

## Vowel variability and contrast in Childhood Apraxia of Speech: acoustics and articulation

Giovanna Lenoci, Chiara Celata, Irene Ricci, Anna Chilosi & Vincenzo Barone

**To cite this article:** Giovanna Lenoci, Chiara Celata, Irene Ricci, Anna Chilosi & Vincenzo Barone (2021) Vowel variability and contrast in Childhood Apraxia of Speech: acoustics and articulation, *Clinical Linguistics & Phonetics*, 35:11, 1011-1035, DOI: [10.1080/02699206.2020.1853811](https://doi.org/10.1080/02699206.2020.1853811)

**To link to this article:** <https://doi.org/10.1080/02699206.2020.1853811>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 16 Dec 2020.



Submit your article to this journal [↗](#)



Article views: 3087



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)

# Vowel variability and contrast in Childhood Apraxia of Speech: acoustics and articulation

Giovanna Lenoci<sup>a</sup>, Chiara Celata<sup>b</sup>, Irene Ricci<sup>a</sup>, Anna Chilosi<sup>c</sup>, and Vincenzo Barone<sup>a</sup>

<sup>a</sup>SMART Lab, Scuola Normale Superiore, Pisa, Italy; <sup>b</sup>Dipartimento di Studi Umanistici (DISTUM), Università Degli Studi Di Urbino 'Carlo Bo', Urbino, Italy; <sup>c</sup>IRCCS Fondazione Stella Maris, Pisa, Italy

## ABSTRACT

Phoneme production may be affected by limited speech motor control in Childhood Apraxia of Speech (CAS), with a general instability of acoustic targets across multiple repetitions of speech stimuli. This acoustic and Ultrasound Tongue Imaging (UTI) study shows that increased variability and reduction of contrast in vowel production is found in native Italian speakers with CAS, particularly as far as the height dimension is concerned. The data suggest that vowel production should play a major role in CAS diagnosis and treatment. Moreover, this study shows that a combined acoustic and articulatory approach allows direct observation of lingual dynamics together with an estimation of changes in the acoustic dimension. The two dimensions are shown not to correspond in a straightforward way in the speech of children with CAS, and encourage consideration of articulatory compensation strategies aimed at saving the acoustic identity of vowels.

## ARTICLE HISTORY

Received 2 June 2020  
Revised 14 November 2020  
Accepted 17 November 2020

## KEYWORDS

Vowel production;  
Childhood Apraxia of  
Speech; phonetic variability;  
phonemic contrast;  
ultrasound tongue imaging;  
Italian

## Introduction

### *Speech variability in Childhood Apraxia of Speech*

Childhood Apraxia of Speech (CAS) is a neurological speech motor disorder with a core deficit in planning and programming the spatio-temporal parameters of movement sequences (American Speech-Language-Hearing Association, 2007a; Shriberg et al., 2012; Maassen et al., 2010; Shriberg, 2010). However, diagnosis of CAS in children is challenging because of the potential presence of comorbid conditions and the lack of validated instruments to identify diagnostic features that differentiate CAS from other types of speech sound disorders (e.g., those due to phonological or neuromuscular impairments). According to past literature, children with CAS exhibit deficits across all levels of speech production, including lexical representation (Marquardt et al., 2004), phonological encoding (Thoonen et al., 1997), speech motor planning and speech motor programming. As such, CAS speech is often characterized by impaired motor control phenomena such as errors in the production of consonants and vowels, simplifications of clusters and complex syllables, prosodic anomalies and higher than normal token-to-token variability (see Chilosi et al., 2015; Forrest, 2003; Grigos et al., 2015; Iuzzini & Forrest, 2008; Iuzzini-Seigel et al., 2016; Jacks et al., 2006; Lewis et al., 2004; Marquardt et al., 2004). Among the symptoms, the observed inconsistency of speech sound articulation is essential for identifying children

**CONTACT** Giovanna Lenoci  [giovanna.lenoci@sns.it](mailto:giovanna.lenoci@sns.it)  Scuola Normale Superiore, Piazza Dei Cavalieri, 7, Pisa 56126.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

with (or at risk for) CAS, considering that token-to-token variability has been found by some scholars to be more marked in CAS than in other conditions of speech delay (e.g., Grigos et al., 2015; Iuzzini, 2012; Marquardt et al., 2004), particularly in the case of isolated multisyllabic words (whereas at the phrase level, children with CAS were not significantly different from children with language impairment, according to Iuzzini-Seigel et al., 2016). Token-to-token variability is usually assessed in terms of phonemes, that is, substituted or distorted phonemic targets across repetitions of the same phoneme in one and the same word context (e.g., Iuzzini, 2012; Iuzzini-Seigel et al., 2016); substitutions and distortions are perceptually identified by trained phoneticians and speech pathologists. Comorbid language disorder is regularly associated with CAS (e.g., Iuzzini, 2012; Lewis et al., 2004), which makes CAS symptoms difficult to isolate and treat. As a matter of fact, the linguistic and motor systems are strongly interconnected (e.g., Arbib, 2006) and typical development implies the simultaneous acquisition of well organized linguistic structures and motor coordinative commands (e.g., Nip et al., 2011). Children with language disorders thus tend to show decreased mastery of the articulators (e.g., Alcock et al., 2000) and children with CAS tend to show impaired production of long stretches of speech (Shriberg et al., 2003), impaired speech perception (Maassen et al., 2003) and a generally low phonological awareness, as revealed e.g., by difficulties in counting syllables and manipulating their phonological shape (Marquardt et al., 2002). This should be taken into account when studying articulation in this population; speech tasks that can successfully be accomplished by the children and with minimal involvement of other language abilities (such as speech comprehension or production of new words and sentences) should be used for elicitation.

### ***Vowels in Childhood Apraxia of Speech: variability and distinctiveness***

In speech development, it is generally agreed that vowels appear early (Donegan, 2013; Oller, 2000; Vihman, 1996; Vorperian & Kent, 2007) and accuracy of vowels approaches adult levels well before consonants (Pollock, 2002). Children with CAS are reported to experience difficulties producing vowels and vowel errors have been documented in various auditory and acoustic studies. For instance, Pollock and Hall (1991) revealed that children with CAS (N = 5) had problems with rhoticised vowels and diphthongs. The typology of errors was highly variable across sounds and included laxing, backing, lowering and tensing, diphthong reduction, vowel harmony and derhoticisation. Similar results were obtained in a longitudinal study investigating the connected speech of three speakers with CAS (Davis et al., 2005). These results showed that although speakers with CAS were able to produce almost all vowels, they did show a high number of vowel errors, resulting in much lower accuracy levels than what was expected for their age. Derhoticisation was among the most frequent types of errors. Since vowel errors (particularly derhoticisation) are also present in the speech of children with phonological disorders, it is unclear whether children with CAS differ from other clinical populations in this respect. An acoustic study by Blech et al. (2007) investigated three German children with suspected CAS and showed that they produced abnormal formant patterns compared to controls (children with no known disorder of speech), for both monophthongs and diphthongs and across a variety of speech conditions (isolated vowels, syllables, words).

Other studies have documented higher than normal acoustic variability in the production of vocalic targets. For instance, mean coefficients of variation for F1 and F2 were nearly

twice as great for a child with CAS than for typically developing and phonologically delayed children (B. Smith et al., 1994). The authors suggested that increased variability in multiple repetitions of vowels, in isolation as well as in real words, could be a potential diagnostic marker of CAS. In their study, Nijland et al. (2002) showed that a group of nine Dutch children with CAS displayed more variable F2 frequencies and less F2 distinction at the mid-point of /i/ and /u/, compared to control peers and adults, during a pseudo-words repetition task. Other studies documented impairments in appropriately producing temporal distinctions. For instance, children with CAS were found to make vowels longer, relative to normally speaking children (Nijland et al., 2003; N = , p. 6) and to reduce the temporal distinction between tense and lax vowels (Peter & Stoel-Gammon, 2005; N = , p. 2).

Most studies targeting vowel production in the speech of children with CAS have been focused on English or German. Considering that vowels are rarely reported as being impaired in speech disorders, Jacks et al. (2013) suggested that difficulties in the production of vowels should be central to the diagnosis of CAS. Their review of vowel characteristics in studies on CAS showed a variety of vowel errors, including in particular errors in the production of rhotic vowels, diphthongs and the tense/lax distinction. Given this background, it is useful to broaden the investigation to include non-Germanic languages, where vowel spaces are typologically different. As a matter of fact, in non-German languages vowel spaces tend to be overall less crowded, diphthongs tend to be more distinct from one another, rhoticity and the tense/lax distinction tend to be less essential as phonetic and phonological features. Once vowel impairments will be documented for speakers with such radically different vowel systems, the importance of taking into account vowels in formal diagnosis will show up with even more evidence.

The current study is an attempt at filling this gap with detailed phonetic data on the production of a subset of vowels by children with CAS who are native speakers of Italian, a language with seven vowel phonemes (/i e ε a ɔ o u/) in stressed syllables (the number reduces to five in unstressed syllables, where the distinction between mid-high and mid-low is neutralized and only /e/, /o/ occur). In particular, this study addresses the questions of variability and distinctiveness in vowel production. These two questions are related to one another as it is apparent from the consideration that increased token-to-token variability, as documented for CAS and for other speech motor and phonological disorders, may have an impact on the quality of the distinctions between sounds and on the overall structuring of the sound system the speakers rely on during speech production. So the question of token-to-token variability and that of phonetic space distinctiveness are two strictly related questions.

In order to address the overall issue, we first relied on perceptual analysis only to exclude distorted or substituted phonemic targets; then, we opted for a more fine-grained level of analysis focused on the instrumental detection of acoustic and articulatory variation across multiple repetitions of correctly produced phonemic targets. The general hypothesis was that an increased instability of articulatory movements could be the source of inconsistent productions of phonemic targets in the speech of children affected by CAS.

