		Signs relating to prosody		

Note. F = female; n.s. = not significant.

[a]All nonsignificant *p* values > .07.

Figure 2. Spectrogram of a participant repeating "baby," illustrating several features of CAS and of the acoustic indices. A broad transcription appears at the top, with formant tracks below. See text for details.



vs. 5.9 for Cluster 2, $p > .05$). Cluster 1's mean score for *inconsistent errors,* which was associated with token-to-token (between-token) variability, was also insignificantly higher than that for Cluster 2 (3.3 vs. 1.8, $p > .05$). Clusters did not differ on mean scores for *consonant distortion, voicing error, nasality error,* or *vowel error,* signs associated with phoneme distortion; all $p > .07$. Finally, Cluster 2 had significantly higher mean scores than Cluster 1 for the signs *intrusive schwa, syllable segmentation, stress error,* and *slow rate,* which were associated with prosodic disturbances. Based on the findings that Cluster 1 showed higher perceptually rated within- and between-token variability than Cluster 2, we matched Cluster 1 with DIVA's MPD condition and Cluster 2 with DIVA's MPD + APD condition.

Study 2: Case Study

Case Study Participants

We employed a case-study approach for the acoustic analyses for two reasons. First, of the 38 participants in our overall sample, 27 children had been selected to participate in an intonation-based therapy (described in Wan et al., 2011 and Chenausky et al., 2016) involving simultaneous tapping of tuned drums and production of sung syllables. Since baseline assessments for these children were administered in this modality, these stimuli were not suitable for acoustic analysis.

Second, the remaining 11 participants' broad age range (3;9–8;5) meant that their vocal tracts, and thus their vowel spaces, would differ in size because of age alone. This precluded the use of group acoustic analyses using formant values. Therefore, we simply selected the two participants (one from each cluster) with the largest number of acoustically analyzable tokens and performed the same acoustic analyses used in Terband et al. (2014) on their speech. Table 3 gives case study participants' demographic information. Note that both children were male and, consistent with Terband et al. (2014), both met criteria for CAS.

Case Study Stimuli

For the acoustic analysis, target syllables containing the corner vowels /a/, /æ/, /i/, and /u/ and the mid-central vowel /ʌ/ (synonymous in this context with schwa) in various consonant contexts were used. These syllables were taken from the set of 30 bisyllabic words/phrases used in the baseline assessments. Syllables were analyzed separately because both participants showed syllable segregation (i.e., they produced the stimuli with a pause between syllables, as illustrated in Figure 2). The Cluster 1 participant produced 170 analyzable syllables, and the Cluster 2 participant produced 256 analyzable syllables. Example transcriptions of each child's responses to stimuli containing corner vowels are included as Supplemental Material S1. For simplicity, below, we refer to the Cluster 1 participant as C1 and the Cluster 2 participant as C2.

Case Study Analyses

Formant-based acoustic measures used in the case study were adapted from the acoustic indices of phoneme distortion, searching articulation, and token-to-token variability from Terband et al. (2009) and Terband et al. (2014). We also calculated vowel space areas for participants C1

Table 3. Case study participant characteristics.

Participant (cluster no.)	Age	MSEL VR[a]	ADOS[b]	KSPT[c]	PPVT-4[d]	NDW[e]
1	7;7	43	23	7	18	0
2	6;2	28	28	6	7	1

Note. Both case study participants were male and met criteria for CAS.

[a]MSEL VR = Mullen Scales of Early Learning, Visual Reception subtest. A maximum of 50 raw score points is possible. [b]ADOS = Autism Diagnostic Observation Schedule algorithm score. Scores ≥ 16 indicate a diagnosis of autism. [c]KSPT = Kaufman Speech Praxis Test, raw score for Sections 1 and 2. A maximum of 74 raw score points is possible. [d]PPVT = Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), raw score. [e]NDW = number of different words produced during a natural language sample.

and C2 (Vorperian & Kent, 2007) and compared them to normative values from Assmann and Katz (2000). Specifics for each measure appear below.

Formants were extracted automatically using WaveSurfer (Sjölander & Beskow, 2000). Each token was trimmed at consonant release (for tokens beginning with stops) or at the onset of voicing (for tokens not beginning with stops), and at the cessation of periodic voicing. Note that, because of participants' voicing errors, all consonant-initial tokens in the acoustic analysis set contained voiced consonants, regardless of the voicing status of the targets. Depending on the analysis, information from the consonant and the vowel, or just the vowel, was used (details below).

We selected WaveSurfer based on the analysis in Derdemezis et al. (2016) showing no significant differences between fundamental frequency (F0) values for 5- to 10-year-old male children with Down's syndrome obtained with WaveSurfer's automatic formant tracking feature and those obtained by consensus measurements from two experienced speech acousticians. We used default settings in WaveSurfer's formant track extraction feature, except to extract three formants instead of four, use a pre-emphasis factor of .99 instead of .7, and use LPC (linear predictive coding) order 12 instead of 10. When necessary, F1 tracks were hand-edited to ensure that they captured the first formant instead of F0. This was done by comparing the formant tracks to wide- and narrowband spectrograms of the tokens, listening to the tokens, and taking into account the expected formant value for the tokens as they were actually produced, as recommended in Derdemezis et al. (2016). Tokens where the automatically extracted F1 matched the actual F1 of the utterance were used unedited.

