Journal of Fluency Disorders 76 (2023) 105975

Contents lists available at ScienceDirect

Journal of Fluency Disorders

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

# Temporal organization of syllables in paced and unpaced speech in children and adolescents who stutter

Mona Franke[a,b,c,d,*], Philip Hoole[a], Simone Falk[b,c,d]

[a] Institute for Phonetics and Speech Processing, Ludwig Maximilian University of Munich, Germany
[b] Faculté des arts et des sciences – Départment de linguistique et de traduction, Université de Montréal, Canada,
[c] International Laboratory for Brain, Music and Sound Research (BRAMS), Montréal, Canada
[d] Centre for Research on Brain, Language and Music (CRBLM), Montréal, Canada

A R T I C L E I N F O

*Keywords:*
Stuttering
Paced speech
Timing
Syllable organization
C-center effect

A B S T R A C T

*Purpose:* Speaking with an external rhythm has a tremendous fluency-enhancing effect in people who stutter. The aim of the present study is to examine whether syllabic timing related to articulatory timing (c-center) would differ between children and adolescents who stutter and a matched control group in an unpaced vs. a paced condition.
*Methods:* We recorded 48 German-speaking children and adolescents who stutter and a matched control group reading monosyllabic words with and without a metronome (unpaced and paced condition). Analyses were conducted on four minimal pairs that differed in onset complexity (simple vs. complex). The following acoustic correlates of a *c-center effect* were analyzed: vowel and consonant compression, acoustic intervals (time from c-center, left-edge, and right-edge to an anchor-point), and relative standard deviations of these intervals.
*Results:* Both groups show acoustic correlates of a c-center effect (consonant compression, vowel compression, c-center organization, and more stable c-center intervals), independently of condition. However, the group who stutters had a more pronounced consonant compression effect. The metronome did not significantly affect syllabic organization but interval stability improved in the paced condition in both groups.
*Conclusion:* Children and adolescents who stutter and matched controls have a similar syllable organization, related to articulatory timing, regardless of paced or unpaced speech. However, consonant onset timing differs between the group who stutters and the control group; this is a promising basis for conducting an articulatory study in which articulatory (gestural) timing can be examined in more detail.

## 1. Introduction

### 1.1. Syllable structure and stuttering

Stuttering is a neurodevelopmental speech fluency disorder that approximately 5–8% of preschool children develop (Yairi & Ambrose, 2013). However, the majority of children spontaneously recover from stuttering in childhood, giving a prevalence of

* Correspondence to: International Laboratory for Brain, Music and Sound Research (BRAMS), Pavillon Marie-Victorin, Université de Montréal, 90 Vincent D'Indy Ave Local A-108, Outremont, Quebec H2V 2S9, Canada.
*E-mail address:* mona.franke@phonetik.uni-muenchen.de (M. Franke).

https://doi.org/10.1016/j.jfludis.2023.105975
Received 30 September 2022; Received in revised form 28 March 2023; Accepted 8 May 2023
Available online 11 May 2023

approximately 1% in the adult population (Yairi & Ambrose, 2013). Furthermore, girls are more likely to recover from stuttering than boys; while boys and girls are equally affected, there are more men who stutter than women who stutter (ratio approximately 4:1) (Yairi & Ambrose, 2013).

The speech of persons who stutter (PWS) is characterized by involuntary interruptions due to repetitions of sounds, syllables, or words, involuntary blocks, or prolongations of sounds (World Health Organization, 2015). These symptoms do not occur randomly within a syllable. Most often, they occur at the word or syllable onset. Harrington's (1987) auditory, acoustic, and electropalatographic study suggests that stuttering symptoms appear before or within the acoustic onglide of the vowel and never in the rhyme (that is, the nucleus and offset consonants of a syllable). Disfluencies that occur at the end of words or syllables (i.e., echodysphemia) are nowadays considered as a subgroup of developmental dysfluency, in addition to stuttering and cluttering (MacMillan, Kokolakis, Sheedy & Packman, 2014).

Speech fluency breakdowns in stuttering have been generally linked to an atypical development of speech motor processes (Smith & Weber, 2016). An explanation for the breakdown characteristics of stuttering at syllable or word onsets was proposed by Wingate (1988) over 30 years ago. Wingate postulated the so-called Fault-Line Hypothesis; this hypothesis is based on the idea that stuttering symptoms occur at the transition from the initial consonant to the stress-bearing vowel (Wingate, 1988). Therefore, he posits that this would lead to a divide or "Fault-Line" at the point of syllable onset and rhyme integration which is caused by the delayed encoding of the vowel (Wingate, 1988). Hence, Wingate's Fault-Line Hypothesis claims that the timing relationship between consonants and vowels is atypical in PWS, especially in stressed syllables.

There is evidence from experimental research that onset-vowel timing is particularly challenging for PWS because their speech motor program breaks down at the point where these specific timing relations usually occur, even in perceptually fluent speech. For instance, adults who stutter show more variable vowels when producing monosyllabic nonwords comprising a consonant, vowel, and a consonant (such as in "vep") and more variable fricatives (/s/) in nonwords with the structure: vowel, /s/, vowel (such as in "asi") (Di Simoni, 1974). They also produce more variability in stop gap durations and voice onset time related to syllable onsets in monosyllabic words (Max & Gracco, 2005), as well as more variability in voice onset times of the same stop consonant in syllable initial position in a trisyllabic word (Jäncke, 1994). Furthermore, it was found that when producing word-initial bilabials adults who stutter have longer bilabial closing intervals, measured from the movement onset of the lip closing and the peak velocity of the closing movement to the offset of vocal fold vibration for the preconsonantal vowel (Max & Gracco, 2005). Another study found a similar result, namely overall longer bilabial closing durations and a higher peak velocity when releasing the bilabial onset of nouns (Max, Caruso, & Gracco, 2003). Adults who stutter also have larger amplitudes of upper lip movement when producing syllable onset bilabials (Namasivayam & van Lieshout, 2008). Heyde, Scobbie, Lickley, and Drake (2016) found that adults who stutter transition from the onset to the vowel with a decreased peak velocity, compared to the control group, which indicates that adults who stutter have lower acceleration/deceleration in releasing the constriction of the consonant. The authors concluded that these differences support Wingate's Fault-Line Hypothesis, since PWS might have difficulty integrating syllable rhymes with their onsets. Another study identified two general strategies that promoted fluency in PWS, namely the reduction of speech rate and the reduction of coarticulation (Zmarich, Balbo, Galatà, Verdurand & Rossato, 2013). However, in general, adults who stutter are less consistent in inter-articulator coordination across mono- and multisyllabic nonwords (Smith, Sadagopan, Walsh & Weber-Fox, 2010; Wiltshire, Chiew, Chesters, Healy & Watkins, 2021).

Less research has been dedicated to the articulatory properties of children's speech in stuttering. A few studies show that children who stutter also display speech motor control differences compared to children who do not stutter. For example, children who stutter show greater articulatory coordination variability across mono- and multisyllabic nonwords (4–5 year-olds, Smith, Goffman, Sasisekaran & Weber-Fox, 2012) and sentences (5–7 year-olds, Usler, Smith, & Weber, 2017; 6–12 year-olds, Usler & Walsh, 2018). These investigations used the lip aperture variability index, which reflects the inter-articulatory coordination of lip and jaw movements – articulators that were involved in inter-gestural timing between many onset (bilabials) and vowel combinations that were present in the stimuli used by the authors. Interestingly, children who recovered from stuttering did not differ from the control group in terms of articulatory coordination (Usler et al., 2017). Furthermore, children who stutter ranging from 8 to 12 years of age displayed more variable voice onset times compared to a matched control group at all three complexity levels – onsets with a single consonant, a two-consonant cluster, and a three-consonant cluster whereby variability was the largest at the single consonant level, followed by the three-consonant onset, and then the biconsonantal onset (De Nil & Brutten, 1991). Moreover, voice onset time in children who stutter (6.5–7.5 year-olds) was more variable in syllable-initial position when produced in a disyllabic word and when included in a sentence, as well (Dokoza, Hedever, & Sarić, 2011). To summarize, these studies lead to the assumption that onset-vowel coupling might differ between children and adolescents who stutter (test group) and children and adolescents who do not stutter (control group), even when considering perceptually fluent speech.

There is evidence that rhythm can affect the timing between two individual articulatory movements (also referred to as inter-gestural timing). For instance, Tilsen (2009) found that the inter-gestural timing between two consonant gestures in a CCV syllable structure was more stable when the phrase in which the word occurred was produced with an easier rhythm (nearer to low-order harmonic ratios). The rhythm was controlled by a metronome. When rhythmic timing was more variable, the inter-gestural timing also became more variable. This is an interesting finding, since an external rhythm works as a fluency-enhancing condition for PWS. For instance, speech fluency tremendously increases in PWS when they synchronize their speech to a metronome (Wingate, 1969; Andrews, Howie, Dozsa & Guitar, 1982). This fluency-enhancing effect is accompanied by a normalization of hyper- and hypo-activation in neural circuits mediating temporal processing and movement initiation, such as the basal ganglia and the cerebellum (e.g., Toyomura, Fuji, & Kuriki, 2011), indicating better coupling of auditory and motor systems (Stager, Jeffries, & Braun, 2003). Therefore, metronome-paced speech also leads to a more stable speech motor coordination (Van Lieshout, Namasivayam, 2010).

The aim of the present study is therefore, to analyze temporal organization of syllables in children and adolescents who stutter and matched control participants, speaking with and without an external rhythm. We examine syllabic timing related to articulatory timing by analyzing the *c-center effect,* which is described in the following section.

### 1.2. Framework for analyzing temporal syllable structure

From a purely articulatory point of view, an elementary task of fluent speech production is the coordination of movements among groups of articulators (Browman & Goldstein, 1991, 1992). In general, the internal structure of the syllable is asymmetrical, which is crucial for inter-articulatory coordination of the speech gestures involved (captured, for example, in the Coupled Oscillator Model of Syllable Structure, see e.g., Nam & Saltzman, 2003; Goldstein & Pouplier, 2014). Gestures, defined in the framework of Articulatory Phonology, are distinct vocal tract actions (Browman & Goldstein, 1992). In Articulatory Phonology, there are two main gesture phasing patterns that have been suggested to describe the timing relationships between consonants and vowels: In-phase and anti-phase coupling. While onset consonants are supposed to be timed in-phase with the vowel, coda consonants are presumed to be timed anti-phase with the vowel. More specifically, in-phase coupling means that the articulatory gesture of the onset consonant and the articulatory gesture of the vowel start at the same time; by contrast anti-phase coupling means that the coda consonant gesture begins when the vowel gesture reaches its peak (e.g., Hall, 2010). However, there are situations of conflict where gestures simultaneously aim to reach opposing goals, e.g. a lip closure to produce a [p] and jaw opening to produce an [a]. In this case, there is a competition between the ideal phasing relationship (in-phase) and reaching the acoustic goal. The consonant will try to get as close to the ideal phasing relationship with the vowel as possible while still reaching the acoustic goal.

Byrd (1996) discovered that there is less variability at the level of gestural overlap in onsets compared to codas. In particular, it appears that onset consonant gestures and vowel gestures together form a much more cohesive unit than vowel gestures and coda consonant gestures (e.g., Hoole & Pouplier, 2015). This asymmetry underlies the so-called *c-center effect* according to which there is a constant temporal relationship between the temporal center of the onset and the following vowel regardless of the number of consonants contained in the onset (Browman & Goldstein, 1988). Describing this phenomenon with the coupling model, consonants forming a complex onset (e.g., CC) are coupled to each other anti-phase, while each consonant in the onset is coupled in-phase with the following vowel (e.g., Browman & Goldstein, 2000). In a complex coda on the other hand, there is only anti-phase coupling between the vowel and the coda consonant as well as between the consonants themselves. Therefore, there is no constant temporal relationship between the vowel and the coda.