### ***Acoustic-articulatory analysis of speech variability in disordered speech***

Instability of movement patterns can be defined as repeated productions of words or phonetic sequences that are different in the absence of contextual variation (Lenoci &

Ricci, 2018; Miller, 1992; Smith, 2006; A. Smith et al., 2000). There is increasing evidence that children with speech sound disorders (including CAS) exhibit less mature speech motor patterns, in particular for jaw and lip coordination, compared to control peers (Grigos et al., 2010; Grigos & Kolenda, 2010; Grigos et al., 2015, 2011; Moss & Grigos, 2012; Terband et al., 2011). Furthermore, various studies have documented a reduction of vowel space areas across a variety of speech disorders; it is not uncommon for persons with hearing impairment, dysarthria or apraxia to produce speech movements that are reduced in size or amplitude (Kent & Rosenbek, 1983; Vorperian & Kent, 2007; Higgins & Hodge, 2002; see Ciocca & Whitehill, 2013 for a review) and in some cases, there is a clear correlation between the reduction of the acoustic space and a lowered degree of intelligibility. As a matter of fact, target articulatory undershoot may be at the origin of a reduction of sound distinctions in vowel systems, as indexed by acoustic measures such as vowel space areas (Hardcastle & Tjaden, 2008).

The current study focuses in particular on the production of several repetitions of lexically stressed /a/, /i/ and /u/. It proposes a multi-level phonetic approach, according to which both vocal tract resonances (vowel formants) and tongue displacements along the midsagittal plane (tongue imaging data) are analysed in order to determine the extent of token-to-token variability and vowel space distinctiveness in both dimensions of acoustic output and physiological dynamics at the tongue level.

Vowels /i/, /a/ (or /ɑ/) and /u/ can be described as corner or point vowels and their vocal tract resonances are used in the literature to map out the acoustic boundaries of vowel articulation of a given speaker (Hixon et al., 2008; Kuhl et al., 1997). The emergence of corner vowels in infants and young children is studied to determine the consequences of anatomical changes on the acoustic structure of speech sounds (e.g., Rvachew et al., 2008; Vorperian & Kent, 2007). Corner vowels correspond to the extremes of tongue height and tongue advancement parameters that shape vowel patterns across languages. In Italian and various other languages, lip configuration is an additional parameter mostly correlated to the tongue advancement parameter, to the extent that back vowels (and particularly [u]) are produced with lip protrusion and rounding.

Corner vowels have been said to be ‘quantal’ vowels since Stevens (1989) because they tend to remain acoustically more stable across a wider range of articulatory variation, compared to non-point vowels. However, the relationship between articulatory variation and acoustic output is far from linear. This is shown by a range of documented phenomena such as the influence of  $f_0$  on vowel formant structure (e.g., Carrell et al., 1981), effects of anatomical asymmetry between the front and the back portions of the oral cavity (e.g., Stone, 2010) or the articulatory adjustments realized to compensate for changes in a given dimension (‘motor equivalence’, e.g., Perkell, 1997).

The non-linear relationship between acoustic and articulatory changes is a crucial issue particularly in the domain of speech motor control anomalies, where the speakers’ vocal tract does not achieve the articulatory targets with the necessary precision and consistency. In that case, besides acoustic analysis, instrumental investigation of articulatory dynamics proves essential to determine the maturity of the speakers’ motor control system in a way that is independent from the acoustic-auditory output (e.g., Howard & Heselwood, 2013). Since we know that the same acoustic output can be achieved with different articulatory configurations, it may be important to rely on an independent estimation of articulators’

variability and acoustic variability, particularly in the case of pathological speech. For this reason, the integration of acoustic data with physiological data is an advantage in speech developmental research (Vorperian & Kent, 2007).

Among the various physiological techniques used in speech production research, Ultrasound Tongue Imaging (UTI) has been successfully used to explore lingual activity in vowel production by adults and children because it allows mapping the cross-sectional shape of the vocal tract from the tongue tip to the tongue back and root, thus providing very clear and complete data on the dynamics of speech sound lingual production. Compared to more invasive techniques such as Electromagnetic Articulography (EMA), the UTI equipment has the advantage of being portable and more easily tolerated by speakers, including children, which explains its success also in the domain of speech rehabilitation (e.g., Bernhardt et al., 2005; Cleland et al., 2015), including in the case of CAS (Preston et al., 2013, 2016). UTI can be successfully used for tracing and analysing tongue contours in the production of all lingual sounds, including comparisons of corner vowels (e.g., Chen & Lin 2011), and to provide a statistically reliable measure of the difference existing between pairs of curves (Davidson, 2006). UTI features have been used to detect covert contrasts in the pathological speech of children, that is, to show that auditorily indistinguishable productions of different phonemes were produced with different lingual configurations (e.g., Zharkova, 2017). To the best of our knowledge, the current study is the first application of a UTI-based analysis to CAS speech, for which we detect tongue movements and evaluate the degree of correspondence between the acoustic and the lingual correlates of vowel production.

### ***Non-contextual variability***

In assessing token-to-token variability across vowels in native Italian speakers with CAS, we also considered the issue of non-contextual variation. In non-contextual variation, when the phonetic context and the speaker are kept constant, vowels have been observed to vary across multiple repetitions in a way which is dependent on their quality and particularly their height (e.g., Pisoni, 1980 on English; Recasens & Espinosa, 2006 on Catalan). More specifically, /i/ shows reduced token-to-token variability with respect to the first formant (F1) but a much higher degree of variability with respect to the second formant (F2), compared to both /u/ and /a/. By contrast, low vowels such as /a/ are more susceptible than high vowels to variation along the F1 dimension, at the same time showing a smaller amount of variation along the F2 dimension. This effect should be a consequence of the fact that for the production of /i/, a more accurate tongue raising gesture is needed, whereas the precision of the constriction along the anteroposteriority dimension is less essential – and vice-versa for the vowel /a/. However, the question arises of whether this general pattern is also observed in the context of increased token-to-token variability levels, as those assumed for CAS speech. If it does, then the increased amount of token-to-token variability under pathological conditions may still be said to follow common patterns of vowel-specific variability levels. If, however, it does not, we will have to conclude that variability in CAS speech is not only quantitatively, but also qualitatively different from variability in non-pathological speech. We reasoned that it was important to address the issue of variability in CAS speech also from this perspective, in order to ascertain if perturbations of those speech



motor commands that are assumed to characterise CAS speech allow maintaining the most general patterns of articulatory variation that are found in normative speech or, rather, the latter are also perturbed.

### **Summary and experimental hypotheses**

The current study contributes with acoustic and articulatory (tongue imaging) data to the identification of features characterizing CAS speech. As speech inconsistency is considered one of the most relevant features of CAS speech, token-to-token variability is explored. The area of research is that of vowel production, a relatively less studied domain in the speech pathology literature, but still important to ascertain if increased token-to-token variability is associated with a reduction in the distinctiveness of the contrasts among phonemes. Moreover, the children are native speakers of Italian, a language with a much less crowded vowel system compared to English, and therefore potentially more subject to reductions of the vowel phonetic space. We provide instrumental data about phoneme variability and distinctiveness and compare the acoustic with the articulatory (tongue movements) dimensions of variability and distinctiveness. To better specify the nature and extent of token-to-token variability in the speech of our subjects with CAS, we also analysed their non-contextual variability and ascertained whether these patterns were consistent with those reported in the phonetic literature for non-pathological speech.

We test three major hypotheses. The first hypothesis concerns token-to-token (or within-category) variability, and can be divided into two sub-hypotheses. According to hypothesis (1a), based on the literature on CAS speech reviewed above, token-to-token variability is expected to be higher in the speech of children with CAS compared to typically developing peers. According to hypothesis (1b), F1 variability is expected to be higher for/a/ than for/i, u/ and F2 variability is expected to be higher for/i/ than for/a, u/ in the speech of control speakers. Children with CAS are expected to show deviant patterns of non-contextual variation (either qualitatively, quantitatively, or both), under the hypothesis that increased levels of within-category variability alter the normal patterns of non-contextual variability.

The third hypothesis of the study concerns vowel distinctiveness and the dimension of vowel space areas. Based on the literature reviewed above, the hypothesis predicts that vowels are less distinct from each other in CAS speech than in the speech of typically developing peers. The resulting vowel space area is therefore expected to be smaller in the former speakers' group than in the latter. The two aspects (phonemic distinctiveness and vowel space areas) are seen as related to each other since increased differentiation between vowel categories implies that the resulting vowel space area is larger.

## **Methodology**

### **Participants**

The study tested 10 native Tuscan Italian children aged 9–12 years and matched for age, sex and dialectal area. Five of them were affected by CAS (ages from cas01 to cas05: 10;5, 10;2, 10;4, 8;10, 12;1; mean age: 10;4) according to formal diagnosis by a multidisciplinary team at IRCCS Stella Maris, Pisa (Italy), in accordance with the three ASHA criteria: 1) inconsistent errors on consonants and vowels in repeated productions of syllables or words; 2) lengthened

and disrupted coarticulatory transitions between sounds and syllables; 3) inappropriate prosody, especially in the realization of lexical or phrasal stress (American Speech-Language-Hearing Association, 2007a). All children with CAS underwent a comprehensive assessment that included neurological and speech and language evaluation. The identification of the diagnostic features was based on perceptual analysis performed by two independent observers on each child's speech samples, which were video recorded during spontaneous interaction and the administration of formal tests. These children's expressive language was poor and they were therefore considered to be affected by CAS and comorbid language impairment. According to previous studies, based on a variety of experimental methods, token-to-token variability (as assessed in terms of phoneme substitutions or distortions in multisyllabic speech stimuli) is not statistically different in children with CAS compared to children with CAS *and* language impairment (e.g., Iuzzini-Seigel et al., 2016). We therefore considered our speaker sample as a sufficiently good representative sample for the population of children affected by CAS. Oral motor skills were examined through the VMPAC (Hayden & Square, 1999); for most of the children, the speech pathologist reported intermittently visible compensatory movements, such as lip rounding that might affect tongue movements.