To compare the acoustic measures between participants C1 and C2, two-tailed *t* tests for equal or unequal variances were used, as appropriate.

Calculation of Acoustic Indices

Because Terband et al. (2014) used stimuli in the form [V_1CV_2], which differed from the CV (consonant–vowel) stimuli produced by our participants, we adapted their indices of phoneme distortion, searching articulation, and token-to-token variability to our stimuli. Predicted values from Terband et al. (2014) for the three indices were estimated from plots in their Figure 7 (p. 25), which show the value for each index as a function of noise level. We used values of each index at 10% noise (the extreme right side of each plot) for the relevant phonemes ($C+V_2$ or just V_2) in our analyses. "Noise level," here, refers to the amount of Gaussian noise added to DIVA's motor, somatosensory, and/or auditory state maps, as described earlier. Ten percent noise was chosen as it was the highest level that resulted in realistic speech movements; model performance breaks down dramatically at higher noise levels (Terband et al., 2014).

Phoneme distortion measures how much a token's actual formant values differ from target values. The predicted level of phoneme distortion was obtained by estimating the value of this index for V_2 at the 10% noise level from Figure 7 (lower left panel) in Terband et al. (2014) for the MPD and MPD + APD conditions. Values for V_2 were used because this vowel follows a consonant, as did those in our stimuli. At 0% noise (extreme left side of plot), reflecting no disorder, the phoneme distortion index is approximately 0.02.

The actual Phoneme Distortion Index was obtained by comparing our participants' formant values for the corner vowels and schwa to normative figures from Assmann and Katz (2000). These authors estimated F1, F2, and F3 for three 7-year-olds from measurements on six repetitions each of /hVd/ syllables. They made five measurements, beginning at 33% of each vowel's duration, and took the median of these. They then averaged the medians for F1, F2, and F3 for each vowel over all 18 tokens to derive their normative averages.

We employed a parallel measurement strategy. First, for direct comparison to Terband et al. (2014), we log-normalized formant values from our participants and from Assmann and Katz (2000). Next, as in Assmann and Katz, we located the center five formant values for each of our participants' tokens and took the median. Then, following Terband et al. (2014), we found the absolute difference between each participants' log-normalized F1, F2, or F3 value and the corresponding log-normalized normative value. Finally, again following Terband et al. (2014), we averaged the absolute differences for each participant over all formants and all vowels to yield the Phoneme Distortion Index.

Searching articulation is a measure of within-token variability. The predicted Searching Articulation Index was estimated by taking values at the 10% noise level from Figure 7 (upper right panel) in Terband et al. (2014) and averaging over $C+V_2$ for the MPD and MPD + APD conditions. At 0% noise, the Searching Articulation Index is approximately 0.00375.

To calculate the actual Searching Articulation Index on our participants' tokens, we first log-normalized F1, F2, and F3 values for each token. Next, we calculated the standard deviation over the entire formant track (C+V) for each formant in each token. We then averaged the standard deviations over all formants and all tokens for each participant to yield the actual Searching Articulation Index.

Token-to-token variability is a measure of variability between different attempts at the same target (i.e., between-token variability). The predicted value was estimated at the 10% noise level for V_2 (postconsonantal) from Figure 7 (lower right panel) in Terband et al. (2014) for the MPD and MPD + APD conditions. At 0% noise, the Token-to-Token Variability Index is zero.

To calculate the actual token-to-token variability, we used the median log-normalized values for F1, F2, and F3 for each participant's vowels. We calculated the standard deviation of these median values for each vowel and then averaged over all formants and all vowels to yield each participant's actual Token-to-Token Variability Index.

Figure 2 also illustrates the constructs of phoneme distortion, within-token variability, and between-token variability. For example, phoneme distortion is evident

in both productions of /b/, which is realized as [bw] in the first syllable and as [d] in the second. This example also illustrates between-token variability, since the first and second attempts differ. The lengthened formant transitions of the first attempt at /b/ also illustrate within-token variability.

The final acoustic measure we employed was *vowel space,* which measures the acoustic distinctiveness of a speaker's vowels in terms of F1 and F2. To assess this, F1 and F2 values for the corner vowels and [ʌ] for both participants were plotted and compared to the Assmann and Katz (2000) values for typically developing 7-year-olds. First, we compared participants' F1 and F2 values for [ʌ] to ensure that there were no differences in vocal tract size that would affect the results. Then, we created two vowel spaces for each participant. One used mean F1 and F2 values for each corner vowel; the other used the extreme F1 and F2 values for each corner vowel (e.g., the [i] token with the lowest F1 and highest F2). Vowel space area for both the mean and extreme cases was calculated using the equation Area = 0.5 x [(F2i x F1æ + F2æ x F1a + F2a x F1u + F2u x F1i) – (F1i x F2æ + F1æ x F2a + F1a x F2u + F1u x F2i)] (Vorperian & Kent, 2007).

Case Study Results and Discussion

Two- tailed *t* tests for unequal variance showed no between-participant differences in mean F1 and F2 values for [ʌ]. Mean F1 was 855 Hz (*SD* = 262) for participant C1 and 820 Hz (*SD* = 107) for participant C2 ($p = .43$). Mean F2 was 1968 Hz (*SD* = 376) for C1 and 2083 Hz (*SD* = 284) for C2 ($p = .12$). Thus, participants' vocal tracts were commensurate in size and their vowel spaces could be appropriately compared.