The following figure (Fig. 1) illustrates the coupling relationships between the consonants (C1, C2) and the vowel in an onset and in a coda organization. The competitive coupling topology in the onset organization (i.e. the combination of in-phase and anti-phase coupling) can be shown to lead to a shift of the rightmost consonant in an onset cluster towards the vowel as more consonants are added to the onset, leading to the *c-center*.

While traditional articulatory studies, e.g. using electromagnetic articulography, provide the most direct evidence for gestural organization, they are, of course, time-consuming and hence often go in hand with small participant sample sizes. In fact, the acoustic signal – which is far more efficient to obtain – can also provide evidence for a *c-center effect*. For example, following a gestural approach to syllable organization (Browman & Goldstein, 1988) and summarized by Katz (2010), the (acoustic) duration of the vowel should be shorter in syllables with a complex onset compared to syllables with a simple onset due to the shift of the rightmost onset consonant towards the vowel. Note that we refer to this phenomenon as vowel compression from here on. Moreover, it is suggested that gestural overlap, in general, would cause compression to arise. Hence, acoustic compression should also be observed in the onset, when we compare a single onset to the same consonant (C) in $C_2$ position, e.g. [l] in [klaʊd] vs. in [laʊd]. According to this assumption, [l] in $C_2$ would be acoustically shorter than [l] in $C_1$ because the gesture of [l] in the syllable with the complex onset is more shifted towards the vowel gesture. Henceforth, this will be referred to as consonant compression.

Several studies found vowel compression (e.g., Marin & Pouplier, 2010; Katz, 2012, Brunner, Geng, Sotiropoulou & Gafos, 2014; Marin & Bučar Shigemori, 2014; Peters & Kleber, 2014), as well as consonant compression (e.g., Katz, 2010; Marin & Bučar Shigemori, 2014; Gibson, Fernández Planas, Gafos & Remirez, 2015) in complex vs. simple syllables (for instance CCV vs. CV). Moreover, there are studies that showed the usefulness of acoustic methods to measure stability patterns of word-initial consonant clusters (e.g., for English: Selkirk & Durvasula, 2013; for Jazani Arabic: Ruthan, Durvasula, & Lin, 2019). The following figure (Fig. 2) displays a sketch of two different stability patterns that can be observed for languages allowing complex onsets (left panel) and languages that do not allow complex onsets (right panel).

Selkirk & Durvasula (2013) replicated results on English with acoustic data, showing that there is more timing stability (measured with the relative standard deviation [RSD][1]) in the c-center to anchor interval (c-center in this case was defined as the midpoint of the onset and the anchor was the end of the vowel) compared to other intervals (right-edge and left-edge). For Jazani Arabic it was found that the right-edge (the midpoint of the right-most consonant in the onset) to anchor interval was most stable, which led to the conclusion that this specific dialect has a simple onset organization (Ruthan et al., 2019), contrary to English and German, for example.

In summary, this articulatory framework is particularly interesting for studying individuals who stutter, as it provides insights into

[1] The RSD is a common measure of interval stability used in studies that examine gestural timing (e.g., Shaw, Gafos, Hoole & Zeroual, 2011; Brunner et al., 2014; Ruthan et al., 2019), since it takes into account the fact that shorter durational intervals typically have a lower absolute standard deviation. Using the latter would thus bias the results towards greater stability for the right-edge to anchor interval since this is by definition the shortest of the three intervals considered.

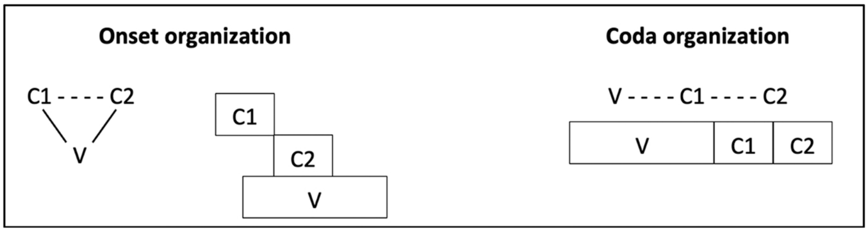


**Fig. 1.** Phasing relations in a complex onset and complex coda. Dashed lines display an anti-phase coupling, solid lines display an in-phase coupling.

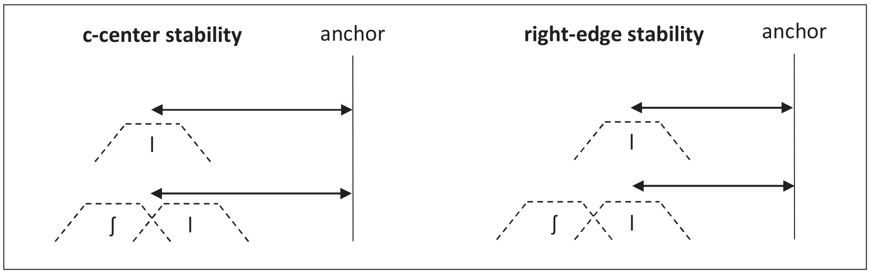


**Fig. 2.** Temporal organization of complex (c-center stability) and simplex (right-edge stability) onset organization.

syllabic organization and associated speech planning. Also, the approach provides clues to articulatory timing processes as well as articulatory control – an area that is particularly challenging for PWS as pointed out in the previous section.

### *1.3. Hypotheses*

Since German admits complex onsets, we would expect to find similar results to the study on English (Pouplier, 2012), where the c-center to anchor interval was most stable. On the one hand, we expect PWS to show c-center stability but overall more variability than persons who do not stutter (PWNS). On the other hand, using a metronome to regulate speech timing could enhance the stability of this interval in PWS.

In particular, we aimed to examine whether syllabic timing related to articulatory timing (c-center) would differ between groups in an unpaced vs. a paced condition.

We hypothesize to find acoustic cues for a *c-center effect* in both groups (by comparing complex and simple syllables), with a difference between the paced and unpaced condition, and between the control group and the test group.

More precisely, we hypothesized that we would find

(1) consonant compression in both groups and
(2) vowel compression in both groups

with a group difference. As for this difference, either direction is possible. Less compression in the test group could derive from less coarticulation, for example, as a strategy to maintain fluency (Zmarich et al., 2013). It is another possibility that compression effects will be more pronounced in the test group, if gestures of PWS were to overlap too much (as for example predicted by Harrington, 1988). We hypothesize that

(3) a pacing condition will amplify the compression effects and that the groups will not differ in the paced condition.

Furthermore, we hypothesize to see evidence for a *c-center effect* by finding

(4) more stability in the interval from the acoustic *c-center* to the end of the vowel compared to the left-edge or right-edge interval, together with an increase in stability in the paced condition, whereby the group who stutters will benefit more from the paced condition leading to a higher increase in stability.

Analyzing the perceptually fluent speech of PWS can give us clues on whether PWS might have difficulties in the coordination of onset (consonant) and nucleus (vowel) timing, and thus, syllabic organization. In addition, acoustic analyses enable us to analyze a larger participant sample size and since this is the first study that examines a potential *c-center effect* in PWS and a matched control group, the present study can serve as a basis for further articulatory investigations.

## 2. Materials and methods

### 2.1. Participants

48 PWS (42 males, 6 females, Mean age = 13.03, SD = 2.55, range = 9–18) and 48 age- and gender-matched controls (Mean age = 12.94, SD = 2.46, range = 9–18) participated in the experiment. All participants spoke German at a native level. All except five participants reported an absence of language, hearing and/or neurological deficits (other than developmental stuttering which was diagnosed by a speech therapist). These five participants had Dyslexia or ADHD in addition to stuttering. All participants and the parents of the underaged participants signed for informed consent. The participants who stutter were recruited through the intensive therapy course "Stärker als Stottern" (staerker-als-stottern.de), during which their stuttering was assessed by trained speech therapists. The stuttering severity was determined with the SSI-3 by trained speech therapists based on recordings on the day the data were recorded (Table 1). All participants who stutter were participating in the intensive therapy course but recordings were done prior to the start of the course. The typically fluent participants were recruited through schools.

### 2.2. Material

Participants were asked to read two separate wordlists that either contained monosyllabic words with simple onsets ($W_{simple}$) or monosyllabic words with complex onsets ($W_{complex}$). Words were either nouns or adjectives. The syllable structure of the monosyllabic words was either consonant vowel consonant (CVC) in the $W_{simple}$ wordlist or CCVC in the $W_{complex}$ wordlist, whereby vowels were either a short vowel or a diphthong. Each list contained a set of practice items at the beginning (5 words in total) to familiarize participants with the wordlist reading pattern. These words were followed by 12 nouns and 12 adjectives which occurred twice in the same order in the same wordlist. The words were printed on paper in landscape format (DIN A4, in 14-point Arial font) in rows of 12 words.

As target words, we chose 4 word pairs that only differed in onset complexity (see Table 2*)* and that were suitable for segmentation solely based on the acoustic signal. Since our focus was on the first consonant ($C_1$) in the onset in words with a single onset ($W_{simple}$) and the second consonant ($C_2$) in the onset of words with a complex onset ($W_{complex}$), we chose words with an [l] in these positions as it is fairly easy to detect (in comparison to a plosive, for instance, where we could not detect the closure phase in words with a single onset). Target words were not situated at the margins (beginning or end) of the wordlist to avoid patterns like phrase-final lengthening or a different intonation contour.

### 2.3. Procedure

Participants were comfortably seated at a table with the experimenter present in the room. The wordlist was placed in front of them on the table by the experimenter who also presented the second wordlist after they finished the first. Participants were recorded with a Zoom H4N recorder (44.1 kHz, 16 bit) via an external headset microphone (beyer dynamic opus 54.16/3) in a quiet room.

Every wordlist was read twice per participant in each condition (unpaced and paced). In the unpaced condition, participants were instructed to read the wordlist in their preferred speech tempo. In the paced condition, they were asked to read the wordlists along with a metronome that had an inter-onset-interval (IOI) of 900 ms. In this condition, every word was to be timed with one metronome beep. Half of the participants of each group started with the $W_{simple}$ wordlist and the other half with the $W_{complex}$ wordlist (randomized by the experimenter). Paced reading always followed unpaced reading because otherwise the paced reading could have impacted the speech rate of the unpaced reading.