The other five children were typically developing peers (henceforth, TYP; ages from typ01 to typ05: 10;3, 10;5, 10;2, 9;3, 12;5; mean age 10;5). They were reported by parents to have no hearing or language disorders. TYP and CAS were matched by age, geographic origin (all were native speakers of a Tuscan variety of Italian), and sex (both groups including four males and one female). All children were tested on separate days. Acoustic and kinematic data collection was carried out at the experimental linguistics laboratory of SMART lab, Scuola Normale Superiore, Pisa. The research protocols were approved by the Ethical Committee of Regione Toscana; written informed consent was obtained from all parents and children before the recording sessions.

### **Speech materials**

Three nonwords, distinguished by the quality of the stressed medial vowel, were used: these were *Adaba* [a'daba], *Adiba* [a'diba] and *Aduba* [a'duba]. The children in the study were told that each nonword was the proper name of one of three pets, whose drawings were shown to the children before and during the test phase on a computer screen. Each nonword was embedded in a short carrier sentence formed by the name and a disyllabic paroxytone verbal form, such as *Adaba scende* 'Adaba goes down'.

We opted for proper nouns (= nonwords) in order to keep the phonetic context as controlled as possible across stimuli. Although multisyllabic words are generally produced with more errors by children with CAS than monosyllabic words (e.g., Davis et al., 1998), our stimuli were multisyllabic in order to sound sufficiently plausible to native Italian speakers as proper names of pets. On the other hand, the phonological and prosodic structure of the nonwords was kept as simple as possible in order to prevent production difficulties by children with CAS, by making particular reference to the specific recommendations provided by the speech pathologist. Given that nonwords may be more difficult to produce compared to real words for children with language disorders (e.g., Archibald & Gathercole, 2007), all the children in the current study familiarized with the pets' pictures, their stories and their names for at least 15 minutes before executing the test.



## **Procedure and equipment**

A sentence repetition task with pre-recorded auditory stimuli was used to elicit the production of the target items. The repetition task was chosen as it is generally considered a low language load task (e.g., Baddeley et al. 2009) given that it does not involve higher order language abilities (such as semantic and grammatical processes) compared to e.g., picture naming. As a matter of fact, imitation was found not to induce increased variability in children with language impairments compared to typically developing children (Vuolo & Goffman, 2018). The children were asked to sit in front of a computer; whenever a drawing appeared on the computer screen, an adult female voice was played saying the target sentence, which was a description of the scene. The child had to repeat the sentence, and then the experimenter pressed a button and skipped to the following drawing/sentence. Each nonword was repeated eight times non-consecutively by each child at a self-paced rate, each time in association with a different verb. A total amount of 240 vowel tokens (3 vowels  $\times$  8 repetitions  $\times$  10 speakers) were collected for the current study. Audio and tongue profiles were recorded via a Mindray UTI system (sampling frequency: 30 Hz) associated to a 65EC10EA microconvex transducer and a Shure unidirectional microphone. Audio synchronization was guaranteed by a Synch Bright-up unit. The participants wore a head stabilization unit (helmet, Articulate Instruments Ltd, 2008) to ensure that the UTI transducer was kept in a stable position during the test and was oriented to display a mid-sagittal configuration of the tongue. The recording session lasted about 25 minutes.

## **Annotation**

All fluent/dV/syllables were segmented and labelled using Praat (version 6.0.15 by Boersma & Weenink, 2016). Three intervals were annotated for each syllable, i.e. the consonantal closure, the VOT and the vowel; for the current study, only the vowel interval was analysed. Vowel boundaries were manually identified based on visible formant structure. The annotations were then imported into the AAA software (version 2.14 by Articulate Instruments Ltd, 2012). The acoustic midpoint of the vowel and its closest ultrasound frame were identified for each repetition of each vowel. Tokens that were produced as incorrect were excluded from the analysis. More specifically, this happened for 6 stimuli which were produced by children with CAS with a velar instead of an alveolar stop (phoneme substitution). Although the vowel was correctly produced (perceptual rating), the stimuli were excluded in order to exclude variability induced by the phonetic context. All other vowels were produced correctly and with correct stress, and were therefore included. Children from the TYP group did not produce any perceptible errors in their production and so we did not exclude any items from their repetitions.

## **Acoustic analysis**

F1, F2 and F3 values were automatically extracted at vowel acoustic midpoint via a Praat script. Maximum formant range was set at 8000 Hz and the number of formants was set at 5. The values extracted from the script were checked through manual correction by calculating F1 and F2 values with the Fast Fourier Transform at the same temporal instant that was considered by the script. For each selected syllable, we measure F1 and F2 from narrow-band spectral cross-sections. Window lengths from which the narrowband spectra were generated were 30 ms.

F1 and F2 mean (in Hz) and standard deviation values were calculated for each vowel uttered by each speaker. Then formant estimates were converted to Equivalent Rectangular Bandwidth (ERB) values, with the Traunmuller's formula (Traunmuller, 1997), to reduce the effect of anatomical differences across participants and median F1 and F2 ERB values were calculated for each vowel uttered by each talker. The ERB normalization method transforms formant values to be similar to those perceived by the auditory system in relative terms, and was chosen over other available methods for not necessarily requiring a full set of vowels for the transformation (Calamai, 2006; Pettinato et al., 2016). Based on normalized median values, the vowel space area (VSA) of each speaker and the Euclidean distance between each pair of vowels as produced by each speaker were calculated. In particular, the VSA was calculated for each speaker according to Heron's formula (see Sapir et al., 2010); the Euclidean distances between /a/and/i/, /a/and/u/, and /i/and/u/, were calculated for each speaker according to the methodology of Pettinato et al. (2016).

Token-to-token variability in the acoustic domain was assessed through inspection of standard deviation values for each vowel as produced by each speaker (Recasens & Espinosa, 2006, p. 662). Vowel space distinctiveness was assessed by calculation of VSAs and Euclidean distances, again on a speaker-by-speaker basis.

### **Articulatory analysis**

For the UTI frame which most closely corresponded to the acoustic midpoint of the vowel, a tongue contour fitting procedure was implemented semi-automatically with the help of a tongue profile tracker in AAA followed by manual correction. All the fitted tongue contours were then exported as a series of (x,y) coordinates into the RStudio software (version 1.2.1335) for the computation of both token-to-token variability and vowel space distinctiveness.

Token-to-token variability in the articulatory domain was assessed by means of the Nearest Neighbor Distances metric (Zharkova & Hewlett, 2009). This metric is used to quantify the overall difference between a pair of tongue curves, based on the mean of all the Euclidean distances between each point on one curve and its nearest neighbour on the other. According to Zharkova et al. (2011) and Frisch et al. (2016), the average distance between curves corresponding to different repetitions of the same vowel context (henceforth, within-set distances: WS) can be interpreted as a measure of speech motor stability for individual speakers: the greater the separation between curves belonging to the same category, the higher the variability of tongue movements across repetitions.

Vowel space distinctiveness was assessed by comparing the tongue profiles of /a/, /i/and/ u/ of each speaker through smoothing spline analysis of variance (SSANOVA; Davidson, 2006) and Bayesian confidence intervals (using the *ssanova* function in the *gss* package). This technique allows determining significant differences between the smoothing splines that are the best fits for three curve sets (one for each vowel). We expected to find significant interaction terms of the models for each speakers of both groups (the main group effects corresponding to the smoothing splines for each vowel dataset), under the assumption that tongue shapes for the production of /a/, /i/and/u/ were significantly different and did produce an articulatory contrast. Visual inspection of the plots showing smoothing spline estimates and 95% confidence intervals allowed comparing the different realization of tongue shape contrasts across subjects. To further determine if there were specific sections of the curves

(corresponding to specific portions of the tongue) in which children with CAS produced the contrast differently from the typically developing peers, interaction effects with 95% Bayesian confidence intervals were also plotted.

## Results

### Vowel variability

#### Formant data

Before analysing token-to-token variability in details, we inspected mean formant data of our two groups of speakers, in order to determine at least on a very general scale if they were consistent with reference data for the Italian language.

Mean (in Hz) and standard deviation values for /a/, /i/ and /u/ as produced by five groups of Italian speakers are thus provided in Table 1. TYP and CAS refer to the two groups of children of the current study. In addition, reference data for 10 adult male and 10 female speakers of Italian and for typically developing children are provided in the Table. The adult and children data were taken from Zmarich and Bonifacio (2003). The latter data were collected from a group of about 70 children (the exact number of children changed across vowels) aged 8–10 and coming from a Northern Italian city, Padua (Ferrero et al., 1996). Although there are differences across Italian dialects as far as typical vowel quality is concerned, these are generally said to pertain to the domain of mid vowels only, leaving /a/, /i/ and /u/ unaffected by geographic variation (Albano Leoni & Maturi, 2003). The speech elicitation task in the case of the children was the same as the one used in the current study (repetition of words following an adult model) whereas adult speech was collected through a word reading task.

In Table 1, all children regardless of group showed higher F1 and F2 mean values than adults, confirming results from past literature about a gradual decrease of formant frequencies with age (e.g., Peterson & Barney, 1952; Vorperian & Kent, 2007). The group with CAS

**Table 1.** Mean and standard deviation values for each of the three vowels/ a/, /i/ and /u/ as produced by the CAS and TYP children of the study, two groups of adult Italian speakers (males and females) and a group of typically developing Italian children. Details about the source data are given in the text.