The calculated value of the Phoneme Distortion Index for participant C1 was 0.10, versus 0.11 for participant C2. A two-tailed *t* test for unequal variances showed that the difference in means of 0.01 was not significantly different ($p = .93$), indicating that the two participants showed similar amounts of phoneme distortion. This agrees with the prediction of similar amounts of phoneme distortion for the two DIVA cases (0.01 for both the MPD and the MPD + APD conditions).

The calculated value of the Searching Articulation Index for participant C1 was 0.25, compared to 0.02 for participant C2. A two-tailed *t* test for unequal variances showed that the difference in means of 0.22 was statistically significant ($p < .001$). This agrees with the prediction that the MPD case would show more searching articulation (0.02) than the MPD + APD case (0.015).

Finally, the calculated value of the Token-to-Token Variability Index was 0.09 for C1 and 0.06 for C2. A two-tailed *t* test for unequal variances showed a statistically significant difference in means, $\Delta = 0.04$ ($p = .04$). This agrees with the prediction that there would be more token-to-token variability in the MPD case (0.015) than in the MPD + APD case (0.010). Calculated and predicted values for all three indices appear in Table 4.

Figure 3 shows vowel space (F1 vs. F2) plots for (A) typically developing 7-year-old speakers (derived from Assmann & Katz, 2000), (B) C1 (the high-variability participant from Cluster 1), and (C) C2 (the low-variability participant from Cluster 2). For each participant, two versions of the vowel space area are shown. The "mean vowel space area" was calculated from the mean F1 and F2 values for each corner vowel. The "extreme vowel space area" was calculated from the most extreme token for each corner vowel.

The vowel space area of the typical 7-year-olds was 680314 Hz^2. The mean vowel space area for participant C1 was 30876 Hz^2, and his extreme vowel space area was 1471577 Hz^2. The mean vowel space area for participant C2 was 1895 Hz^2. However, his extreme vowel space area was just 358631 Hz^2, smaller than the mean vowel space for typically developing 7-year-olds and the extreme vowel space of participant C1. This is consistent with the prediction that the vowel space area for C2 would be highly centralized.

General Discussion

In this article, we tested whether predictions of the DIVA model, when it was given either an MPD alone or an MPD combined with an auditory processing disorder, were consistent with perceptual and acoustic analyses of speech produced by minimally verbal children with ASD.

Cluster Analysis

In Study 1, k-means analysis was used to divide 38 participants into two clusters based on perceptually identified signs of CAS. The clusters showed similar mean scores

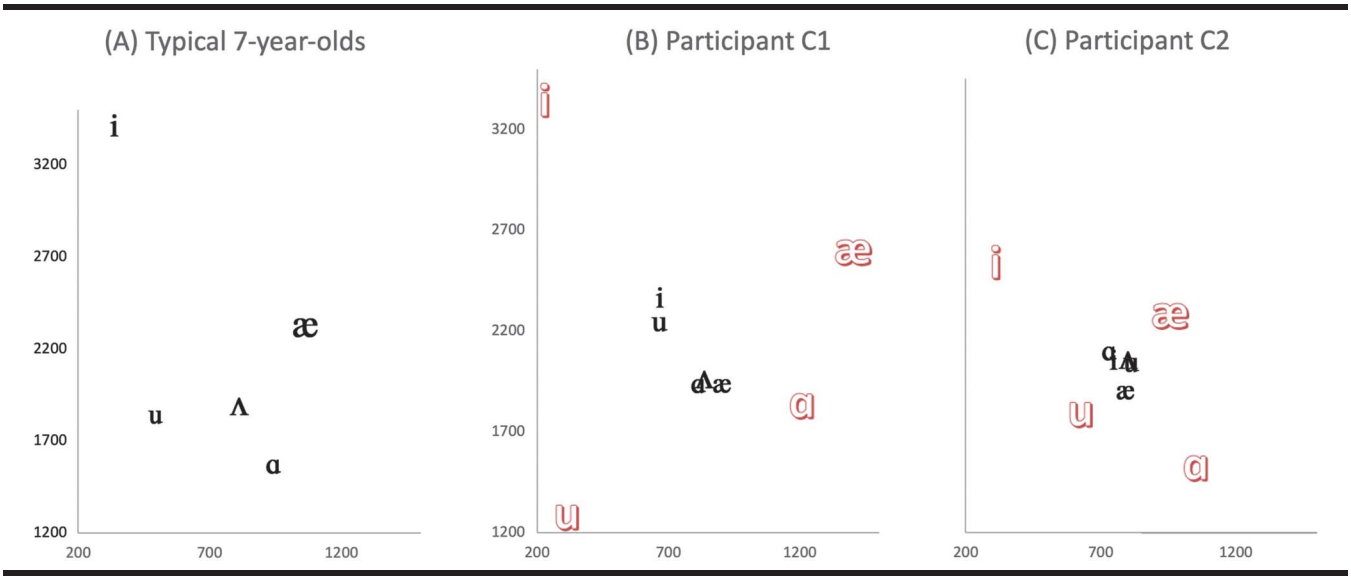
Table 4. Acoustic indices of Phoneme Distortion, Searching Articulation, and Token-to-Token Variability for Study 2 participants, compared to model predictions.