### 2.4. Analyses

The segmentation of every word and the corresponding segments was based on the acoustic signal and the oscillogram using Praat (version 6.1) (Boersma & Weenink, 2019) and was done by phonetics students that were trained in acoustic segmentation. Words were excluded from analyses if they displayed a markedly increased tonus or speech rate, a blockade, prolongation or repetition of sounds. Incorrectly read words were also excluded. Hence, only the perceptually fluent speech was analyzed. Exclusions were based on the assessment of the first author and checked by a trained speech therapist. The rules for segmentation were defined as follows:

Laterals were segmented at the first positive zero crossing of the first recognizable period, and fricatives at the beginning of frication in the signal. Note that [klaŋ] is a special case since the start of the plosive is not measurable (the closure duration cannot be detected based on the acoustic recordings) and the [l] may be largely voiceless. For this reason, the word pair *Klang-lang* was excluded

**Table 1**
Distribution of stuttering severity at the recording day across participants.

| Stuttering severity | Very mild | Mild | Mild-Moderate | Moderate | Moderate-severe | Severe | Severe - very severe | Very Severe |
|---|---|---|---|---|---|---|---|---|
| **No. of participants** | 2 | 8 | 3 | 11 | 2 | 10 | 1 | 11 |
| **Mean SSI score (SD)** | 8 (1) | 15.25 (2.86) | 19.67 (4.5) | 22.36 (0.98) | 22.5 (0.5) | 30.4 (1.56) | 35 | 41.18 (5.11) |

**Table 2**
Target words with simple (left column) and complex onsets (right column). One row displays one word pair.

| $W_{simple}$ | $W_{complex}$ |
|---|---|
| Leim [laɪ̯m] (Engl. *glue*) | Schleim [ʃlaɪ̯m] (Engl. *slime*) |
| Lamm [lam] (Engl. *lamb*) | Schlamm [ʃlam] (Engl. *mud*) |
| Lauch [laʊ̯x] (Engl. *leek*) | Schlauch [ʃlaʊ̯x] (Engl. *tube*) |
| lang [laŋ] (Engl. *long*) | Klang [klaŋ] (Engl. *sound*) |

for some analyses. The beginning of a vowel was segmented at the second zero crossing of the first recognizable period and the end of the vowel was determined by the beginning of the coda consonant. In this case, a nasal was segmented at the time when antiresonances were present and/or at the first clearly visible change in the periodic pattern in the oscillogram at the first positive zero crossing. The segmentation of a fricative in coda position was segmented at the beginning of frication, as described for segmentation in onset position. The following figure (Fig. 3) displays a segmentation example of the word pair *Schleim-Leim*.

The following table displays the number of excluded words out of 384 in total per condition and per group. Table 3.

*2.4.1. Acoustic correlates of a c-center effect*

We analyzed three different correlates of a *c-center effect*, namely consonant compression, vowel compression (following Katz, 2010) as well as three intervals (left-edge to vowel offset, right-edge to vowel offset, and the c-center to vowel offset) which are associated with (articulatory) syllable organization.

*2.4.1.1. Consonant compression.* To analyze consonant compression, we extracted [l]-durations of $W_{simple}$ and $W_{complex}$ and ran a linear mixed model (see *3.2. Consonant compression*) to compare [l]'s that were produced as $C_1$ in $W_{simple}$ with [l]'s that were produced in $C_2$ in $W_{complex}$.

*2.4.1.2. Vowel compression.* For the analysis of vowel compression, we extracted vowel durations of $W_{simple}$ and $W_{complex}$ and compared them following the procedure for consonant compression.

*2.4.1.3. Intervals (acoustic c-center).* The methodology for this part follows that of Ruthan et al. (2019). They proposed an acoustic method to calculate three intervals that were measured in an articulatory study (e.g., by Shaw et al., 2011). Note that the definition of these intervals was slightly different in articulatory studies (e.g., Shaw et al., 2011; Brunner et al., 2014). The intervals in the present study are defined as follows:

(1) Left-edge to anchor: This interval was calculated as the duration from the midpoint of the left-most consonant (= left-edge) to the end of the vowel (= anchor).
(2) Right-edge to anchor: This interval was calculated as the duration from the midpoint of the right-most consonant (= right-edge) to the end of the vowel (= anchor).
(3) C-center to anchor: This interval was determined by calculating the duration from the mean of the midpoints of the two onset consonants (c-center) to the end of the vowel (=anchor).

For better illustration, Fig. 4 displays the calculation of the intervals.

If an onset only has a single onset consonant, the three intervals (left-edge to anchor, right-edge to anchor, and c-center to anchor) do not differ, as can be seen in Fig. 4.

To identify the most stable interval (it is hypothesized that the c-center is the most stable interval in our study), we followed previous studies (Shaw et al., 2011; Brunner et al., 2014; Ruthan et al., 2019) and used the relative standard deviation (RSD), which is defined as *100 * (standard deviation of the duration/mean duration)*. It is a measure that refers to the variability of an interval over word pairs (e.g., Brunner et al., 2014; Ruthan et al., 2019) or triads (e.g. Shaw et al., 2011). Therefore, RSDs were calculated across word pairs per group.[2]

We also compare the RSD between groups and analyze whether the test group and the control group differ significantly in their stability patterns.

*2.4.2. Statistical analyses*

All statistical analyses were conducted with R Version 4.0.2 (R Core Team, 2020). We used the package tidyverse (Wickham et al., 2019) for data processing and lme4 (Bates, Maechler, Bolker & Walker, 2015) to perform linear mixed effects analyses. Linear mixed effects models were calculated to estimate [l] and vowel duration (with regards to compression effects), and a linear mixed effects model was also run to estimate interval duration for the c-center, left-edge, and right-edge interval, as well as the RSD. Variables that were included in the models as random or fixed effects were GROUP (test group vs. control group), CONDITION (paced vs. unpaced),

[2] We calculated the RSD per group and not per participant because we had maximally 2 repetitions per word and condition per participant.

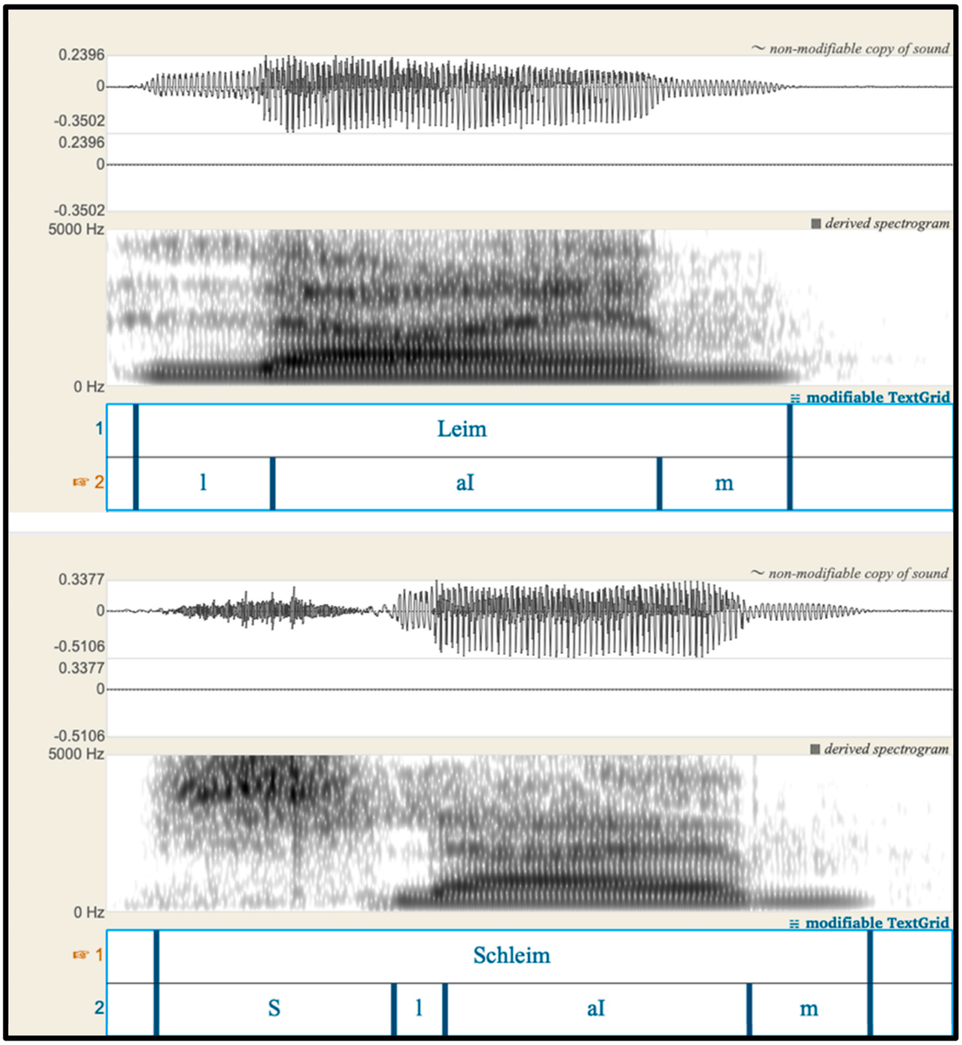


**Fig. 3.** Segmentation example from Praat with oscillogram and spectrogram. Tier 1: Word, tier 2: Phones (Sampa transcription).

**Table 3**
Target words excluded per group and condition.

| | PWNS | PWS |
|---|---|---|
| **$W_{simple}$ unpaced** | 0 | 49 |
| **$W_{simple}$ paced** | 4 | 44 |
| **$W_{complex}$ unpaced** | 2 | 39 |
| **$W_{complex}$ paced** | 3 | 29 |

ONSET COMPLEXITY (one or two segment(s)), WORD PAIR (4 in total, see Table 2), WORD (8 in total, see Table 2) and PARTICIPANT.

We followed the same procedure for all models:

We started all linear mixed models with a full model including a three-way interaction term between the fixed factors (GROUP, CONDITION, ONSEST COMPLEXITY in the case of predicting compression effects and interval durations, and INTERVAL, CONDITION, ONSET COMPLEXITY in the case of predicting RSD) as well as random intercepts and random slopes (intercepts for PARTICIPANT and WORD PAIR, by-participant random slopes for CONDITION and ONSET COMPLEXITY/INTERVAL and by-word-pair random slopes for GROUP and CONDITION). Likelihood-ratio tests were performed using the R-function *anova*, to compare several models with the intention to find the best fit model. Model fit was assessed with BIC and the variance that the model explains was estimated using the function *r2_nakagawa*.

The complexity of all models could be reduced to a two-way interaction term between the fixed factors and by excluding all by-word-pair random slopes (because of perfect correlations of WORD PAIR and GROUP, as well as of WORD PAIR and CONDITION

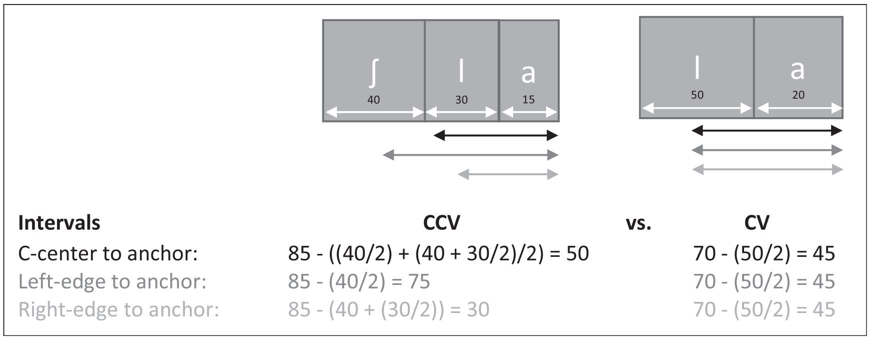


**Fig. 4.** Example calculation of intervals in ms for a CCV and a CV syllable structure.

and singular fit). Hence, the final models included the fixed effects with a 2-way interaction between all variables, as well as intercepts for PARTICIPANT and WORD PAIR with by-participant random slopes for CONDITION and ONSET COMPLEXITY (in the RSD model only by-participant random slope for CONDITION). All models were first calculated for all participants and then without the five participants with comorbidities to check for the robustness of the results.