		F1 mean Hz (s.d.)	F2 mean Hz (s.d.)
/a/	TYP	1043 (47)	1723 (67)
	CAS	962 (70)	1655 (99)
	Male adult	715 (32)	1274 (46)
	Female adult	898 (65)	1467 (95)
	Children	1046 (59)	1737 (108)
/i/	TYP	372 (23)	3039 (75)
	CAS	415 (35)	2948 (109)
	Male adult	273 (11)	2111 (127)
	Female adult	312 (16)	2524 (107)
	Children	342 (46)	3135 (121)
/u/	TYP	497 (22)	1016 (50)
	CAS	495 (34)	997 (73)
	Male adult	269 (31)	649 (98)
	Female adult	291 (26)	684 (83)
	Children	416 (46)	859 (57)

showed a higher/a/(lower F1 values) and a less anterior/i/(lower F2 values) compared to both TYP (current study) and the children of the Zmarich and Bonifacio (2003) study. Moreover, the children with CAS showed a less posterior/u/(higher F2 values) compared to the children of that study. Although these were raw, non-normalized data and can only be considered with great caution, they nevertheless seem to indicate that the vowels in our speakers with CAS tended to occupy less peripheral positions: low vowels were less low than expected, and at the same time, high vowels were not as low as in the speech of children with typical development.

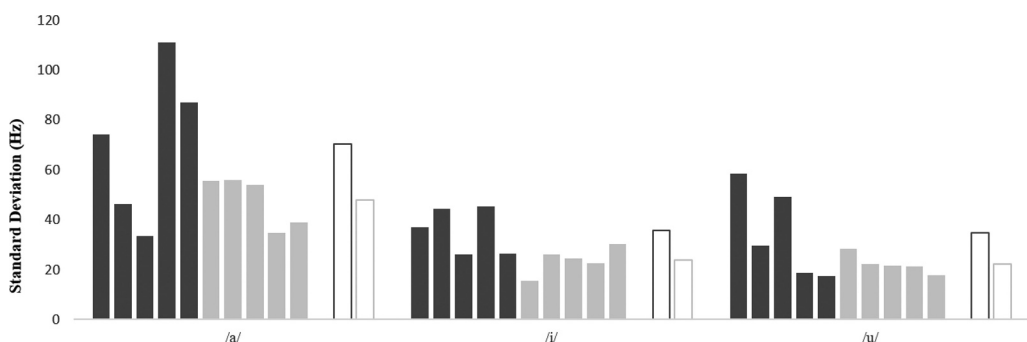
We now turn to the analysis of the more central issue of token-to-token variability. Figures 1 and 2 show the formants' standard deviation values for each vowel of each child of this study; average values by group are also provided for reference.

In particular, Figure 1 shows that the standard deviation values for F1 were higher for the CAS group compared to the TYP group in the production of all three vowels. Individual values were not always consistent with the group values, to the extent that some speakers with CAS could show lower standard deviation values compared to some TYP speakers.

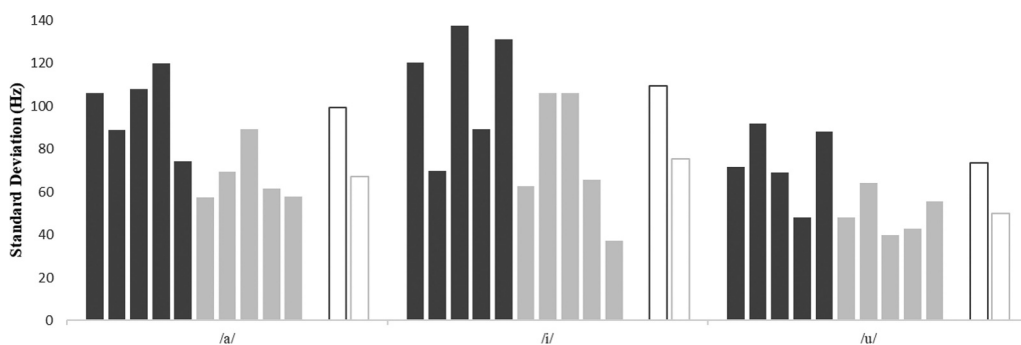
Still with respect to the first formant, standard deviations were overall higher for/a/ than for the two other vowels, thus suggesting that there was more variability in the acoustic correlate of vowel height for the low vowel, compared to/i/and/u/, in the production of both groups of speakers. At the individual level, the pattern was found for all five TYP speakers but for only three of the speakers with CAS (cas01, cas04, cas05). The two other speakers with CAS, i.e. cas02 and cas03, showed a rather low degree of variability for/a/; standard deviation values for this vowel were close to (or even lower than) those found in TYP speech; at the same time, they did not show any decrease of F1 variability in the production of/i/and/u/.

F1 acoustic evidence thus suggested that high vowels were more constrained than/a/for tongue height and constriction location for all TYP speakers but for only three of the speakers with CAS.

Concerning the second formant, Figure 2 shows that the standard deviation values for F2 were higher for the CAS group compared to the TYP group. The pattern was found for all three vowels at the group level. The differences between the three vowels were not as strong as in the case of F1; nevertheless, the average standard deviation values were higher for/i/



**Figure 1.** F1 standard deviation values as a function of vowel and speaker. Dark grey = CAS speakers (from cas01 to cas05, left to right), light grey = TYP speakers (from typ01 to typ05, left to right). Average group values (white bars with dark and light grey lines for, respectively, CAS and TYP groups) are also reported for each vowel.



**Figure 2.** F2 standard deviation values as a function of vowel and speaker. Dark grey = CAS speakers (from cas01 to cas05, left to right), light grey = TYP speakers (from typ01 to typ05, left to right). Average group values (white bars with dark and light grey lines for, respectively, CAS and TYP groups) are also reported for each vowel.

and/a/compared to/u/for both the CAS and the TYP group. To be precise, the values were very slightly higher for/i/than for/a/; however, the difference was a very small one.

At the individual level, as in the case of F1, there was variability; some speakers with CAS could show lower standard deviation values compared to some TYP speakers. Recall that according to the English and Catalan data reviewed above (Recasens & Espinosa, 2006),/i/ was expected to be the *least* constrained vowel for tongue posteriority, that is, the most tolerant vowel with respect to token-to-token F2 excursions. In our data, it seemed that both/a/and/i/turned out to be less constrained than/u/in the majority of the speakers. In particular, there was a tendency for/u/to show less variable F2 values compared to both/a/and/i/in four TYP and three speakers with CAS. By contrast, in one TYP (i.e. typ05) and one speaker with CAS (i.e. cas02), the pattern was exactly the opposite of the expected one, with vowel/i/showing the least variation.

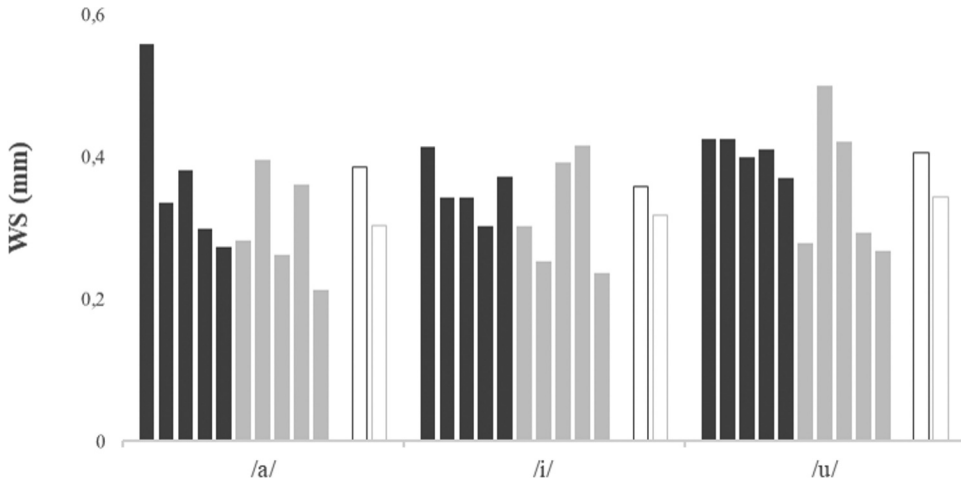
In conclusion, as far as F2 variability was concerned, the expected pattern was not as clearly represented in our data as it was in the case of F1 variability (see above). Moreover, we found that speakers with CAS were only slightly different from TYP speakers with respect to the amount of F2 variability across vowels.

### **Lingual profile data: Within-set distances (WS)**

Token-to-token variability in the articulatory domain was estimated through an analysis of the average distance between lingual curves belonging to the same vowel context (WS distances). In Figure 3, greater articulatory variability across repetitions is reflected in higher bars.

Articulatory variability was found to be higher in CAS speech compared to TYP speech for all three vowels ( $t(367) = 5.603, p < .001$ ). The difference was bigger in the case of/a/( $t(114) = 3.007, p = .003$ ) and/u/( $t(118) = 3.582, p < .001$ ) and a bit smaller, but still significant, in the case of/i/( $t(125) = 2.510, p = .013$ ).

At the individual level we found more variable behaviours. In the case of vowels/a/and/i/, at least two TYP speakers showed a higher degree of articulatory variability compared to the majority of speakers with CAS; in the case of/u/, one TYP speaker showed much more variability than all speakers with CAS and another TYP speaker showed values comparable to those of the majority of the speakers with CAS.



**Figure 3.** WS distance values as a function of vowel and speaker. Dark grey = CAS speakers (from cas01 to cas05, left to right), light grey = TYP speakers (from typ01 to typ05, left to right). Average group values (white bars with dark and light grey lines for, respectively, CAS and TYP groups) are also reported for each vowel.