Values	Participant/condition	Phoneme distortion	Searching articulation	Token-to-token variability
Calculated	C1	0.100	0.250	0.090
	C2	0.110	0.020*	0.060**
Predicted	MPD	0.010	0.020	0.015
	MPD + APD	0.010	0.015	0.010

Note. MPD = motor processing disorder; APD = auditory processing disorder.

*Significant difference between participants, $p = .001$. **Significant difference between participants, $p = .04$.

Figure 3. Vowel space plots for Study 2. Solid black symbols represent mean vowel spaces; red outlined symbols represent extreme vowel spaces (details in text). (A) Typically developing 7-year-olds (Assmann & Katz, 2000), (B) Participant C1 (high-variability participant), and (C) Participant C2 (low-variability participant).



for signs corresponding to phoneme distortion (*consonant distortion, voicing error, nasality error,* and *vowel error*). Cluster 1 showed significantly greater mean scores than Cluster 2 for two of the three signs corresponding to searching articulation (*difficulty with coarticulation* and *groping*), but a nonsignificantly greater score for *addition error*. Cluster 1 also showed a nonsignificantly greater score for *inconsistent errors* (token-to-token variability) than Cluster 2. Based on this, we equated Cluster 1, the high-variability group, to DIVA's MPD condition and Cluster 2, the low-variability group, to DIVA's MPD + APD condition. It should be noted, however, that these matches are qualitative rather than quantitative. Although the direction of the differences between the MPD and MPD + APD model indices are the same as the subjects', the magnitudes of the indices and their differences are often substantially larger in the participants compared to the model. One possible reason for this is that the model's performance becomes highly unnatural at noise levels greater than 10% (Terband et al., 2014), likely due to implementational details in the model such as time step size and limits on cell activation imposed by the software. Additionally, the equating of neural impairment to simple Gaussian noise within the model's cells is a simplification that is necessary to make the results interpretable in this initial study. Future modeling, using more sophisticated implementations of neural impairments, may more fully capture the behavior of some impaired populations. This issue will be discussed in more detail below.

The prediction that clusters would differ according to token-to-token variability (*inconsistent errors*) was not upheld. However, the inability to reach a significant difference between clusters for this sign may have arisen from the way it was scored. The range of possible scores for this sign is smaller than for the others: While most signs can be rated multiple times per token, *inconsistent errors* is only rated once per target when at least two tokens containing different errors occur. Scores for this sign thus range only between 1 and 5, 5 being the number of targets examined in our perceptual study. Increasing statistical power by adding more targets, which allows for a larger mean difference between groups than found here, or by including a larger number of participants, may be needed to demonstrate statistical significance. The lack of statistical significance in the between-group difference for *addition error* may also have been due to low statistical power.

The finding of significantly larger mean scores for prosodic signs (*syllable segmentation, stress error,* and *slow rate*) for Cluster 2 than for Cluster 1 also deserves comment. DIVA makes no predictions about the effect of developmental speech disorders on prosody because it models speech production only at the segmental level; suprasegmental prosodic characteristics are beyond its scope. However, prosodic impairment is one of the three consensus criteria for CAS (ASHA, 2007), along with inconsistent errors and distorted transitions, so it is not surprising that both clusters showed evidence of prosodic impairment. That said, no research on the effect of comorbid auditory processing disorders on signs or severity of CAS exists, but it is possible that children with this comorbidity experience greater prosodic disturbances than children with CAS alone. This is also a testable hypothesis.

Acoustic Analyses

Next, we discuss the findings from the case study acoustic analyses. Both participants, though selected for their number of analyzable tokens, met criteria for CAS. The Phoneme Distortion Index was similar for the two participants,

as predicted. Indices of searching articulation (within-token variability) and token-to-token (between-token) variability were larger for C1 than for C2, also predicted. Finally, the prediction that because of coalescence to schwa, C2's vowel space would be much smaller than that of typically developing age-matched speakers was upheld.

How might the different putative comorbidities (i.e., MPD, with or without APD) have affected the vowel space results? No specific predictions about vowel space were made for DIVA's MPD-only case (Cluster 1). Though participant C1's mean and extreme vowel spaces were larger than those of C2, C1's mean vowel space area was still much smaller than that of typically developing children. However, C1 also showed a very large extreme vowel space and highly variable distortions—in some sense, he was wrong in a different way each time he tried the same target. On the other hand, C2's vowels were more consistent, though generally wrong unless the target was [ʌ]. These between-participant differences were reflected in larger standard deviations for C1's formants than C2's. That is, C1's compressed mean vowel space seems to have arisen through averaging tokens with widely different formant values for each target (errors above and below the mean canceling each other out), while C2's compressed mean vowel space appears more a consequence of an extremely limited vowel inventory.

A related question is why an APD did not increase phoneme distortion when added to an MPD. A more detailed discussion appears in Terband et al. (2014), but for current purposes, it is sufficient to note that "noisy" auditory representations have the effect of making errors between target and actual output harder to detect, so that online corrective action (searching articulation) is not induced. Noisy auditory representations do cause errors in DIVA's motor state representations during its developmental stages, but because the deficit is implemented with zero-mean noise, these errors tend to cancel each other out over time. Thus, the auditory deficit has little net effect on DIVA's speech output.