Residual plots were visually checked for homoscedasticity of normality before reporting the results. Main effects are reported, using the R-function *anova,* and interactions were analyzed with Post-hoc Tukey corrected t-tests, using the package *emmeans* (Lenth, 2020). Pairwise comparisons were done using the contrast function in the package *emmeans,* and correlations were done using Spearman-rho correlations. Before examining acoustic correlates of a *c-center effect*, word duration was analyzed in order to reveal potential group differences in speaking rate in children and adolescents who stutter and matched peers.

## 3. Results

### 3.1. Word duration

Since acoustic correlates of a *c-center effect* are mirrored in durations (durational compression and interval duration), we analyzed word duration beforehand in order to be able to interpret the following results with this information in mind. Fig. 5 displays the word duration grouped by syllable structure (simple vs. complex onsets) and condition (paced vs. unpaced) for the control group (PWNS)

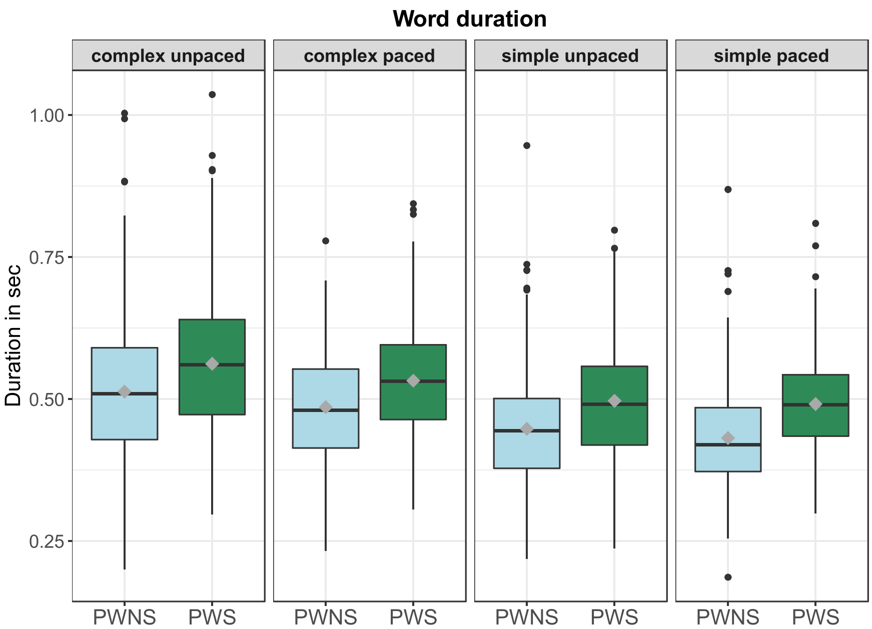


**Fig. 5.** Word duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.

and test group (PWS).

To predict word duration, we followed the procedure described in 2.4.2 Statistical analyses except that we replaced the random intercept WORD PAIR with WORD. Residuals were slightly right-skewed but since the model predicts real word durations of groups who might differ in speech tempo, this is expected. A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 1: Word duration*).

The model (Conditional $R^2 = 0.758$, Marginal $R^2 = 0.130$) revealed that GROUP ($F[1, 92.01] = 20.56$, $p < .001$) and CONDITION ($F[1, 68.8] = 11.2$, $p = .01$) significantly predicted word duration. As can be seen in Fig. 5, the group who stutters showed longer word durations than the control group and words were longer in the unpaced condition compared to the paced condition. Moreover, there was a significant interaction between ONSET COMPLEXITY and CONDITION ($F[1, 6.09] = 7.06$, $p = .04$). Post-hoc tests showed that the paced condition only affected the duration of words with complex onsets ($p < .001$) but not words with simple onsets, indicating that word duration significantly decreased in $W_{complex}$ produced along with a metronome but not in $W_{simple}$.

The results did not change when running the same model without the five participants who had comorbidities. Based on the results regarding word duration, significant main effects of condition (paced vs. unpaced) and group (PWS vs. PWNS) are also expected for the acoustic correlates of a c-center. Therefore, these main effects will not be reported in detail, since they only display the speech rate differences we found in the word duration analysis. Moreover, residuals are also expected to be slightly right and potentially left-skewed. Log transforming the durations ([l]-duration, vowel duration, and the interval duration) did not improve the skewness. For this reason, the distribution of residuals for the following statistical models will no longer be reported.

### 3.2. *Consonant compression*

If consonant compression takes place, we would expect the duration of [l] to depend on ONSET COMPLEXITY (shorter [l]-duration in $W_{complex}$ than in $W_{simple}$). Moreover, if the groups differ in consonant compression, we would expect to find a significant interaction between GROUP and ONSET COMPLEXITY. Furthermore, we examined potential effects of metronome pacing (CONDITION). Fig. 6 displays [l]-duration in seconds for $W_{complex}$ and $W_{simple}$ for each condition per group.

A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 2: l duration*). The model revealed a strong correlation between the participant intercept and onset complexity (see Fig. 7).

The significant correlation between participant intercept and estimated deviation for the complex onset ($R = -0.89$, $p < .001$) indicates that participants who produced a long [l]-duration in $W_{simple}$, produced shorter [l]-durations in $W_{complex}$. Hence, these particpants produced a bigger change from $W_{simple}$ to $W_{complex}$.

The linear mixed effects model (Conditional $R^2 = 0.528$, Marginal $R^2 = 0.281$) showed a significant effect of GROUP ($F[1, 89.17] = 13.97$, $p < 0.001$), ONSET COMPLEXITY ($F[1, 92.41]= 315.54$, $p < .001$), and a significant interaction between GROUP and ONSET COMPLEXITY ($F[1, 92.41] = 4.32$, $p = .04$). Pairwise comparisons between GROUP and ONSET COMPLEXITY revealed that the

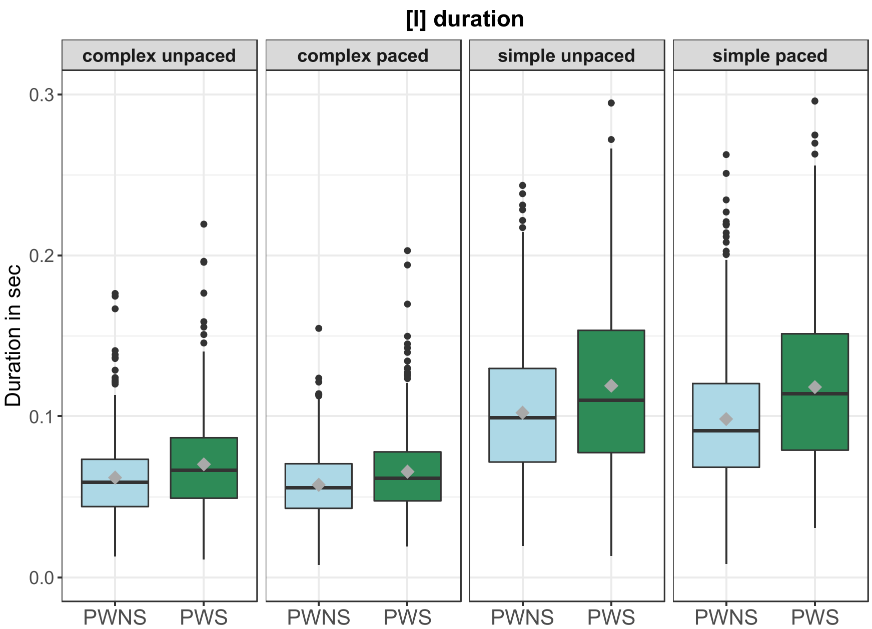


**Fig. 6.** [l]-duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.

groups differed significantly in the duration of [l] between $W_{simple}$ and $W_{complex}$ ($p = .04$). That is, participants who stutter showed a significantly bigger difference between the [l]-duration of $W_{simple}$ and $W_{complex}$.

Whether participants read at their preferred tempo or with a metronome did not affect [l]-duration significantly ($F[1, 79.39] = 3.85$, $p = .053$). No further interactions became significant. These results suggest that, although the test group produced significantly longer [l]-durations compared to the control group (given the slower speech rate in this group), they also showed more pronounced consonant compression.

Running the same model without the five participants who stutter who had comorbidities (Conditional $R^2 = 0.538$, Marginal $R^2 = 0.278$) did not change the significant main effects. GROUP ($F[1, 84.76] = 12.06$, $p < .001$) and ONSET COMPLEXITY ($F[1, 87.38] = 285.82$, $p < .001$) were still significant predictors of [l]-duration. However, the interaction between GROUP and ONSET COMPLEXITY was no longer significant, indicating that the groups did not differ in consonant compression anymore.

In conclusion, these results suggest that we observe consonant compression in both groups, whereby the group who stutters produced more consonant compression. Moreover, the results suggest that the pacing condition did not have a significant effect on consonant compression, nor [l]-duration in general since the model did not reveal a significant interaction with CONDITION and ONSET COMPLEXITY, nor a significant main effect.

### 3.3. Vowel compression

For vowel compression to occur, vowels of $W_{complex}$ must be shorter than $W_{simple}$. If the groups differ significantly in vowel compression, we would expect to find a significant interaction between ONSET COMPLEXITY and GROUP. Fig. 8 displays vowel duration in seconds for $W_{complex}$ and $W_{simple}$ for each condition per group.

The model (Conditional $R^2 = 0.848$, Marginal $R^2 = 0.031$) revealed that GROUP ($F[1, 91.72] = 5.20$, $p = .025$) and ONSET COMPLEXITY ($F[1, 92.24] = 132.28$, $p < .001$) were significant predictors of vowel duration. As can be seen in Fig. 8, vowel durations are longer in $W_{simple}$ compared to $W_{complex}$. Furthermore, the model revealed a significant interaction between ONSET COMPLEXITY and CONDITION ($F[1, 2679.48] = 5.45$, $p = .02$). Post-hoc Tukey corrected t-tests support vowel compression as it was found that vowels in words with a simple onset were always longer than in words with a complex onset, regardless of condition ($p < .001$). However, pairwise comparisons revealed that vowel compression significantly differed between the paced and the unpaced condition, indicating more vowel compression in the paced condition ($p = .02$). Nonetheless, the low marginal $R^2$ points out that most of the variance cannot be explained with the fixed effects. Running the same model without the five participants who had comorbidities (Conditional $R^2 = 0.848$, Marginal $R^2 = 0.032$) did not change the results.

Since there was no significant interaction between GROUP and ONSET COMPLEXITY, it can be assumed that the groups do not differ in vowel compression. To summarize, we can conclude that vowel compression occurred in both groups; furthermore, the pacing condition did not affect vowel duration but did affect vowel compression, as the latter was significantly more pronounced in the paced condition.

### 3.4. Acoustic c-center (intervals and RSDs)

#### 3.4.1. Intervals

In this section we analyze whether the participants had timing patterns characteristic of a complex syllable organization (it is assumed that the intervals of $W_{simple}$ and $W_{complex}$ differ less in the c-center interval compared to the other intervals), whether the groups differ in syllabic organization and whether the pacing condition affects this. Therefore, we compared the intervals of c-center,

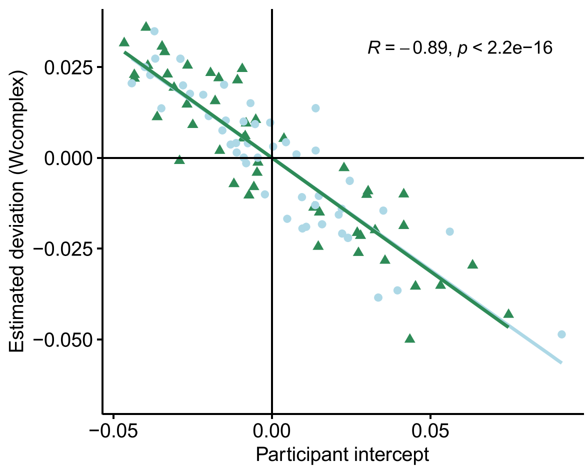


**Fig. 7.** Correlation between participant intercept and estimated deviation for $W_{complex}$ (units in seconds). 0 displays the mean value for all participants. Participants who do not stutter = blue and circles, participants who stutter = green and triangles.