**Vowel contrast**

**Euclidean distances and vowel space areas (VSAs)**

The Euclidean distances between/a/and/i/and between/a/and/u/were smaller in CAS compared to TYP speech (/a/-/i/: 7.68 vs 8.24 ERB;/a/-/u/: 6.24 vs 6.81 ERB). This suggested that the vowel space of speakers with CAS showed less distinction as far as the height dimension was concerned. In contrast, the Euclidean distance between the two high vowels/u/and/i/ was slightly larger for CAS (8.83 ERB) than for TYP speech (8.51 ERB).

VSAs were further analysed to determine the overall shape and extension of the acoustic space delimited by the three corner vowels. Individual values and group means are shown in Table 2; F1× F2 plots for each speaker are provided in Figure 4.

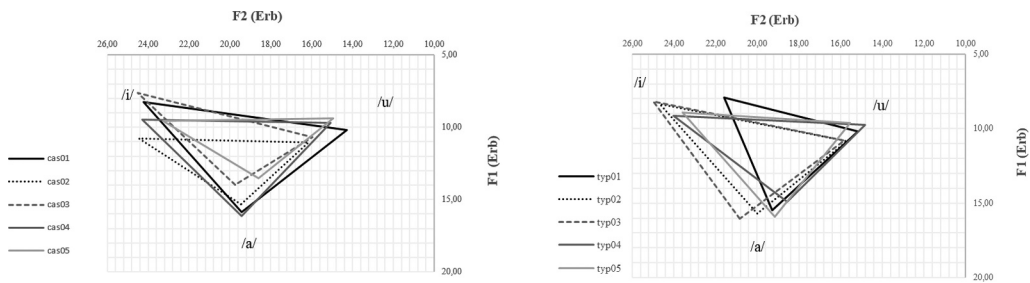
Data in Table 2 show that the acoustic vowel space of speakers with CAS was on average smaller than that of TYP speakers. This suggested that distinctions across vowels were overall less consistent, or less sharp, in the former population compared to the latter, although the Euclidean distances pointed to a reduction in only one of the two acoustic dimensions. However, as the individual data in Table 2 reveal, there was much cross-subject variation. For instance, cas01 produced the largest VSA in the total sample. Nevertheless, three speakers with CAS showed very low (< 20) VSA values, thus lower than any TYP value.

F1x F2 plots in Figure 4 show that speakers with CAS were extremely variable within group as far as vowel height was concerned; for most of them, height distinctions were reduced. This could happen according to three different strategies: either raising the/a/(as in cas03, whose/a/ had an F1 of 13.99 ERB), or lowering one or both high vowels (as in cas02 or cas04; for

**Table 2.** VSA values for each speaker and mean values for groups.

CAS	cas01	cas02	cas03	cas04	cas05	Mean
TYP	typ01	typ02	typ03	typ04	typ05	Mean
	33.33	18.25	18.89	30.18	16.97	23.41
	21.56	27.08	30.49	24.82	26.63	26.04





**Figure 4.** Normalised F1× F2 VSAs for CAS (left panel) and TYP children (right panel).

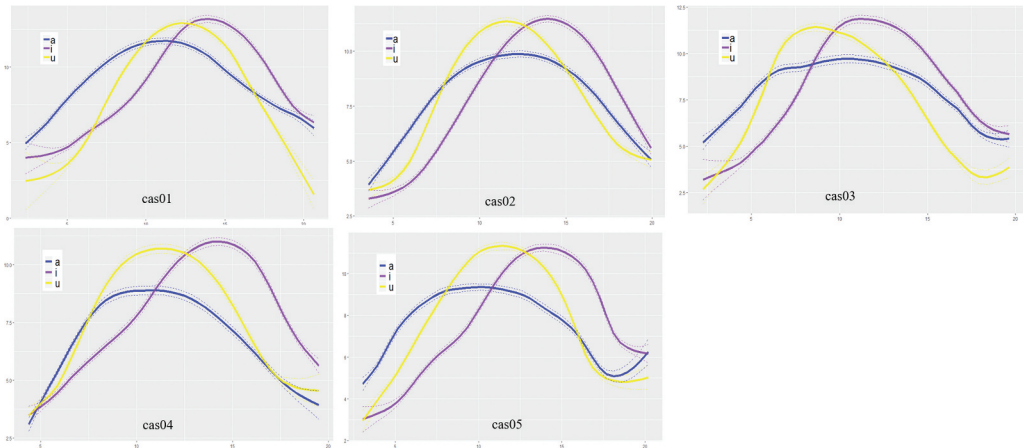
instance, F1 was 10.78 ERB for /i/ and 11.06 for /u/ in cas02 speech), or even raising the /a/ and lowering the high vowel(s) at the same time (as in cas05). Considering F1, /u/ was less variable across subjects than /a/ and /i/ (ranging from an 11.06 ERB in cas02 to 9.39 ERB in cas05).

By contrast, F1 targets were much less variable across TYP speakers. All three vowels varied approximately to the same extent. The subject with the smallest height distinctions was typ04. This speaker also showed the most posterior /a/ vowel, with an F2 value of 18.57 ERB. The most anterior /a/ was produced by typ03 (F2: 20.84 ERB). There was more cross-subject variability along the F2 axis in the /a/ of the TYP group, compared to the /a/ of the CAS group. Extremely variable along the anteroposteriority dimension was also /i/ in the TYP group, although variation here was mostly due to one subject, namely, typ01 (F2: 21.58 ERB). The realization of /i/ by typ01 probably explains why the /i/-/u/ mean Euclidean distance was found to be smaller for TYP speakers than for speakers with CAS (see above). As a matter of fact, when typ01 was excluded, TYP speakers were not highly variable within group as far as the F2 targets of /i/ and /u/ were concerned. By contrast, there was more cross-subject variation in the CAS group, especially for /u/; F2 values varied from very posterior /u/ in cas01 (14.26 ERB) to more anterior /u/ in cas02.

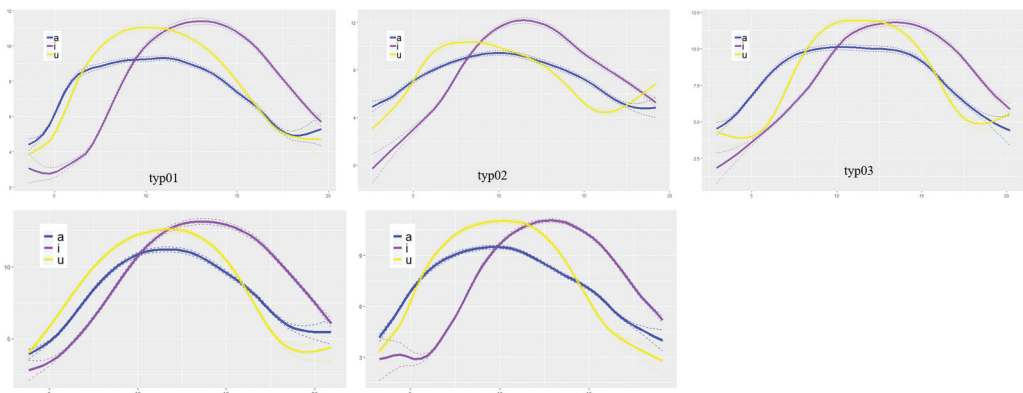
To sum up, VSAs showed that, as a group, speakers with CAS produced more centralized corner vowels than TYP speakers. Centralization was primarily the consequence of concomitant /i/ and /u/ lowering and /a/ raising, but it was also achieved by /u/ fronting. Centralization was reflected in a more crowded central area in Figure 4, left panel. TYP speakers showed more distinct vowel targets and an uncrowded central area in Figure 4, right panel, especially when typ01 abnormally posterior /i/ was excluded.

### **Lingual profile data: SSANOVAs**

SSANOVA provides statistical models for determining whether or not there are significant differences between the smoothing splines that are the best fits for data sets being compared. Figures 5 and 6 show the estimated mean tongue profiles for each CAS and TYP speaker, respectively, during the reiterated production of /a/, /i/ and /u/; the shadowed area along the mean profile represents the standard error. Figure 7 additionally shows the interaction effects with Bayesian confidence intervals for two of the subjects, one from each group (i.e. cas03 and typ05), basically visualizing the contribution of each vowel to the overall three-vowel model of each subject (Davidson, 2006).



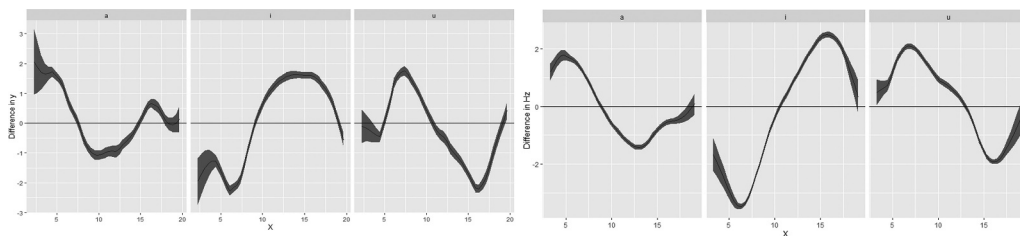
**Figure 5.** Smoothing splines of tongue shapes in the production of CAS speakers.



**Figure 6.** Smoothing splines of tongue shapes in the production of TYP speakers.

In the CAS group, three of the speakers (i.e. cas01, cas04 and cas05) showed an extremely fronted position of the tongue in the production of the /u/ vowel, as revealed by the fact that the anterior part of the tongue is in a front position for /u/ than for /a/. In the TYP group, this pattern was visible in only one speaker, typ01. Moreover, /u/ was produced with a more fronted tongue back than /a/ in four out of five speakers with CAS (the only exception being cas04), whereas this pattern was found in only one TYP speaker (namely, typ03). Thus, it is possible to hypothesize that /u/ is produced by speakers with CAS with a forward displacement of the tongue, that is visible in our data at both tongue back and tongue front. Considering that the neighboring consonants are alveolars it is possible to further hypothesize that children with CAS exhibit stronger effects of coarticulation on this vowel than on the other vowels.