Further pursuing the issue of disordered auditory perception, one might speculate whether children like participant C2, whose speech is highly centralized, also experience poorer phoneme discrimination. As implemented in DIVA with an MPD + APD, noisy auditory state maps blur distinctions between phonemes, making them sound more similar to each other. Thus, Terband et al. (2014) predict that an MPD + APD would make speech perception less categorical. If this is true, we would expect that minimally verbal children with CAS whose speech is highly centralized would show poorer between-category discrimination than minimally verbal children with CAS whose speech is highly variable. While this has not been tested in minimally verbal children with ASD, as mentioned in the introduction, Stewart et al. (2018) showed that at least some verbal children with ASD have poorer between-category phoneme discrimination than typical peers. Several researchers have also found that children with CAS who do not have ASD show poorer discrimination of minimal pair continua than typically developing peers (Groenen et al., 1996; Maassen et al., 2003; Nijland, 2009), though others have reported similar phoneme discrimination between children with CAS and typical peers (Zuk et al., 2018).

Since the two case study participants were of different ages, it is reasonable to ask whether the differences in their speech profiles were age related. The design of this study cannot rule that out. However, clinical observation suggests that the differences were not due to age differences. Recall that both C1 and C2 participated in a treatment study, the total duration of which (from baseline to posttreatment assessment) was approximately 6 months. The speech of both participants remained qualitatively similar after their participation in the treatment study: C1's speech was still highly variable, while C2's speech continued to be highly centralized. Thus, it is unlikely that the differences between participants were due to their different ages at baseline.

Conclusions

Taken all together, the current findings provide at least preliminary evidence for the existence of two groups of minimally verbal children with ASD who may experience additional comorbidities of MPD (Cluster 1) or MPD and APD (Cluster 2). They are consistent with previous findings in an independent group of minimally verbal children with ASD that a subgroup meets criteria for CAS (Chenausky et al., 2019). Thus, we conclude provisionally that Cluster 1 may exemplify a subphenotype of minimally verbal children with ASD characterized by comorbid CAS, and that Cluster 2 may exemplify a subphenotype characterized by comorbid CAS and auditory processing disorder. These are hypotheses that are testable behaviorally (see Limitations and Future Work section for further discussion of this point).

Note that Terband et al. (2014) equate the MPD-only case to CAS plus a phonological component and the MPD + APD case to a typical apraxic speech disorder (i.e., CAS alone). In their view, the presence of an MPD during development causes deterioration in the language-specific phonemic mappings between articulatory and auditory goals, introducing a phonological component to the apraxia. By contrast, they see the presence of both motor and auditory processing disorders during development as impairing the language-independent mappings between motor commands and their sensory consequences, corresponding to CAS without a phonological component.

However, the question of how CAS should be instantiated in the DIVA model is far from decided, despite the initial parallels made by Terband et al. (2014). The current perceptual findings suggest that the symptom complex we refer to as CAS may correspond more closely to the MPD-only case of Terband et al. (2014), while the MPD + APD case may represent CAS plus an additional auditory-processing comorbidity. Though the core impairment in CAS is described as being one of planning or programming the spatiotemporal parameters of speech movements (ASHA, 2007), it is further possible that there are subcomponents

of movement planning and programming that may be separately disordered and still give rise to the perceptual signs that define CAS. A full treatment of these theoretical points is beyond the scope of this article, but see Shriberg et al. (1997) for one such discussion. Ultimately, disentangling the component processes that make up "planning or programming the spatiotemporal parameters of speech movements" and relating them to disordered modules of computational models of speech production must be the work of future research.

Clinical Implications

The results of this study support the clinical hypotheses that subgroups of minimally verbal children with ASD experience comorbid motor speech disorders and/or auditory processing disorders. They also suggest that the degree of variability in children's speech will correlate with these additional comorbidities.

The present results further suggest a mechanism for how comorbid CAS and/or APD may affect speech development in minimally verbal children with ASD and inform treatment planning. If an APD diminishes children's ability to distinguish acoustically similar but phonetically different tokens, children with highly centralized utterances may need training in identification of auditory-perceptual contrasts in speech (e.g., minimal pair training) in addition to treatment designed to address the underlying motor planning impairment. Minimal pair training has been used successfully in verbal children with phonological disorders (Rvachew et al., 1999). Treating only the motor aspects of speech in children with CAS + APD may improve consistency of production but not intelligibility (i.e., it may make them more consistent in their errors). In contrast, therapy for children with highly variable speech might focus on tasks designed to increase both accuracy and consistency.

Limitations and Future Work

Because of the high degree of phenotypic heterogeneity in the speech of minimally verbal children with ASD, our results must be considered preliminary. In particular, the acoustic results should be replicated in a larger group of children in order to determine whether the between-subjects differences are statistically significant. The arbitrariness of the selection criterion for the current acoustic analysis (choosing the child from each cluster with the largest number of analyzable tokens) may lend some credence to the present results, but more work must be done to understand how prevalent the two profiles delineated here are in the population of minimally verbal children with ASD overall. In addition, careful and thorough clinical assessment of the speech production and speech perception abilities of a larger sample of minimally verbal children with ASD should be carried out to obtain a better prevalence estimate of comorbid CAS and/or APD.