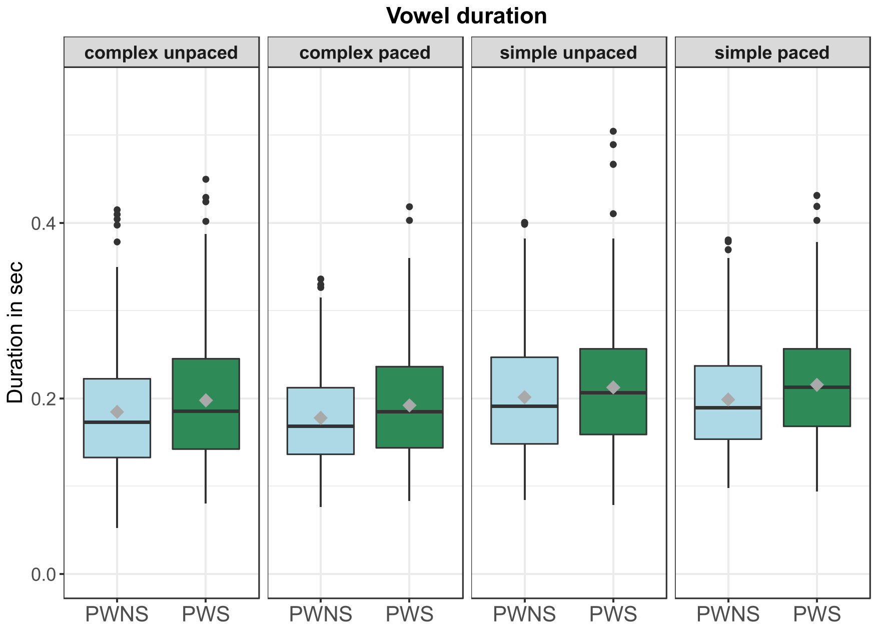


**Fig. 8.** *Vowel duration for each group per condition. Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers. Participants who do not stutter = blue, participants who stutter = green, mean values per group and condition are displayed with grey diamonds. Words with complex onsets are in the left two columns and words with a simple onset are in the right two columns.*

left-edge, and right edge of $W_{complex}$ with the $W_{simple}$ interval (recall that the latter was the same across intervals; see *2.4.1.3 Intervals (acoustic c-center)*). The results are displayed in Fig. 9. For this analysis, however, we had to exclude the word pair *Klang-lang*, since the onset of /k/ could not be determined based on the acoustic signal. Thus, the results on this section only include 3 word pairs.

Note that for $W_{simple}$, the interval for c-center, left-edge, and right-edge is the same. To examine whether the duration of the interval differs between $W_{simple}$ and $W_{complex}$, whether the groups differ in interval duration, and whether the metronome affects the duration of the intervals, we performed a linear mixed effects analysis for each interval separately. Therefore, the dependent variable varied with respect to the interval (duration of c-center, left-edge or right-edge interval).

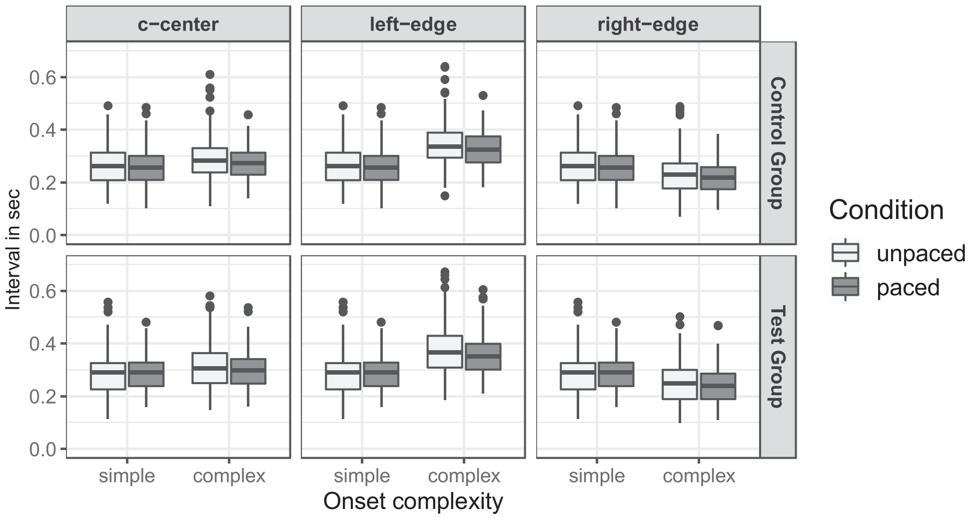


**Fig. 9.** Intervals in sec to the vowel offset anchor point for c-center, left-edge, and right-edge (columns) per condition (unpaced = beige, paced = brown) for different onset complexity (x-axis), separated by group (rows). Within each box, the median is denoted with black lines; boxes extend from the 25th to the 75th percentile of each group's distribution of values; the ends of the whiskers denote 1.5 interquartile range beyond the 25th and 75th percentile of each group; dots display observations outside the range of whiskers.

*3.4.1.1. C-center interval.* The model that predicted the duration of the c-center interval had a conditional $R^2$ of 0.812, and a marginal $R^2$ of 0.042. From this low marginal $R^2$ it can be concluded that the fixed effects do not explain much variation. Since the conditional $R^2$ value is quite high in comparison to the marginal $R^2$, it can be assumed that most of the variance can be explained with the random effects. A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 4: C-center*).

Nonetheless, the model did show that GROUP ($F[1, 89.79] = 10.23$, $p = .002$), ONSET COMPLEXITY ($F[1, 89.06] = 46.5$, $p < .001$), and CONDITION ($F[1, 91.75] = 5.81$, $p = .02$) were significant predictors of the c-center interval duration, reflecting the speech rate effects found in relation to word duration (see Result Section 3.1 Word duration). Moreover, the model revealed a significant interaction between ONSET COMPLEXITY and CONDITION ($F[1, 1944.93] = 12.30$, $p < .001$). Pairwise comparisons showed that the c-center interval did differ significantly between $W_{simple}$ and $W_{complex}$ in the unpaced condition ($p < .001$), and the paced condition ($p < .001$), whereby $W_{complex}$ intervals were longer than $W_{simple}$ intervals. While speaking along with a metronome did not affect the c-center interval in $W_{simple}$, the interval became significantly shorter in the paced condition in $W_{complex}$ ($p = .004$). Thus, the c-center interval did not differ between $W_{simple}$ that were produced in the unpaced condition and $W_{complex}$ that were produced in the paced condition ($p = 0.4410$) because the interval duration was similar. The results did not change when running the model without the five participants who had comorbidities.

*3.4.1.2. Left-edge interval.* A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 5: Left-edge*). The model that predicted the duration of the left-edge interval (Conditional $R^2 = 0.825$, Marginal $R^2 = 0.200$) revealed that the group who stutters had longer interval durations ($F[1, 89.18] = 10.84$, $p = .001$) and that $W_{complex}$ intervals were longer than $W_{simple}$ intervals ($F[1, 88.61] = 655.22$, $p < .001$), mirroring the speech rate differences. Moreover, the metronome significantly decreased interval duration ($F[1, 91.96] = 7.43$, $p = 0.008$). A significant interaction effect between ONSET COMPLEXITY and CONDITION was also found for this interval ($F[1, 1945.25] = 17.87$, $p < .001$). The results mirror those of the c-center interval. However, the pairwise comparisons showed that the intervals of $W_{complex}$ are always longer than those of $W_{simple}$ (when comparing $W_{simple}$ unpaced to $W_{complex}$ unpaced ($p < .001$), $W_{simple}$ unpaced to $W_{complex}$ paced ($p < .001$), and $W_{simple}$ paced to $W_{complex}$ unpaced ($p < .001$), as well when comparing $W_{simple}$ paced to $W_{complex}$ paced ($p < .001$)). The pacing condition did not significantly affect the left-edge interval of $W_{simple}$ but the $W_{complex}$ interval ($p < .001$). The results did not change when running the model without the participants who had comorbidities.

The basic pattern for this section is as would be expected for a c-center organization: the left-edge shifts left for words with a complex vs. a simple onset. The results also suggest that even though participants who stutter had longer left-edge intervals than participants who do not stutter, the groups did not differ in the difference between the $W_{simple}$ and $W_{complex}$, as there was no significant interaction between GROUP and ONSET COMPLEXITY.

*3.4.1.3. Right-edge interval.* A summary table with the estimates and confidence intervals of the model can be found in the Appendix (*A-Model 6: Right-edge*). The model to predict the right-edge interval duration (Conditional $R^2 = 0.833$, Marginal $R^2 = 0.0086$) revealed the same main effects as the models for the c-center and the left-edge interval. However, according to the marginal $R^2$ the fixed effects did not explain much of the variance. Nevertheless, GROUP ($F[1, 90.66] = 9.32$, $p = .003$), ONSET COMPLEXITY ($F[1, 90.03] = 298.87$, $p < .001$), and CONDITION ($F[1, 91.27] = 4.09$, $p = .046$) were significant predictors of the right-edge interval duration pointing towards longer intervals in the group who stutters, longer intervals in $W_{simple}$, and shorter intervals in the pacing condition (as expected due to the speech rate effects). Furthermore, a significant interaction was found between ONSET COMPLEXITY and CONDITION ($F[1, 1944.88] = 6.50$, $p = .01$). Pairwise comparisons also showed the same effects but this time, the intervals of $W_{simple}$ were longer than those of $W_{complex}$ (when comparing $W_{simple}$ unpaced to $W_{complex}$ unpaced ($p < .001$), $W_{simple}$ unpaced to $W_{complex}$ paced ($p < .001$), $W_{simple}$ paced to $W_{complex}$ unpaced ($p < .001$), and $W_{simple}$ paced to $W_{complex}$ paced ($p < .001$)). Furthermore, the interval of $W_{complex}$ became shorter in the paced condition ($p = .03$) but the metronome did not affect the interval duration of the $W_{simple}$ interval. When running the model without the five participants who had comorbidities, CONDITION was no longer a significant predictor of the right-edge interval duration. The other results of the main effects did not change. In general, the results suggest that the pacing condition had the least impact on the right-edge interval.

Taking stock of the interval-based measures reported in this section it can be concluded that $W_{complex}$ c-center intervals were the closest to the $W_{simple}$ intervals, pointing towards a complex syllable onset organization in both groups. C-center does shift slightly (about 18 ms) from simple to complex onset, but the shifts for right and left edge are larger (39 ms and 75 ms).

*3.4.2. RSDs*

The RSD displays the variability of an interval over word pairs, such that the lower the RSD, the lower the variability. A c-center organization would be displayed in a more stable (smaller) RSD for the c-center interval. The following tables display the RSD for the different intervals and the number of word pairs (n) produced per group. The tables are separated by condition.