Additionally, for three speakers with CAS (namely, cas01, cas02 and cas03) the vowel /a/ was articulated with a relatively high tongue dorsum, compared to the position of the tongue dorsum in vowel /i/. This is particularly evident from Figure 7, comparing cas03 (three boxes on the left) with typ05 (three boxes on the right). Consider that in each box, the



**Figure 7.** In each box, the tongue tip is on the right and the tongue back in on the left. When the confidence interval encompasses the zero on the y axis at any point along the interaction curve, there is no difference between the two curves being compared; the interaction at that point is not statistically significant (Davidson, 2006).

tongue tip is on the right and the tongue back is on the left. The figure shows that the contribution to the dimensions of height and fronting is entirely in charge of the /i/ vowel in the production of typ05 (the curve for /i/ was situated above the y axis with highest values at the rightmost positions), whereas there was evidence of tongue fronting and raising also with /a/ in the production of cas03.

## Discussion

The major aim of this study was to explore the acoustic and articulatory correlates of vowel variability and contrast in the speech of children with CAS compared to that of TYP peers. Since CAS has been defined as a speech motor disorder, we expected a less mature control on articulators responsible for speech production. Concerning variability, both token-to-token and non-contextual variability were analysed to ascertain whether children with CAS showed deviant patterns in vowel articulation and acoustics compared to controls. Contrast among vowels was evaluated in terms of vowel space area, again through an analysis of the acoustic output and of tongue articulatory movements as directly observed in UTI.

As a first and fundamental result, we found high levels of cross-subject variation in all measures and for both analyses of variability and distinctiveness. Individual variation was found not only in CAS speech, but also in the speech of typically developing children, consistently with repeated observations that children and adolescents are less stable in speech articulation compared to adults (Goffman & Smith, 1999; Smith & Zelaznik, 2004) and that an increased control of the articulators is a correlate of maturation. We did find group differences in the data (and in some of these cases, they were certified by statistical tests); however, it is important to notice that group differences should not be considered independently of individual variation in order not to overemphasize differences between children groups that are always modulated by the specific characteristics of each child. With this *caveat* in mind, we will now turn to the discussion of the obtained results. We will first comment on variability and then on contrast.

### Variability

Consistently with our initial hypothesis (1a), the group of speakers with CAS was found to produce vowels with a higher degree of variability compared to control peers. This finding

complement observations of inconsistency in CAS speech based on phoneme-level analysis (number of substituted or distorted phonemes, e.g., Iuzzini-Seigel et al., 2016; Lewis et al., 2004), providing evidence of fine-grained acoustic and articulatory variability in the production of one and the same phoneme. F1, F2 and WS distances were examined to assess the relative acoustic and articulatory stability in achieving vowel targets.

Acoustic data revealed that in spite of high levels of cross-speaker variability, mean standard deviation values for both F1 and F2 were higher for the CAS group compared to TYP group for all three vowels. The acoustic output was therefore less stable across multiple repetitions in CAS speech compared to TYP speech. Although this pattern was consistently found across vowels and formants, the individual speakers showed a large amount of cross-individual variation, with TYP speakers exhibiting sometimes higher variability scores than selected speakers with CAS. This circumstance is worth noticing considering that the speaker samples in clinical phonetics studies tend to be reduced for contingent reasons, sometimes with numbers that are inferior to those of the current study. By contrast, the analysis of variation would need large speaker samples. We thus have to conclude by acknowledging that, although the group patterns referred to above were clear and substantially in line with previous observations that can be found in the literature, it would be important to replicate the study on additional groups of speakers, both native speakers of Italian and of different languages, in order to corroborate the validity of the findings discussed here.

Again with respect to variability, the analysis also aimed to ascertain (hypothesis 1b) whether general patterns of non-contextual variation, as revealed by adult Catalan speakers in Recasens and Espinosa (2006), were also found in the speech of our Italian CAS and TYP children. According to Recasens and Espinosa (2006), speakers would show the most variable F1 values during the production of /a/ and the most variable F2 values during the production of /i/.

Our data confirmed that both groups of children showed higher levels of F1 variability for /a/; however, the effect was stronger and more generalized in TYP than CAS speech. This result thus suggested that token-to-token variability tended to be different in CAS speech, compared to TYP speech, not only quantitatively, but also qualitatively. In other words, speakers with CAS, when considered as a group, did not produce variable vowels in the same way as TYP speakers did, when phonetic context was kept constant. In particular, vowels were all equally variable as far as the acoustic output of tongue height was concerned; the high vowel /i/ was not less variable for tongue height than the other vowels, as it could have been predicted based on TYP speech as well as adult reference speech. Non-contextual variability in the case of F2 provided a slightly different picture. In that case, the data only partially reflected the predictions put forth at the beginning. In particular, both /i/ and /a/ were slightly more variable than /u/ in the speech of both groups. The standard deviation values for /i/ were only minimally higher than those of /a/, thus suggesting that the parameter of anteroposteriority was not straightforwardly more relevant for /u/ compared to /a/, nor for /a/ compared to /i/. Most importantly, CAS speech and TYP speech were not clearly divergent in this respect. So it should be concluded that, although CAS speech was both quantitatively and qualitatively different from TYP speech as far as variability along the height dimension was concerned, it was only quantitatively (but not qualitatively) different as far as variability along the anteroposteriority dimension was concerned.

The articulatory measurement of token-to-token variability (UTI-based NND distances) confirmed that speakers with CAS were overall more variable than TYP speakers in articulating all three vowels. Even in this case, subjects could differ from one another idiosyncratically, with selected TYP speakers showing more variable articulation than selected speakers with CAS for selected vowels. For instance, typ02 was highly variable in the production of both/a/and/u/, typ03 in the production of both/i/and/u/and typ04 in the production of both/a/and/i/. However, group differences were always statistically significant. This finding was not consistent with those in Vuolo and Goffman (2018), who found no statistically reliable differences between speakers of CAS and other groups of speakers (including typically developing children) as far as variability in the kinematics of speech movement was concerned. One possibility of such divergence may be related to the fact that Vuolo & Goffman's study focused on lip movements whereas the current study focused on tongue movements. Since lip movements are acquired relatively early in speech development (e.g., Stoel-Gammon & Cooper, 1984) and lips can be moved according to a smaller amount of degrees of freedom compared to tongue (e.g., Beautemps et al., 2001), it could be speculated that the intrinsic difficulty of lingual articulation for the production of vowels provided a more challenging speech target than the articulation of bilabial consonants of Vuolo & Goffman's study, thus allowing groups differences to emerge. However, in the lack of any systematic comparison, this hypothesis remains speculative.

### ***The acoustic-articulatory relationship***

Differently from the acoustic analysis, which evaluated one dimension at time (either height or anteroposteriority, as indirectly revealed by formant values), UTI-based articulatory evidence provided a global picture of tongue profile movements by considering all portions of the tongue midsagittal profile at one time, from the tongue tip to the dorsum and the back. Speakers with CAS thus turned out to have more variable tongue profiles across multiple repetitions of the same vowels than TYP speakers.

It is important to highlight the different nature of the acoustic and articulatory evidence, and the different conclusions that can be drawn from each of them. From the UTI-based articulatory evidence, we were told nothing about potentially different shapes and dimensions of resonant cavities (something that would have a differential impact on formant values instead). By contrast, we were told that global tongue shapes in speakers with CAS were less stable across repetitions than global tongue shapes in TYP speakers. Speakers with CAS were thus found to be more variable not only at the level of the acoustic output (which is naturally affected by a number of different articulatory settings) but also in terms of the measurable accuracy of their overall tongue movements.

### ***Distinctiveness***

The second purpose of the study was to verify whether vowel distinctiveness was reduced and the overall vowel space areas were smaller in CAS compared to TYP speakers, as suggested also by other scholars (e.g., Iuzzini & Forrest, 2008). Formant evidence confirmed that, as far as the acoustic vowel space was concerned, the hypothesis could be maintained only with reference to a general trend, but high levels of variation among individual speakers were also found. In particular, we found that one of the TYP speakers, namely

typ01, showed very low F2 values for/i/, which suggested a reduced distinction along the anteroposteriority dimension with respect to the two other vowels. However, speakers with CAS were consistent in showing generalized reduction of F1 distinctions, which suggested that vowels were more centralized and particularly less distinct along the height dimension. The articulatory evidence confirmed the critical nature of height distinctions for speakers with CAS, particularly as far as/a/is concerned; this vowel was not always produced with a flat tongue dorsum but rather implied some tongue dorsum activity in most of the speakers with CAS. Furthermore, the/i/vowel was not produced with the same raising gesture that was typical of TYP speakers; both the amount of the raising (as comparatively ascertained with respect to the other vowels, see [Figure 7](#)) and the portion of the tongue that was raised were different in speakers with CAS, compared to more canonical high vowel production strategies found in TYP speakers.

The articulatory evidence additionally suggested that/u/was not realized as a canonical back vowel by speakers with CAS. Both the tongue root and the tongue tip were extremely fronted for most of these speakers. Interestingly, this pattern did not emerge with the same clarity from the acoustic analysis; as a matter of fact, F2 values for/u/were in the same range for CAS and TYP speakers (see [Figure 4](#)). Therefore, this finding reinforces the view that acoustics and articulation are in no way isomorphic and that looking at the acoustic output does not necessarily provide detailed information about the articulatory manoeuvres that the speakers are putting in place to achieve a given acoustic output. In the particular case discussed here, it might be hypothesized that speakers with CAS compensate for insufficiently retracted tongue configuration by increasing lip rounding (which notoriously produces lowering of all formants) or other articulatory changes that have the effect of saving the acoustic identity of the back vowel. We should consider that lip protrusion/rounding is one of the earliest and simplest facial gestures that are realized by newborns in face imitation and interaction (e.g., Meltzoff & Moore, 1977; Stoel-Gammon & Cooper, 1984), so it is quite possible that children with speech impairment resort to this articulatory strategy to achieve an acoustic goal normally conveyed by complex tongue retraction mechanisms. As specified in the methodological section, the speech pathologists did indeed report evidence of special lip movements in children with CAS. However, in order to prove a role for lip rounding as a compensatory strategy for F2 lowering, additional testing and quantification will be needed.