Other testable hypotheses suggested by the current findings include determining whether minimally verbal children whose speech is highly centralized also experience auditory processing disorders characterized by poor phoneme discrimination. This could be accomplished by adapting shape- or color-matching tasks that are commonly taught to these children. If a child can be taught to accurately recognize two visually distinct stuffed animals named, for example, "bat" and "dat," it may be feasible to test their discrimination and identification ability of a /bæt/-/dæt/ continuum. Well-designed single-case or group treatment studies are also needed in order to determine whether treatment for CAS is effective in minimally verbal children with ASD + CAS, and to determine the extent to which training in auditory perception of speech may help minimally verbal children with ASD + CAS + APD. We hope that the studies reported on here will be only the start of a longer investigation into these important issues of basic and clinical science.

Author Contributions

Karen V. Chenausky: Conceptualization (Lead), Formal analysis (Lead), Funding acquisition (Supporting), Investigation (Lead), Methodology (Lead), Validation (Equal), Writing – original draft (Lead), Writing – review & editing (Equal). **Amanda Brignell:** Validation (Equal), Writing – review & editing (Supporting). **Angela T. Morgan:** Supervision (Supporting), Validation (Supporting). **Andrea C. Norton:** Data curation (Supporting), Writing – review & editing (Supporting). **Helen B. Tager-Flusberg:** Funding acquisition (Equal), Supervision (Supporting), Writing – review & editing (Supporting). **Gottfried Schlaug:** Funding acquisition (Equal), Resources (Equal). **Frank H. Guenther:** Methodology (Supporting), Supervision (Supporting), Writing – review & editing (Supporting).

Acknowledgments

This work was supported by P50 DC 13027 to H. T. F. (supporting H. T. F., G. S., A. C. N., and K. V. C.), P50 DC 18006 to H. T. F. (supporting H. T. F. and K. V. C.), the Nancy Lurie Marks Family Foundation and Autism Speaks (supporting G. S.), T32 DC 013017 to C. A. Moore (supporting K. V. C.), a Clinical Research Grant from the American Speech-Language-Hearing Foundation (supporting K. V. C.), K99 DC 017490 (to K. V. C.), an National Health and Medical Research Council (NHMRC) Practitioner Fellowship 1105008 (to A. M.), an NHMRC Centre of Research Excellence in Speech and Language Neurobiology 1116976 (to A. M.), and NHMRC Project Grant 1127144 (to A. M.).

We thank the children who participated in the studies leading to this article and their families for allowing them to participate. We are grateful to Jordan Green and Hayo Terband for their suggestions regarding earlier versions of the article. Finally, our thanks also go to two anonymous reviewers from *AJSLP*, whose detailed and incisive comments have greatly improved the article. Data sharing requests should be sent to Gottfried Schlaug at gschlaug@gmail.com.

References

Adams, L. (1998). Oral-motor and motor-speech characteristics of children with autism. *Focus on Autism and Other*

Developmental Disabilities, 13(2), 108–112. https://doi.org/10.1177/108835769801300207

American Psychiatric Association. (2013). *American psychiatric association task force on DSM-5: Diagnostic and statistical manual of mental disorders: DSM-5*. https://doi.org/10.1176/appi.books.9780890425596

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. http://www.asha.org/policy/TR2007-00278/

American Speech-Language-Hearing Association. (2005). *(Central) auditory processing disorders* [Technical report]. https://www.asha.org/policy/tr2005-00043/

Assmann, P., & Katz, W. (2000). Time-varying spectral change in the vowels of children and adults. *The Journal of the Acoustical Society of America, 108*(4), 1856–1866. https://doi.org/10.1121/1.1289363

Baghdadli, A., Pascal, C., Grisi, S., & Aussiloux, C. (2003). Risk factors for self-injurious behaviors among 222 young children with autistic disorders. *Journal of Intellectual Disability Research, 47*(8), 622–627. https://doi.org/10.1046/j.1365-2788.2003.00507.x

Belmonte, M., Saxena-Chandhok, T., Cherian, R., Muneer, R., George, L., & Karanth, P. (2013). Oral motor deficits in speech-impaired children with autism. *Frontiers in Integrative Neuroscience, 7,* Article 47. https://doi.org/10.3389/fnint.2013.00047

Brady, N., Bruce, S., Goldman, A., Erickson, K., Mineo, B., Ogletree, B., & Wilkonson, K. (2016). Communication services and supports for individuals with severe disabilities: Guidance for assessment and intervention. *American Journal on Intellectual and Developmental Disabilities, 121*(2), 121–138. https://doi.org/10.1352/1944-7558-121.2.121

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 Impairments, 4,* 1–12. https://doi.org/10.1177/2396941519856333

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,* 175. https://doi.org/10.3389/fnhum.2017.00175

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), Article e0164930. https://doi.org/10.1371/journal.pone.0164930

Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2018). Behavioral predictors of improved speech output in minimally verbal children with autism. *Autism Research, 11*(10), 1356–1365. https://doi.org/10.1002/aur.2006

Chenausky, K. V., Brignell, A., Morgan, A., Gagné, D., Norton, A., Tager-Flusberg, H., Schlaug, G., Shield, A., & Green, J. (2020). Factor analysis of signs of childhood apraxia of speech. *Journal of Communication Disorders, 87,* 106033. https://doi.org/10.1016/j.jcomdis.2020.106033

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

Choi, B., Leech, K. A., Tager-Flusberg, H., & Nelson, C. A. (2018). Development of fine motor skills is associated with expressive language outcomes in infants at high and low risk for autism spectrum disorder. *Journal of Neurodevelopmental Disorders, 10,* Article 14. https://doi.org/10.1186/s11689-018-9231-3