As can be seen in Table 4 and Table 5, RSD was the lowest in the c-center interval in both groups and in both conditions, indicating that the c-center interval was the most stable one, thus supporting a c-center organization. However, RSD decreased in the paced condition in both groups, pointing towards more stability in the metronome condition. Furthermore, the RSD for all word pairs in the paced condition are less variable in participants who stutter, compared to the control group, indicating that the test group benefits more from the paced condition. While the control group improved stability by only 1.94% in the word pair *Schlamm-Lamm* and 1.84% in *Schlauch-Lauch*, the test group improved stability by almost 4.8% in *Schlamm-Lamm* and 5.3% in *Schlauch-Lauch*. For PWNS, the greatest improvement due to the paced condition in stability happened for the word pair *Schleim-Leim* where they improved by 4.88%.

For the same word pair, PWS improved by 5.01%. The biggest difference between groups in RSD was in the unpaced condition for the word pair *Schlamm-Lamm*. The group who stutters was 3.35% less stable than the control group. In the word pair *Schleim-Leim* PWS were even less variable than PWNS in the paced condition (by 0.97%).

In order to test whether the groups differed significantly in the stability of the relevant intervals and if the pacing condition had an effect on stability, a linear mixed effects model was run to predict RSD. The model (Conditional $R^2 = 0.421$, Marginal $R^2 = 0.110$) revealed that INTERVAL ($F[1, 1455.23] = 130.46$, $p < .001$) and CONDITION ($F[1, 84.06] = 4.95$, $p = .009$) were significant predictors of the RSD, indicating that left-edge and right-edge intervals had a higher RSD compared to the c-center interval and that the paced condition reduced variability significantly. Furthermore, significant interactions were found between GROUP and INTERVAL ($F[2, 1455.23] = 4.95$, $p = .007$) and CONDITION and INTERVAL ($F[2, 1455.23] = 4.31$, $p = .01$). Post-hoc tests showed that the groups did not differ significantly in interval stability of the same intervals. However there were within group differences, namely that the variability of the left-edge interval did not differ from the variability of the right-edge interval in the group who stutters, while in the control group, there was a significant difference in RSD between these two intervals ($p < .001$), pointing towards more variability in the left-edge interval compared to the right-edge interval. The post hoc test regarding the significant interaction between CONDITION and INTERVAL showed that the metronome only significantly improved stability in the left-edge interval ($p = .02$), nearly significantly in the c-center interval ($p = .054$), and not significantly in the right-edge interval ($p = .096$).

Hence, it can be concluded that the variability was the lowest in the c-center interval, pointing towards a c-center organization. Moreover, results suggest that PWS did not differ from PWNS in terms of variability and that the paced condition only increased stability in the left-edge interval but not in the c-center, nor the right-edge interval.

## 4. Discussion

The aim of the present study was to analyze temporal organization of syllables in children and adolescents who stutter and children and adolescents who do not stutter, speaking with and without an external rhythm (metronome). Therefore, participants were asked to read wordlists which contained German monosyllabic words that differed in onset complexity in their own preferred tempo (unpaced condition) and along with a metronome (paced condition). Four minimal pairs (*Klang-lang*, *Schlamm-Lamm*, *Schlauch-Lauch*, *Schleim-Leim*) were analyzed. We focused on syllabic timing related to articulatory timing and analyzed acoustic cues of a *c-center effect* for which our participants indeed showed evidence. Examining the c-center effect in PWS is particularly interesting, as it can provide evidence for difficulties in articulatory control. This study presents novel findings since this is the first study (to our knowledge) that examines these effects in a population who stutters.

We found consonant and vowel compression effects in both groups, as well as support for a complex syllable onset organization as indicated by lower relative standard deviations in c-center intervals compared to left-edge or right-edge intervals. These results point towards a complex syllable onset organization in German children and adolescents who stutter and German children and adolescents who do not stutter, indicating a shift of the rightmost consonant in an onset cluster towards the vowel.

Moreover, speaking along with a metronome did not affect compression effects in consonants, but there was more vowel compression in the paced condition. Furthermore, speaking along with a metronome improved durational interval stability. A group difference was observed with respect to consonant compression indicating that children and adolescents who stutter showed a bigger difference between [l] in words with a simple onset and words with a complex onset than the control group when all participants were included in the analyses.

Our results indicate that the groups do not differ in general (articulatory) syllable organization in perceptually fluent speech, supporting hypothesis (1) and (2), as we did find both consonant and vowel compression in the group who stutters and the control group. These results point towards a c-center organization in syllable articulation. As hypothesized, participants who stutter differed from participants who do not stutter in consonant compression (but not in vowel compression) which suggests that children and adolescents who stutter time onset consonant gestures differently.

According to neurophonetic models of stuttering, such as the GODIVA model (Civier, Bullock, Max & Guenther, 2013), the initiation of the articulatory gestures within a syllable organization is atypical in PWS and thus may lead to stuttering symptoms. Hence, the timing of consonantal onset gestures seems to be particularly challenging for PWS. However, it should be mentioned that the GODIVA model specifically addresses stuttering events while the present study focuses on fluent speech only. In the GODIVA model, stuttering events are interpreted as failures to activate the next syllable's motor program in time (Civier et al., 2013). The neural circuit involved in initiating and terminating syllables consists of basal ganglia, thalamus, and left ventral premotor cortex (Civier et al., 2013). Toyomura and colleagues (2015) found that PWS's basal ganglia activity (which is an indication of motor control)

**Table 4**
RSD in the unpaced condition.

| Word pair | n | Group | RSD c-center | RSD left-edge | RSD right-edge |
|---|---|---|---|---|---|
| Schlamm-Lamm | 198 | PWNS | **19.79** | 25.70 | 32.22 |
| Schlauch-Lauch | 193 | PWNS | **19.78** | 21.94 | 25.20 |
| Schleim-Leim | 194 | PWNS | **22.30** | 24.85 | 27.82 |
| Schlamm-Lamm | 171 | PWS | **23.15** | 27.49 | 31.61 |
| Schlauch-Lauch | 160 | PWS | **21.01** | 23.71 | 26.80 |
| Schleim-Leim | 176 | PWS | **21.33** | 24.09 | 26.88 |

**Table 5**
RSD in the paced condition.

| Word pair | n | Group | RSD c-center | RSD left-edge | RSD right-edge |
|---|---|---|---|---|---|
| Schlamm-Lamm | 192 | PWNS | **17.85** | 22.90 | 29.99 |
| Schlauch-Lauch | 188 | PWNS | **17.90** | 19.66 | 23.15 |
| Schleim-Leim | 190 | PWNS | **17.42** | 19.97 | 23.53 |
| Schlamm-Lamm | 172 | PWS | **17.48** | 21.96 | 26.46 |
| Schlauch-Lauch | 174 | PWS | **15.69** | 17.33 | 20.22 |
| Schleim-Leim | 177 | PWS | **16.32** | 18.58 | 21.75 |

increased to the level of PWNS's after practicing to speak along with a metronome over a period of 8 weeks for at least 15 min per day and at least 5 days per week. In our study, the significant difference between groups in consonant compression underlines differences in motor control, particularly gestural timing of syllable onsets. The group difference was present regardless of speaking with or without a metronome, which indicates that differences in motor control might even be present in a fluency-enhancing condition. A future study could analyze the temporal syllabic structure in long-term fluency-enhancing effects. More specifically, it could be investigated whether the group difference regarding consonant compression would be cancelled out after a period of 8 weeks regular practice of speaking along with a metronome. As Toyomura et al. (2015) showed, basal ganglia activity did not differ between PWS and PWNS after this period of time, suggesting that this had led to similar syllable initiation patterns.

Since the groups did not differ in vowel compression, but children and adolescents who stutter showed more consonant compression than the control group, it points towards more gestural overlap between the right-most consonant in an onset cluster and the following vowel in the group who stutters. This could support Harrington's (1988) model of stuttering in which he suggests that stuttered speech occurs because individuals who stutter incorrectly apply their perceptual predictive timing to their own speech production output. According to his theoretical viewpoint, PWS expect the time of sensory feedback to occur earlier than it actually does and, thereby, they would erroneously correct for the moment of their actual segmental production; this then would lead to stuttering because the articulatory movements are too much in conflict with each other (e.g., simultaneous instruction to close the lips and to lower the jaw) (Harrington, 1988). Higher overlap between [l] and the following vowel in words with a complex onset could indicate altered predictive timing mechanisms, resulting in an atypical inter-gestural timing. However, Harrington's theory remains to be tested for syllables with different onset complexities since the model is mainly based on the coupling of one onset consonant and the following vowel.

If PWS would coarticulate less, for instance as a strategy to speak more fluently (e.g., Zmarich et al., 2013), this would have led to less consonantal compression in the group who stutters. Another articulatory explanation that would lead to a greater consonant compression is shortening of the [l] gesture in words with a complex onset. This could be either a consequence of altered articulatory timing in PWS or even a strategy for PWS to speak more fluently. Conducting an articulatory study, using for instance electromagnetic articulography, could clarify whether more overlapping between the right-most consonant gesture and the vowel gesture causes more consonant compression in PWS, or whether it is the shortening of the second consonant gesture in a complex onset (CC) that leads to more compression in the group who stutters.

With respect to our results, we should keep in mind that the group difference was not very strong as consonant compression did not differ between groups anymore when the five participants who stutter with comorbidities (Dyslexia, ADHD) were excluded. However, including the five participants who had comorbidities displays a more realistic reflection of the population who stutters as more than 60% of children who stutter have co-occurring speech, language, or non-speech-language disorders, as a study with 2628 American children revealed (Blood, Ridenour, Qualls & Scheffner Hammer, 2003). Non-speech-language disorders, which affect around 34.3% of the children, include for instance attention deficit disorders (5.9% of the 34.3%) and literacy disorders (8.2% of the 34.3%) (Blood et al., 2003). In our view, an articulatory study (i.e., using articulography) will clarify the strength of consonant compression in young persons who stutter. In addition, it would be especially interesting to look into the phenomenon that participants who produced long [l]-durations in words with a simple onset, produced shorter [l]-durations in words with a complex onset. These were individuals who displayed greater consonant compression, which was observed independently of group.

The result of more consonant compression in the group who stutters suggests that stuttering may be related to issues with the coordination of speech motor movements, specifically with the timing and sequencing of the movements required for producing onset consonants. This finding may also be relevant for the neurosciences studying the underlying neural mechanisms of stuttering. For instance one might further investigate how the basal ganglia-thalamo-cortical circuit is involved in initiating syllables with different onset complexity in different rhythmic conditions. Clinically, the knowledge of increased consonant compression in individuals who stutter may inform the development of targeted therapy interventions. For example, speech therapists currently teach techniques that prolong the onset of syllables to modify or prevent stuttered disfluencies. It could be a complementary avenue for clinical research to explore the efficiency of techniques that specifically target the coordination of speech motor movements in fluent speech, such as training paced and unpaced fluent speech over a longer period of time (like in the study by Toyomura et al., 2015), to foster objective and subjective articulatory control in individuals who stutter.

Furthermore, contrary to our hypothesis, children and adolescents who stutter were as stable in their syllable organization as children and adolescents who do not stutter since the groups did not differ significantly in interval stability. This means that the group who stutters and the control group had similar durational variability in the c-center to anchor, left-edge to anchor, and right-edge to anchor intervals.