## Conclusions

This study has shown variability and reduction of contrasts particularly in the height dimension, and non-linear relationships between acoustics and articulation. The latter suggested alternative ways of realizing contrasts in CAS speech. Moreover the study has shown that vowel production is an important domain to be looked at in CAS analysis, even for speakers of languages with less crowded vowel systems, such as Italian.

In this respect, the current study supports the observation that an accurate estimation of vowel production strategies might be useful information for the diagnosis of this speech motor control disorder (Jacks et al., 2013). The current study also has obvious limitations, starting with the size of the speaker sample and number of word tokens that were analysed. Future investigation will have to deal with additional parameters (such as vowel length and rounding) and segments (such as mid vowels



and diphthongs) in order to provide a clearer picture. Moreover, a comparison with the speech of children whose diagnosis is a phonological disorder will have to be included in future research; comparative data in that case will be useful for differential diagnosis. Our study did not include children with language impairment; therefore, it's difficult to exactly ascertain how much of the observed variability is due to CAS and how much to concomitant language impairment. It should be recalled that we carefully selected the quality of the speech materials and the experimental setup in order to be minimally challenging for our children (§2.2–2.3). Moreover, the literature reports no statistically relevant differences in the amount of variability between children with CAS and children with language impairment and CAS (§2.1; Iuzzini-Seigel et al., 2016). Nevertheless, the lack of a control group of children with language impairment remains an objective limitation of the current study.

For the time being, we found a tendency that, when compared to matched controls, speakers with CAS were more variable across multiple repetitions of the same stimulus; this variability encompassed both tongue movements when analysed through imaging techniques and oral resonances in the acoustic domain. Their variability was not only quantitatively more widespread, but also qualitatively different to the extent that high vowels were as variable as low vowels along the height dimension, whereas normal adult and child production would inhibit high vowels from varying consistently along the height dimension. When compared to matched controls, speakers with CAS also showed smaller vowel spaces and less distinct vowel contrasts, again in both articulation and acoustics. As in the domain of variability, also in that of contrast the height dimension seemed to be the most affected in CAS production, particularly at the acoustic level (F1 distinctions). However, the inspection of tongue movements also showed difficulties in the realization of tongue retraction for/u/. It was important to distinguish direct articulatory observation from acoustics-based estimation of vowel variation to the extent that this allowed visualizing differences in articulatory targets and deviant tongue configuration dynamics even when they had only slight consequences on formant values. Our speakers with CAS were likely to introduce alternative strategies for their disrupted tongue configuration dynamics and this kind of knowledge might be useful for clinical treatment.

## Acknowledgments

This work was funded by Fondazione Pisa, project “Speech motor disorders and visual feedback: 3D simulation of articulatory movements” granted to V. Barone (2017-2020).

## Disclosure of interest

The authors report no conflict of interest.

## Funding

This work was supported by the Fondazione Pisa [progr. n. 132/16].

## References

- Albano Leoni, F., & Maturi, P. (2003). *Manuale di fonetica*. Carocci (1a ed. Roma, NIS, 1995).
- Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and Language*, 75(1), 17–33. <https://doi.org/10.1006/brln.2000.2322>
- American Speech-Language-Hearing Association (2007a). *Childhood apraxia of speech* [technical report]. <http://www.asha.org/policy>.
- Arbib, M. A. (2006). The mirror system hypothesis on the linkage of action and languages. In M. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 3–47). Cambridge University Press.
- Archibald, L. M., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, 14(5), 919–924. <https://doi.org/10.3758/BF03194122>
- Articulate Instruments Ltd. (2008). *Ultrasound stabilisation headset user's manual: Revision 1.13*. Author.
- Articulate Instruments Ltd. (2012). *Articulate assistant advanced user guide: Version 2.14*.
- Baddley, A. D., Hitch, G. J. & Allen, R. J. (2009). Working memory and binding in sentence recall. In *Journal of Memory and Language*, 61, 438–456.
- Beautemps, D., Badin, P., & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *The Journal of the Acoustical Society of America*, 109(5), 2165–2180. <https://doi.org/10.1121/1.1361090>
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6/7), 605–617. <https://doi.org/10.1080/02699200500114028>
- Blech, A., Springer, L., & Kroger, B. (2007). Perceptual and acoustic analysis of vowel productions in words and pseudo words in children with suspected childhood apraxia of speech. In *International Congress of Phonetic Sciences, ICPhS, Saarbrücken, Germany*.
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*. <http://www.praat.org>
- Calamai, S. (2006). Intrinsic methods in vowel normalization: Data from Pisa and Florence Italian. In E. Miyares Bermúdez & L. Ruiz Miyares (Eds.), *Linguistics in the twenty first century* (pp. 113–124). Cambridge Scholars Press.
- Carrell, T., Smith, L., & Pisoni, D. (1981). Some perceptual dependencies in speeded classification of vowel colour and pitch. *Perception and Psychophysics*, 29(1), 1–10. <https://doi.org/10.3758/BF03198833>
- Chen, Y. & Lin, H. (2011). Analysing tongue shape and movement in vowel production using SS ANOVA in Ultrasound imaging. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*. 17–21 August 2011, Hong Kong.
- Chilosi, A., Lorenzini, I., Fiori, S., Graziosi, V., Rossi, G., Pasquariello, R., Cipriani, P., & Cioni, G. (2015). Behavioral and neurobiological correlates of childhood apraxia of speech in Italian children. *Brain and Language*, 150, 177–185. <https://doi.org/10.1016/j.bandl.2015.10.002>
- Ciocca, V., & Whitehill, T. L. (2013). The acoustic measurement of vowels. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 113–136). Psychology Press.
- Cleland, J., Scobbie, J. M., & Wrench, A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, 29(8–10), 575–597. <https://doi.org/10.3109/02699206.2015.1016188>
- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120(1), 407–415. <https://doi.org/10.1121/1.2205133>
- Davis, B. L., Jacks, A., & Marquardt, T. P. (2005). Vowel patterns in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, 19(4), 249–274. <https://doi.org/10.1080/02699200410001695367>
- Davis, B. L., Jakielski, K. J., & Marquardt, T. P. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics & Phonetics*, 12(1), 25–45. <https://doi.org/10.3109/02699209808985211>

- Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of Vowels and Vowel Disorders* (pp. 24–60). Psychology Press.
- Ferrero, F. E., Magno Caldognetto, E., & Cosi, P. (1996). Sui piani formantici acustici e uditivi delle vocali di uomo, donna e bambino. In *Atti del XXIV Convegno Nazionale AIA - Associazione Italiana di Acustica. 12-14 Giugno*, Trento, Italy, 169–178.
- Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech language pathologists. *American Journal of Speech-Language Pathology*, 12(3), 376–380. [https://doi.org/10.1044/1058-0360\(2003/083\)](https://doi.org/10.1044/1058-0360(2003/083))
- Frisch, S. A., Maxfield, N., & Belmont, A. (2016). Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. *Clinical Linguistics & Phonetics*, 30(3–5), 277–291. <https://doi.org/10.3109/02699206.2015.1137632>
- Goffman, L., & Smith, A. (1999). Development and phonetic differentiation of speech movement patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 25(3), 649–660. <https://doi.org/10.1037//0096-1523.25.3.649>
- Grigos, M. I., Hayden, D., & Eigen, J. (2010). Perceptual and articulatory changes in speech production following PROMPT treatment. *Journal of Medical Speech-Language Pathology*, 18(4), 46–53. PMID: PMC3442609.
- Grigos, M. I., & Kolenda, N. (2010). The relationship between articulatory control and improved phonemic accuracy in childhood apraxia of speech: A longitudinal case study. *Clinical Linguistics & Phonetics*, 24(1), 17–40. <https://doi.org/10.3109/02699200903329793>
- Grigos, M. I., Moss, A., & Lu, Y. (2015). Oral articulatory control in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 58(4), 1103–1118. [https://doi.org/10.1044/2015\\_JSLHR-S-13-0221](https://doi.org/10.1044/2015_JSLHR-S-13-0221)
- Grigos, M. I., Moss, A., & Tampakis, P. (2011). Language and motor interactions in childhood apraxia of speech. *Paper presented at the 6th International Conference on Speech Motor Control*, Groningen, The Netherlands.
- Hardcastle, B. & Tjaden, K. (2008). Coarticulation and Speech Impairment. In Ball, M. J., Perkins, M. R., Muller, N. & Howard, S. (Eds.), *The handbook of clinical linguistics* (506–524). Blackwell Publishing Ltd.
- Hayden, D., & Square, P. (1999). *Verbal Motor Production Assessment For Children – VMPAC*. Psychological Corporation.
- Higgins, C. M., & Hodge, M. M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology*, 10(4), 271–277.
- Hixon, T. J., Weismer, G., & Hoit, J. D. (2008). *Preclinical speech science: Anatomy, physiology, acoustics, perception*. Plural Publishing Inc.
- Howard, S., & Heselwood, B. (2013). The contribution of phonetics to the study of vowel development and disorders. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 61–111). Psychology Press.
- Iuzzini, J. (2012). *Inconsistency of speech in children with childhood apraxia of speech, phonological disorders, and typical speech*. Unpublished Doctoral Dissertation, Indiana University.
- Iuzzini, J., & Forrest, K. (2008). Acoustic vowel area in phonologically disordered, CAS, and normal speech. *Poster presented at the Conference for Motor Speech*, Monterey, CA.
- Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R. (2016). Speech inconsistency in children with Childhood Apraxia of Speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language and Hearing Research*, 60(5), 1194–1210. [https://doi.org/10.1044/2016\\_JSLHR-S-15-0184](https://doi.org/10.1044/2016_JSLHR-S-15-0184)
- Jacks, A., Marquardt, T. P., & Davis, B. L. (2013). Vowel production in childhood and acquired Apraxia of Speech. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 326–346). Psychology Press.
- Jacks, A., Marquardt, T. P., & Davis, B. L. (2006). Consonant and syllable structure patterns in childhood apraxia of speech: Developmental change in three children. *Journal of Communication Disorders*, 39(6), 424–441. <https://doi.org/10.1016/j.jcomdis.2005.12.005>
- Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, 26(2), 231–249. <https://doi.org/10.1044/jshr.2602.231>

- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684–686. <https://doi.org/10.1126/science.277.5326.684>
- Lenoci, G., & Ricci, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical Linguistics & Phonetics*, 32(12), 1126–1144. <https://doi.org/10.1080/02699206.2018.1510983>
- Lewis, B. A., Freebairn, L. A., Hansen, A. J., Iyengar, S. K., & Taylor, H. G. (2004). School-age follow up of children with childhood apraxia of speech. *Language, Speech, and Hearing Services in Schools*, 35(2), 122–140. [https://doi.org/10.1044/0161-1461\(2004/014\)](https://doi.org/10.1044/0161-1461(2004/014))
- Maassen, B., Nijland, L., & Terband, H. (2010). Developmental models of childhood apraxia of speech. In B. Maassen & P. Van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 243–258). Oxford University Press.
- Maassen, B., Groenen, P., & Crul, T. (2003). Auditory and phonetic perception of vowels in children with apraxic speech disorders. *Clinical Linguistics & Phonetics*, 17(6), 447–467. <https://doi.org/10.1080/0269920031000070821>
- Marquardt, T. P., Jacks, A., & Davis, B. L. (2004). Token-to-token variability in developmental apraxia of speech: Three longitudinal CAS studies. *Clinical Linguistics & Phonetics*, 18(2), 127–144. <https://doi.org/10.1080/02699200310001615050>
- Marquardt, T. P., Sussman, H. M., Snow, T., & Jacks, A. (2002). The integrity of the syllable in developmental apraxia of speech. *Journal of Communication Disorders*, 35(1), 31–49. [https://doi.org/10.1016/S0021-9924\(01\)00068-5](https://doi.org/10.1016/S0021-9924(01)00068-5)
- Meltzoff, A. N., & Moore, M. K. (1977). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6(3–4), 179–192. [https://doi.org/10.1002/\(SICI\)1099-0917\(199709/12\)6:3/4<179::AID-EDP157>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R)
- Miller, N. (1992). Variability in speech dyspraxia. *Clinical Linguistics & Phonetics*, 6(1–2), 77–85. <https://doi.org/10.3109/02699209208985520>
- Moss, A., & Grigos, M. I. (2012). Interarticulatory coordination of the lips and jaw in childhood apraxia of speech. *Journal of Medical Speech-Language Pathology*, 20(4), 127–132. PMID: 26005314.
- Nijland, L., Massen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 16(6), 461–483. <https://doi.org/10.1080/02699200210159103>
- Nijland, L., Massen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 17(1), 1–24. <https://doi.org/10.1080/0269920021000050662>
- Nip, I. S., Green, J. R., & Marx, D. B. (2011). The co-emergence of cognition, language, and speech motor control in early development: A longitudinal correlation study. *Journal of Communication Disorders*, 44(2), 149–160. <https://doi.org/10.1016/j.jcomdis.2010.08.002>
- Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.
- Perkell, J. S. (1997). Articulatory processes. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 333–370). Blackwell.
- Peter, B., & Stoel-Gammon, C. (2005). Timing errors in two children with suspected childhood apraxia of speech (sCAS) during speech and music-related tasks. *Clinical Linguistics & Phonetics*, 19(2), 67–87. <https://doi.org/10.1080/02699200410001669843>
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V. (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, 54, 1–14. <https://doi.org/10.1016/j.wocn.2015.07.002>
- Pisoni, D. B. (1980). Variability of vowel formant frequencies and the quantal theory of speech: A First report. *Phonetica*, 37(5–6), 285–305. <https://doi.org/10.1159/000259999>
- Pollock, K., & Hall, P. (1991). An analysis of the vowel misarticulation of five children with developmental apraxia of speech. *Clinical Linguistics & Phonetics*, 5(3), 207–224. <https://doi.org/10.3109/02699209108986112>

- Pollock, K. E. (2002). Identification of vowel errors: Methodological issues and preliminary data from the Memphis vowel project. In M. J. Ball & F. E. Gibbon (Eds.), *Vowel disorders* (pp. 83–113). Butterworth-Heinemann.
- Preston, J. L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 22(4), 627–643. [https://doi.org/10.1044/1058-0360\(2013/12-0139\)](https://doi.org/10.1044/1058-0360(2013/12-0139))
- Preston, J. L., Maas, E., Whittle, J., Leece, M. C., & McCabe, P. (2016). Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, 30(3–5), 3–5. <https://doi.org/10.3109/02699206.2015.1052563>
- Recasens, D., & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech Communication*, 48(6), 645–666. <https://doi.org/10.1016/j.specom.2005.09.011>
- Rvachew, S., Alhaidary, A., Mattock, K., & Polka, L. (2008). Emergence of the corner vowels in the babble produced by infants exposed to Canadian English or Canadian French. *Journal of Phonetics*, 36(4), 564–577. <https://doi.org/10.1016/j.wocn.2008.02.001>
- Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research*, 53(1), 114. [https://doi.org/10.1044/1092-4388\(2009/08-0184\)](https://doi.org/10.1044/1092-4388(2009/08-0184))
- Shriberg, L. D. (2010). A Neurodevelopmental framework for research in childhood apraxia of speech. In B. Maassen & P. Van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 259–270). Oxford University Press.
- Shriberg, L. D., Green, J. R., Campbell, T. F., McSweeney, J. L., & Scheer, A. R. (2003). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Clinical Linguistics & Phonetics*, 17(7), 575–595. <https://doi.org/10.1080/0269920031000138141>
- Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics*, 26(5), 445–482. <https://doi.org/10.3109/02699206.2012.655841>
- Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39(5), 331–349. <https://doi.org/10.1016/j.jcomdis.2006.06.017>
- Smith, A., Johnson, M., McGillem, C., & Goffman, L. (2000). On the assessment of stability and patterning of speech movements. *Journal of Speech, Language and Hearing Research*, 43(1), 277–286. <https://doi.org/10.1044/jslhr.4301.277>
- Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*, 45(1), 22–33. <https://doi.org/10.1002/dev.20009>
- Smith, B., Marquardt, T. P., Cannito, M. P., & Davis, B. L. (1994). Vowel variability in developmental apraxia of speech. In J. A. Till, K. M. Yorkston, & D. R. Beukelman (Eds.), *Motor Speech Disorders: Advances in assessment and treatment* (pp. 81–89). Paul H. Brookes Publishing Co.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17(1–2), 3–46. [https://doi.org/10.1016/S0095-4470\(19\)31520-7](https://doi.org/10.1016/S0095-4470(19)31520-7)
- Stoel-Gammon, C., & Cooper, J. (1984). Patterns of early lexical and phonological development. *Journal of Child Language*, 11(2), 247–271. <https://doi.org/10.1017/S0305000900005766>
- Stone, M. (2010). Laboratory techniques for investigating speech articulation. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 9–38). Wiley-Blackwell.
- Terband, H., Maassen, B., Van Lieshout, P., & Nijland, L. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders*, 44(1), 59–74. <https://doi.org/10.1016/j.jcomdis.2010.07.003>
- Thoonen, G., Maassen, B., Gabreels, F., Schreuder, R., & Swart, B. (1997). Towards a standardised assessment procedure for developmental apraxia of speech. *International Journal of Language & Communication Disorders*, 32(1), 37–60. <https://doi.org/10.3109/13682829709021455>
- Traunmuller, H. (1997). *Auditory scales of frequency representation*. [Online: <http://www.ling.su.se/staff/hartmut/bark.htm>]
- Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Blackwell.



- Vorperian, H. K., & Kent, R. D. (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\(2007/104\)](https://doi.org/10.1044/1092-4388(2007/104))
- Vuolo, J., & Goffman, L. (2018). Language skill mediates the relationship between language load and articulatory variability in children with language and speech sound disorders. *Journal of Speech, Language, and Hearing Research*, 61(12), 3010–3022. [https://doi.org/10.1044/2018\\_JSLHR-L-18-0055](https://doi.org/10.1044/2018_JSLHR-L-18-0055)
- Zharkova, N. (2017). Voiceless alveolar stop coarticulation in typically developing 5-year-olds and 13-year-olds. *Clinical Linguistics & Phonetics*, 31(7–9), 503–513. <https://doi.org/10.1080/02699206.2016.1268209>
- Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English/t/and/a/. *Journal of Phonetics*, 37(2), 248–265. <https://doi.org/10.1016/j.wocn.2008.10.005>
- Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, 15(1), 118–140. <https://doi.org/10.1123/mcj.15.1.118>
- Zmarich, C., & Bonifacio, S. (2003). Sui piani formantici acustici e uditivi delle vocali di infanti, bambini, e adulti maschi e femmine. In P. Cosi, E. Caldognetto Magno, & A. Zamboni (Eds.), *Voce, Canto, Parlato. Studi in onore di Franco Ferrero* (pp. 311–320). Unipress.