Derdemezis, E., Vorperian, H. K., Kent, R. D., Fourakis, M., Reinicke, E. L., & Bolt, D. M. (2016). Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. *American Journal of Speech-Language Pathology, 25*(3), 335–354. https://doi.org/10.1044/2015_AJSLP-15-0020

Dominick, K., Davis, N., Lainhart, J., Tager-Flusberg, H., & Folstein, H. (2007). Atypical behaviors in children with autism and children with a history of language impairment. *Research in Developmental Disabilities, 28*(2), 145–162. https://doi.org/10.1016/j.ridd.2006.02.003

Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Pearson Assessments. https://doi.org/10.1037/t15144-000

Gernsbacher, M., Sauer, E., Geye, H., Schweigert, E., & Goldsmith, H. (2008). Infant and toddler oral- and manual-motor skills predict later speech fluency in autism. *The Journal of Child Psychology and Psychiatry, 49*(1), 43–50. https://doi.org/10.1111/j.1469-7610.2007.01820.x

Groenen, P., Maassen, B., Crul, T., & Thoonen, G. (1996). The specific relation between perception and production errors for place of articulation in developmental apraxia of speech. *Journal of Speech and Hearing Research, 39*(3), 468–482. https://doi.org/10.1044/jshr.3903.468

Guenther, F. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72*(1), 43–53. https://doi.org/10.1007/BF00206237

Guenther, F., Ghosh, S., & Tourville, J. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 93*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Hartley, S., Sikora, D., & McCoy, R. (2008). Prevalence and risk factors of maladaptive behaviour in young children with autistic disorder. *Journal of Intellectual Disability Research, 52*(10), 819–829. https://doi.org/10.1111/j.1365-2788.2008.01065.x

Iuzzini-Seigel, J., Hogan, T., Guarino, A., & Green, J. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54,* 32–42. https://doi.org/10.1016/j.jcomdis.2015.01.002

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research, 6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kaufman, N. (1995). *The Kaufman Speech Praxis Test for Children.* Wayne State University Press.

Kent, R. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7–23. https://doi.org/10.1044/1058-0360.0503.07

Levy, S., Giarelli, E., Lee, L., Schieve, L., Kirby, R., Cunniff, C., Nicholas, J., Reaven, J., & Rice, C. (2010). Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States. *Journal of Developmental and Behavioral Pediatrics, 31*(4), 267–275. https://doi.org/10.1097/DBP.0b013e3181d5d03b

Lloyd, M., MacDonald, M., & Lord, C. (2011). Motor skills of toddlers with autism spectrum disorders. *Autism, 17*(2), 133–146. https://doi.org/10.1177/1362361311402230

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2002). *Autism diagnostic observation schedule*. Western Psychological Services.

Maassen, B., Groenen, P., & Crul, T. (2003). Auditory and phonetic perception of vowels in children with apraxic speech

disorders. *Clinical Linguistics and Phonetics, 17*(6), 447–467. https://doi.org/10.1080/0269920031000070821

Marco, E., Hinkley, L., Hill, S., & Nagarajan, S. (2011). Sensory processing in autism: A review of neurophysiologic findings. *Pediatric Research, 69*(5, pt. 2), 48R–54R. https://doi.org/10.1203/PDR.0b013e3182130c54

Mody, M., Shui, A., Nowinski, L., Golas, S., Ferrone, C., O'Rourke, J., & McDougle, C. (2017). Communication deficits and the motor system: Exploring patterns of association in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(1), 155–162. https://doi.org/10.1007/s10803-016-2934-y

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. AGS.

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

Prizant, B. (1996). Brief report: Communication, language, social, and emotional development. *Journal of Autism and Developmental Disorders, 26*(2), 173–178. https://doi.org/10.1007/BF02172007

Rapin, I., Dunn, M., Allen, D., Stevens, M., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology, 34*(1), 66–84. https://doi.org/10.1080/87565640802564648

Rvachew, S., Rafaat, S., & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *Journal of Speech, Language, and Hearing Research, 8*(1), 33–43. https://doi.org/10.1044/1058-0360.0801.33

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience, 18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Schopler, E., & Van Bourgondien, M. (2010). *Childhood Autism Rating Scale–Second Edition (CARS2)*. Western Psychological Services.

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. https://doi.org/10.1044/jslhr.4002.273

Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*(4), 405–426. https://doi.org/10.1007/s10803-010-1117-5

Sjölander, K., & Beskow, J. (2000). WaveSurfer—An open source speech tool. In B. Yuan, T. Huang, & X. Tang (Eds.), *Proceedings of ICSLP 2000, 6th International Conference on Spoken Language Processing* (pp. 464–467). International Speech Communication Association.