Concerning the effect of a regular rhythm on the temporal organization of speech, speaking along with a metronome did not affect consonant compression but it enlarged vowel compression. This result can be interpreted in the light of the effect that vowel durations in particular are affected by changes in speech rate in PWS (e.g., Davidow, 2014). In the paced condition, participants may have increased their speech rate in words with a complex onset even more than in words with a simple onset, as they had one sound more to produce and wanted to be in time with the metronome. Hence, vowels in words with a complex onset would become even shorter in the paced condition. In this way, paced speech would not only affect speech rate but also vowel timing patterns. To clarify rate variations and their interaction with syllable timing, a future study could focus on investigating compression effects in longer utterances consisting of multiple syllables. Syllable durations will be produced according to the needs of the utterance, where any particular syllable could be expanded or compressed according to those needs. This would allow the study of consonant and vowel compression in context, as well as the contribution of rate variation.

Moreover, the paced condition reduced variability of durational intervals, matching other studies that found reduced variability in fluency-enhancing conditions, such as metronome-paced speech and singing (e.g., a decreased variability of duration of voiced and voiceless segments in metronome-paced speech [Janssen & Wieneke, 1987] and a decreased variability in voice onset time in word-initial stressed positions when PWS were singing [Falk, Maslow, Thum & Hoole, 2016]). In metronome-timed speech, we found reduced variability especially in the left-edge interval which points towards a reduction in durational variability of the fricative [ʃ]. In addition, the hypothesis that the group who stutters benefits more from the paced condition leading to a higher increase in stability could be confirmed.

The study had limitations with respect to the number and type of word pairs. Our results are mostly based on fricative-lateral [ʃl] onsets in words with a complex onset and on the lateral [l] in words with a simple onset. However, word pairs that differ in onset complexity are not easy to find, especially with the limitation of an acoustic analysis; for instance, onsets with a plosive in $C_2$ position in words with a complex onset had to be excluded since these would be the onset consonant in words with a simple onset and hence, plosive onsets cannot be determined based on the acoustic signal only. Therefore, an articulatory study would allow to include more diverse word pairs. Onsets that start with a plosive might be more difficult to initiate for PWS than onsets that start with a sibilant or lateral. For English it was found that consonant manner and consonant place were predictors of stuttering rate (Howell, Au-Yeung, Yaruss & Eldridge, 2006). The first consonant in a cluster could therefore also impact the following consonant(s). This can be addressed in future work where more word pairs of different cluster combinations should be included. Another aspect that should be taken into account when analyzing the *c-center effect* in the future is that c-center stability was found to be more frequent in words that contain tense vowels than lax vowels or diphthongs (Brunner et al., 2014).

Finally, as in other studies on stuttering in a school-age or older population, we had more male than female participants. This imbalance reflects a general bias in the population with persistent stuttering or with a risk of persistence (Yairi & Ambrose, 2013). Our participants were more likely to show persistent stuttering as they were 9 years and older. Hence, we cannot entirely exclude that the articulatory effects found in the current study are more visible in the male population. For example, it is known from the literature on younger children between the ages of 4 years and 5 years and 11 months that boys who stutter show greater lags in speech motor development than girls who stutter, compared to their peers (Walsh, Mettel, & Smith, 2015). Moreover, there are differences in speech-related brain regions between boys and girls who stutter between the ages of 3 and 10 years, possibly contributing to the fact that girls are more likely to recover (Chang, Zhu, Choo & Angstadt, 2015). Overall, it should be an aim for future research to better understand individual patterns (for example age, education, reading skills) in fluent and disfluent speech of stuttering – a general need for studies on rare populations.

In sum, this study is a promising basis for conducting an articulatory study in which articulatory (gestural) timing can be examined in more detail. This is an important step in the investigation of motor control in stuttering and it will greatly help to understand the underlying motor patterns in PWS.

## 5. Conclusion

The results of our study suggest that the temporal organization of the syllable is similar in children and adolescents who stutter vs. children and adolescents who do not stutter, regardless of speaking in their own preferred speech tempo or along with a metronome. Moreover, paced speech improved durational interval stability in both groups. However, the group who stutters produced more consonant compression than the control group, suggesting differences in articulatory onset timing.

## Funding

This work was supported by DFG grants [FA 901/4–1 and HO 3271/6–1], the Graduate School "Class of Language", Ludwig Maximilian University of Munich, and the German Academic Exchange Service (DAAD). It was also supported by the Canada Research Chair on Interdisciplinary studies on rhythm and language acquisition (SSHRC), Canada.

## Declaration of Competing Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Data Availability

Data will be made available on request.


## Acknowledgements

We thank all our participants for their participation, Ingeborg Meyer, Georg Thum, and the entire SAS team for their help with recruiting, as well as Michele Gubian for his support with statistics.


## Appendix

Description of Predictors that can be found in the following models:
group[*s*]: PWS.
group[*c*]: PWNS.
nonset[2]: $W_{complex}$.
nonset[1]: $W_{simple}$.
met[1]: paced condition.
met[0]: unpaced condition.
contrastpair: word pair.
A-Model 1: Word duration.

| | Word_dur | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.45 | 0.38 – 0.51 | **< 0.001** |
| group [s] | 0.06 | 0.03 – 0.10 | **< 0.001** |
| nonset [2] | 0.07 | -0.02 – 0.16 | 0.131 |
| met [1] | -0.01 | -0.03 – 0.01 | 0.158 |
| group [s] * nonset [2] | -0.01 | -0.02 – 0.01 | 0.267 |
| group [s] * met [1] | -0.00 | -0.03 – 0.02 | 0.834 |
| nonset [2] * met [1] | -0.02 | -0.03 – -0.00 | **0.008** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.01 | | |
| $\tau_{00}$ Word | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\tau_{11}$ Word.met1 | 0.00 | | |
| $\rho_{01}$ participant.nonset2 | -0.20 | | |
| $\rho_{01}$ participant.met1 | -0.80 | | |
| $\rho_{01}$ Word | -0.92 | | |
| ICC | 0.72 | | |
| $N$ participant | 96 | | |
| $N$ Word | 8 | | |
| Observations | 2909 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.130 / 0.758 | | |

*A-Model 2: l duration.*

| | onset_comp | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.10 | 0.09 – 0.11 | **< 0.001** |
| group [s] | 0.02 | 0.01 – 0.03 | **0.002** |
| nonset [2] | -0.04 | -0.05 – - 0.03 | **< 0.001** |
| met [1] | -0.00 | -0.01 – 0.00 | 0.382 |
| group [s] * nonset [2] | -0.01 | -0.02 – 0.00 | **0.038** |
| group [s] * met [1] | -0.00 | -0.01 – 0.01 | 0.855 |
| nonset [2] * met [1] | -0.00 | -0.01 – 0.00 | 0.404 |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.00 | | |

(*continued on next page*)

(*continued*)

| | onset_comp | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| $\tau_{00}$ contrastpair | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\rho_{01}$ participant.met1 | -0.55 | | |
| $\rho_{01}$ participant.nonset2 | -0.85 | | |
| ICC | 0.34 | | |
| N participant | 96 | | |
| N contrastpair | 4 | | |
| Observations | 2909 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.281 / 0.528 | | |

*A-Model 3: Vowel duration.*

| | Vowel_dur | | |
|---|---|---|---|
| Predictors | Estimates | CI | p |
| (Intercept) | 0.20 | 0.15 – 0.25 | **< 0.001** |
| group [s] | 0.01 | 0.00 – 0.03 | 0.075 |
| nonset [2] | -0.02 | -0.02 – - 0.01 | **< 0.001** |
| met [1] | -0.00 | -0.01 – 0.01 | 0.557 |
| group [s] * nonset [2] | -0.00 | -0.01 – 0.00 | 0.626 |
| group [s] * met [1] | 0.00 | -0.01 – 0.01 | 0.716 |
| nonset [2] * met [1] | -0.00 | -0.01 – -0.00 | **0.020** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.00 | | |
| $\tau_{00}$ contrastpair | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\rho_{01}$ participant.met1 | -0.61 | | |
| $\rho_{01}$ participant.nonset2 | -0.35 | | |
| ICC | 0.84 | | |
| N participant | 96 | | |
| N contrastpair | 4 | | |
| Observations | 2909 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.031 / 0.848 | | |

*A-Model 4: C-center.*

| | center | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.27 | 0.20 – 0.33 | **< 0.001** |
| group [s] | 0.02 | 0.00 – 0.05 | **0.016** |
| nonset [2] | 0.02 | 0.02 – 0.03 | **< 0.001** |
| met [1] | -0.01 | -0.02 – 0.01 | 0.353 |
| group [s] * nonset [2] | -0.00 | -0.01 – 0.01 | 0.775 |
| group [s] * met [1] | 0.00 | -0.01 – 0.02 | 0.818 |
| nonset [2] * met [1] | -0.01 | -0.02 – - 0.00 | **< 0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.00 | | |
| $\tau_{00}$ contrastpair | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\rho_{01}$ participant.nonset2 | -0.19 | | |
| $\rho_{01}$ participant.met1 | -0.72 | | |
| ICC | 0.80 | | |
| N participant | 96 | | |
| N contrastpair | 3 | | |
| Observations | 2185 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.042 / 0.812 | | |

*A-Model 5: Left-edge.*

| | left_edge | | |
|---|---|---|---|
| Predictors | Estimates | *CI* | *p* |
| (Intercept) | 0.27 | 0.20 – 0.33 | **< 0.001** |
| group [s] | 0.03 | 0.01 – 0.05 | **0.014** |
| nonset [2] | 0.08 | 0.07 – 0.09 | **< 0.001** |
| met [1] | -0.00 | -0.02 – 0.01 | 0.435 |
| group [s] * nonset [2] | 0.00 | -0.01 – 0.01 | 0.667 |
| group [s] * met [1] | 0.00 | -0.02 – 0.02 | 0.990 |
| nonset [2] * met [1] | -0.01 | -0.02 – - 0.01 | **< 0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.00 | | |
| $\tau_{00}$ contrastpair | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\rho_{01}$ participant.nonset2 | 0.03 | | |
| $\rho_{01}$ participant.met1 | -0.75 | | |
| ICC | 0.78 | | |
| N participant | 96 | | |
| N contrastpair | 3 | | |
| Observations | 2185 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.200 / 0.825 | | |

*A-Model 6: Right-edge.*

| | right_edge | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.27 | 0.20 – 0.33 | **< 0.001** |
| group [s] | 0.02 | 0.00 – 0.04 | **0.020** |
| nonset [2] | -0.03 | -0.04 – - 0.03 | **< 0.001** |
| met [1] | -0.01 | -0.02 – 0.00 | 0.284 |
| group [s] * nonset [2] | -0.01 | -0.01 – 0.00 | 0.266 |
| group [s] * met [1] | 0.00 | -0.01 – 0.02 | 0.634 |
| nonset [2] * met [1] | -0.01 | -0.01 – - 0.00 | **0.011** |
| **Random Effects** | | | |
| $\sigma^2$ | 0.00 | | |
| $\tau_{00}$ participant | 0.00 | | |
| $\tau_{00}$ contrastpair | 0.00 | | |
| $\tau_{11}$ participant.nonset2 | 0.00 | | |
| $\tau_{11}$ participant.met1 | 0.00 | | |
| $\rho_{01}$ participant.nonset2 | -0.41 | | |
| $\rho_{01}$ participant.met1 | -0.70 | | |
| ICC | 0.82 | | |
| N participant | 96 | | |
| N contrastpair | 3 | | |
| Observations | 2185 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.086 / 0.833 | | |

## References


Andrews, G., Howie, P., Dozsa, M., & Guitar, B. (1982). Stuttering: Speech pattern characteristics under fluency-inducing conditions. *Journal of Speech, Language and Hearing Research, 25*, 208–216.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blood, G. W., Ridenour, V. J., Qualls, C. D., & Scheffner Hammer, C. (2003). Co-occurring disorders in children who stutter. *Journal of Communication Disorders, 36*(1), 427–448. https://doi.org/10.1016/S0021-9924(03)00023-6

Boersma, P., & Weenink, D.Praat (2019): Doing Phonetics by Computer [Computer Program]. Version 6.1, 2019. Available online: http://www.praat.org/.

Browman, C., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory. *Phonology Phonetica, 45*, 140–155.

Browman, C. P., & Goldstein, L. (1991). Gestural structures: distinctiveness, phonological processes, and historical change. In I. Mattingly, & M. Studdert-Kennedy (Eds.), *Modularity and the Motor Theory of Speech Perception* (pp. 313–338). New Jersey: Erlbaum.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*, 155–180.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin Délelőtt Louisiana Communication Parlée, 5*, 25–34.

Brunner, J., Geng, C., Sotiropoulou, S., & Gafos, A. (2014). Timing of German onset and word boundary clusters. *Laboratory Phonology, 5*(4), 403–454.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics, 24*, 209–244.

Chang, S. E., Zhu, D. C., Choo, A. L., & Angstadt, M. (2015). White matter neuroanatomical differences in young children who stutter. *Brain: a Journal of Neurology, 138*(Pt 3), 694–711. https://doi.org/10.1093/brain/awu400

Civier, O., Bullock, D., Max, L., & Guenther, F. H. (2013). Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. *Brain and Language, 126*(3), 263–278. https://doi.org/10.1016/j.bandl.2013.05.016

Davidow, J. H. (2014). Systematic studies of modified vocalization: The effect of speech rate on speech production measures during metronome-paced speech in persons who stutter. *International Journal of Language & Communication Disorders, 49*(1), 100–112. https://doi.org/10.1111/1460-6984.12050. PMID: 24372888; PMCID: PMC4461240.

De Nil, L. F., & Brutten, G. J. (1991). Voice onset times of stuttering and nonstuttering children: The influence of externally and linguistically imposed time pressure. *Journal of Fluency Disorders, 16*, 143–158.

Di Simoni, F. G. (1974). Letter: preliminary study of certain timing relationships in the speech of stutterers. *The Journal of the Acoustical Society of America, 56*(2), 695–696.

Dokoza, K. P., Hedever, M., & Sarić, J. P. (2011). Duration and variability of speech segments in fluent speech of children with and without stuttering. *Collegium Antropologicum, 35*(2), 281–288.

Falk, S., Maslow, E., Thum, G., & Hoole, P. (2016). Temporal variability in sung productions of adolescents who stutter. *Journal of Communication Disorders, 62*, 101–114.

Gibson, M., Fernández Planas, A.M., Gafos, A., & Remirez, E. (2015). Consonant duration and VOT as a function of syllable complexity and voicing in a subset of Spanish clusters. Interspeech 2015. [online] https://www.ling.uni-potsdam.de/~gafos/papers/Gibson-Gafos-Interspeech2015.pdf (Accessed 9 September 2022).

Goldstein, L., & Pouplier, M. (2014). The temporal organization of speech. In F. M. Goldrick, & M. Miozzo (Eds.), *The Oxford Handbook of Language Production* (pp. 210–227). Oxford: Oxford University Press.

Hall, N. (2010). Articulatory phonology. *Language and Linguistics Compass, 4*(*9*), 818–830. https://doi.org/10.1111/j.1749-818x.2010.00236.x

Harrington, J. M. (1987). Coarticulation and stuttering: an acoustic and electropalatographic study. In H. Peters, & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 381–392). New York: Springer Verlag.

Harrington, J. M. (1988). Stuttering, delayed auditory feedback, and linguistic rhythm. *Journal of Speech and Hearing Research, 31*, 36–47.

Heyde, C. J., Scobbie, J. M., Lickley, R., & Drake, E. K. E. (2016). How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. *Clinical Linguistics & Phonetics, 30*(3–5), 292–312. https://doi.org/10.3109/02699206.2015.1100684

Hoole, P., & Pouplier, M. (2015). Interarticulatory coordination - speech sounds. ch. 7. In M. Redford (Ed.), *The Handbook of Speech Production* (pp. 133–157). Hoboken, New Jersey: John Wiley & Sons. ch. 7.

Howell, P., Au-Yeung, J., Yaruss, J. S., & Eldridge, K. (2006). Phonetic difficulty and stuttering in English. *Clinical Linguistics & phonetics, 20*(9), 703–716. https://doi.org/10.1080/02699200500390990

Janssen, P., & Wieneke, G. (1987). The effects of fluency inducing conditions on the variability in the duration of laryngeal movements during stutterer's fluency speech. In H. F. M. Peters, & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 337–344). New York: Springer-Verlag.

Jäncke, L. (1994). Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders, 19*(1), 21–37.

Katz, J. (2010). Compression effects, perceptual asymmetries, and the grammar of timing. Dissertation www.researchgate.net/publication/265231502, (Accessed 9 September 2022).

Katz, J. (2012). Compression effects in English. *Journal of Phonetics, 40*, 390–402. https://doi.org/10.1016/j.wocn.2012.02.004

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.2–1.https://CRAN.R-project.org/package=emmeans.

MacMillan, V., Kokolakis, A., Sheedy, S., & Packman, A. (2014). End-Word dysfluencies in young children: A clinical report. *Folia Phoniatrica et Logopaedica, 66*, 115–125. https://doi.org/10.1159/000365247

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control, 14*(3), 380–407.

Marin, S., & Bučar Shigemori, L.S. (2014). Vowel compensatory shortening in Romanian. [online], https://www.phonetik.uni-muenchen.de/universals/pub/ISSP2014_VCS_Marin_BucarShigemori.pdf (Accessed 23 September 2022).

Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech Language, and Hearing Research, 46*(1), 215–232. https://doi.org/10.1044/1092-4388(2003/017)

Max, L., & Gracco, V. L. (2005). Coordination of oral and laryngeal movements in the perceptually fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research, 48*(June), 524–542.

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator of syllable structure. *Proc XVth International Congress of Phonetic Sciences, Barcelona*, 3-9 Aug 2003.

Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders, 33*, 32–51.

Peters, S., & Kleber, F. (2014). Articulatory mechanisms underlying incremental compensatory vowel shortening in German. ISSP 2014. [online] https://www.phonetik.uni-muenchen.de/personen/mitarbeiter/kleber_felicitas/publikationen/peters_kleber14_issp.pd (last accessed: 09/23/22).

Pouplier, M. (2012). The gestural approach to syllable structure: universal, language- and cluster-specific aspects. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 63–96). Berlin: Peter Lang.

R. Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ruthan, M.Q., Durvasula, K., & Lin, Y.-H. (2019). Temporal coordination and sonority of Jazani Arabic word-initial clusters. Proceedings of the Annual Meetings on Phonology, 7, 10.3765/amp.v7i0.4485.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. (2011). Dynamic invariance in the phonetic expression of syllable structure: A case study of Moroccan Arabic consonant clusters. *Phonology, 28*(3), 455–490. https://doi.org/10.1017/S0952675711000224

Selkirk, E., & Durvasula, K. (2013). Acoustic correlates of consonant gesture timing in English. *The Journal of the Acoustical Society of America, 134*(5), 4202-4202.

Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders, 35*, 1–18.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of Fluency Disorders, 37*(4), 344–358. https://doi.org/10.1016/j.jfludis.2012.06.001

Smith, A., & Weber, C. (2016). Childhood stuttering: Where are we and where are we going. *Seminars in Speech and Language, 37*(4), 291–297. https://doi.org/10.1055/s-0036-1587703

Stager, S. V., Jeffries, K. J., & Braun, A. R. (2003). Common features of fluency-evoking conditions studied in stuttering subjects and controls: An H215O PET study. *Journal of Fluency Disorders, 28*, 219–336.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science, 33*, 839–879.

Toyomura, A., Fuji, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. *Neuroimage, 57*, 1507–1516.

Toyomura, A., Fujii, T., & Kuriki, S. (2015). Effect of an 8-week practice of externally triggered speech on basal ganglia activity of stuttering and fluent speakers. *NeuroImage, 109*, 458–468. https://doi.org/10.1016/j.neuroimage.2015.01.024

Usler, E., Smith, A., & Weber, C. (2017). A lag in speech motor coordination during sentence production is associated with stuttering persistence in young children. *Journal of Speech, Language, and Hearing Research, 60*(1), 51–61. https://doi.org/10.1044/2016_JSLHR-S-15-0367

Usler, E. R., & Walsh, B. (2018). The effects of syntactic complexity and sentence length on the speech motor control of school-age children who stutter. *Journal of Speech, Language, and Hearing Research, 61*, 2157–2167. https://doi.org/10.1044/2018_JSLHR-S-17-0435

Van Lieshout, P. H. H. M., & Namasivayam, A. K. (2010). Speech motor variability in people who stutter. In B. Maassen, & P. H. H. M. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 191–214). Oxford, England: Oxford University press.

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of neurodevelopmental Disorders, 7*(1), 27. https://doi.org/10.1186/s11689-015-9123-8

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Müller, K., Ooms, J., Robinson, D., Paige Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiltshire, C. E. E., Chiew, M., Chesters, J., Healy, M., & Watkins, K. E. (2021). Speech movement variability in people who stutter: A coval tract magnetic resonance imaging study. *Journal of Speech, Language, and Hearing Research, 64*, 2438–2452.

Wingate, M. E. (1969). Sound and pattern in "artificial" fluency. *Journal of Speech Language and Hearing Research, 12*, 677–686.

Wingate, M. E. (1988). *The Structure of Stuttering (a Psycholinguistic Analysis)*. New-York, NY: Springer Verlag.

World Health Organization (WHO) (2015). The ICD-10 Classification of Mental and Behavioral Disorders; F98.5 Stuttering; Geneva, Switzerland: WHO.

Yairi, E., & Ambrose, N. G. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders, 38*, 66–87.

Zmarich, C., Balbo, D., Galatà, V., Verdurand, M., & Rossato, S. (2013). The production of syllabes in stuttering adults under normal and altered auditory feedback. In V. Galatà (Ed.), *Multimodalità e Multilingualità: la Sfida più Avanzata della Comunicazione Orale* (9th ed..,, pp. 463–474). Rome, Italy: Bulzoni editore.

**Mona Franke** is a Ph.D. student in Phonetics at the Ludwig Maximilian University Munich, supervised by Dr. Philip Hoole, and a Ph.D. student in Linguistics at the University of Montreal, supervised by Dr. Simone Falk. She is exploring the role of rhythm in speech articulation and perception, as well as in communication.

**Dr. Philip Hoole** is Senior Lecturer at the Institute for Phonetics and Speech Processing, LMU Munich. His main research interests include linguistic phonetics, speech motor control, laryngeal articulation and auditory feedback processing. He has a special focus on instrumental techniques such as electromagnetic articulography and real-time magnetic resonance imaging.

**Dr. Simone Falk** is a professor in neurolinguistics at the University of Montreal. Since her PhD (in linguistics, LMU Munich), she is investigating the links between musical and verbal capacities, and in her recent projects, the role of rhythm and timing in stuttering.