Soke, G., Maenner, M., Christensen, D., Kurzius-Spencer, M., & Schieve, L. (2018). Prevalence of co-occurring medical and behavioral conditions/symptoms among 4- and 8-year-old children with autism spectrum disorder in selected areas of the United States in 2010. *Journal of Autism and Developmental Disorders, 48*(8), 2663–2676. https://doi.org/10.1007/s10803-018-3521-1

Stewart, M. E., Petrou, A. M., & Ota, M. (2018). Categorical speech perception in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 48,* 72–82. https://doi.org/10.1007/s10803-017-3284-0

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6), 468–478. https://doi.org/10.1002/aur.1329

Terband, H., Maassen, B., Guenther, F., & Brumberg, J. (2009). Computational neural modeling of childhood apraxia of speech (CAS). *Journal of Speech, Language, and Hearing Research, 52*(6), 1595–1690. https://doi.org/10.1044/1092-4388(2009/07-0283)

Terband, H., Maassen, B., Guenther, F., & Brumberg, J. (2014). Auditory-motor interactions in pediatric motor speech disorders: Neurocomputational modeling of disordered development. *Journal of Communication Disorders, 47,* 17–33. https://doi.org/10.1016/j.jcomdis.2014.01.001

Vorperian, H., & Kent, R. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50*(6), 1510–1545. https://doi.org/10.1044/1092-4388(2007/104)

Wan, C., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., Norton, A., & Schlaug, G. (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, 6*(9), Article e2550. https://doi.org/10.1371/journal.pone.0025505

Wolk, L., & Giesen, J. (2000). A phonological investigation of four siblings with childhood autism. *Journal of Communication Disorders, 33*(5), 371–389. https://doi.org/10.1016/S0021-9924(00)00021-6

You, R., Serniclaes, W., Rider, D., & Chabane, N. (2017). On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in Developmental Disabilities, 61,* 158–171. https://doi.org/10.1016/j.ridd.2016.12.009

Zuk, J., Iuzzini-Seigel, J., Cabbage, K., Green, J., & Hogan, T. (2018). Poor speech perception is not a core deficit of childhood apraxia of speech: Preliminary findings. *Journal of Speech, Language, and Hearing Research, 61*(3), 583–592. https://doi.org/10.1044/2017_JSLHR-S-16-0106

Appendix

Operational Definitions of Signs of CAS

Definitions are taken from K. V. Chenausky et al. (2020), adapted from Iuzzini-Seigel et al. (2015).

1. **Consonant distortion**: A consonant production error in which a speech sound is recognizable as a specific phoneme but is not produced exactly correctly (e.g., an /s/ that is produced with lateralization or dentalization). Also includes substitutions (e.g., [to] for "so") and omissions (e.g., [ba] for "bob"). Not scored if the only consonant error is voicing or nasality. Thus, consonant distortions include manner/place distortions or substitutions, and omissions.
2. **Voicing error**: A sound is produced as its voicing cognate (e.g., a /p/ that is produced as a [b]). In addition, this could also describe productions which appear to be in between voicing categories (e.g., blurring of voicing boundaries). Note that glottal stop is considered neither voiced nor voiceless, so substitution of a glottal stop for another consonant does not trigger this error.
3. **Nasality error**: Sounds either hyponasal (not enough airflow out of nose/"stuffy") OR hypernasal (too much airflow out of nose for non-nasal phonemes such as plosives). Nasality errors can also occur if an oral stop is substituted for a nasal (e.g., [do] for "no"), if a nasal is substituted for an oral stop (e.g., [mi] for "bee"), or if a vowel in a word with no nasal consonant is heavily nasalized.
4. **Vowel error:** A vowel production error in which the vowel is substituted for another phoneme OR in which the vowel is recognizable as a specific phoneme but is not produced exactly correctly (e.g., it is not a prototypical production but may sound like it is in between two vowels). It is not considered an error if the vowel is substituted with another phoneme that is consistent with an adultlike model or a regional accent (e.g., /hɑtdɑg/, /hɑtdɔg/).
5. **Intrusive schwa** (e.g., in clusters): A schwa is added between consonants. For example, it may be inserted in between the consonants in a cluster (e.g., /blu/ becomes /bəlu/). This is NOT considered a "vowel error." Intrusive schwa may also occur before an initial consonant (e.g., [əbʌni] for "bunny") or adjacent to a vowel (e.g., [noə] for "no").
6. **Syllable segregation**: Brief or lengthy pause between syllables within a word which is not appropriate.
7. **Stress error**: An error in which the appropriate stress is not produced correctly. For example: conDUCT and CONduct have different stress patterns. It is considered an error if the stress is inappropriately equalized across syllables or placed on the wrong syllable. Addition of syllables (as in [dædədi] for "daddy") or deletion of syllables (as in [tɛfon] for "telephone") also count as stress errors, since they change the metrical structure of the word.
8. **Slow rate**: Speech rate is not typical. It is slower during production of part (e.g., zzziiiiiiper/zipper) or the whole word (e.g., tooommmmaaatoooo/tomato). Syllable segregation also triggers the "slow rate" error.
9. **Difficulty with coarticulation**: Initiation of utterance or initial speech sound may be difficult for child to produce and may sound lengthened, uncoordinated, or excessively effortful. Also, child may evidence lengthened or disrupted coarticulatory gestures or movement transitions from one sound to the next. For example, heavily prevoiced stops or words with a glottal stop inserted at the beginning fall into this category.
10. **Groping**: Prevocalic (silent) articulatory searching prior to onset of phonation, possibly in an effort to improve the accuracy of the production. Video is needed to assess this feature.
11. **Variable errors**: The same target is produced with different errors each time. Note that if a child produces an errored token once and a correct version once, this does not count as a variable error. The child must produce at least two distinct errored versions in order to trigger this error.
12. **Additions** (of phonemes other than schwa): The token contains phonemes or syllables that are not in the target. For example, [mɑmbi] for "mommy" would contain [b] as an addition (and would also trigger a "difficulty with coarticulation" error). Addition of syllable(s) also triggers the "stress" error.