



HAL
open science

Rhythm and Stuttering in Speech Perception and Production

Michela Russo, Fabrice Hirsch

► **To cite this version:**

Michela Russo, Fabrice Hirsch. Rhythm and Stuttering in Speech Perception and Production. Lars Meyer (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig) et Antje Strauss (University of Konstanz) (éd.). Rhythms of Speech and Language, Cambridge University Press (CUP), In press. hal-04430109

HAL Id: hal-04430109

<https://hal.science/hal-04430109v1>

Submitted on 9 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

SPEECH RHYTHM IN STUTTERING: PERCEPTION AND PRODUCTION

Michela Russo^{1,2} and *Fabrice Hirsch*³

¹ *CNRS, SFL UMR 7023, University of Paris 8, France*

² *University of Jean Moulin Lyon III, France*

³ *Paul Valéry University, Montpellier 3, France & CNRS Praxiling UMR 5267*

Corresponding authors:

m.russo@univ-paris8.fr

fabrice.hirsch@univ-montp3.fr

1. INTRODUCTION

1.1. THE NATURE OF RHYTHM AND ITS MEASURABILITY

'Rhythm' in the study of spoken language is used in many different contexts – for example, as an aesthetic property, as a manifestation of foreign accent, or as a feature of language typology (Hoeqvist 1983; Koreman et al. 2009; Barry et al. 2009). What is meant by 'rhythm' in these different contexts is primarily an auditory property¹.

Phonetic studies of rhythm have recently focused on measurable properties, more specifically on the durational characteristics of speech (Low 1998; Low & Grabe 1999; Low et al. 2001; Grabe et al. 1999; Grabe 2002; Deterding 2001; Gibbon & Gut 2001; Asu & Nolan 2005; Russo & Barry 2010; see also Fuchs in this volume).

This is immediately understandable and plausible in the light of the early, auditorily based statements. Additionally, the concept of an acoustic foundation underpinning auditory language-rhythm discrimination further supports this idea, which reduced rhythmic differences between languages to syllable-timed, that is, with a claim of approximately equal syllabic intervals, or syllabic isochrony, and stress-timed, that is, with a claim of roughly equal foot intervals and reduced syllable durations between the accented syllables (Lloyd 1940; Abercrombie 1965).

It is now generally accepted, that none of the many attempts to find a physical reflex of isochrony have been successful in the past (Dauer 1983, 1987; Bolinger 1965; Wenk & Wioland 1982; Roach 1982; Manrique & Signorini 1983; Eriksson 1992; Deterding 2001; Wagner 2010; Gibbon 2001). Rhythm is no

¹ See the original auditorily based observations by Lloyd (1940), who initially suggested the auditory distinction between languages characterized by a primarily Morse code rhythm (stress-timed) and those characterized by a primarily machine gun rhythm (syllable-timed) and by Pike (1945), who refers to syllable-timed vs. stress-timed languages; see discussions on the auditorily cues underlying rhythm distinction in Barry et al. (2009); Barry & Andreeva (2010); Lee et al. (2004); Dellwo (2008) on the the perceived rhythmic regularity/irregularity.

longer considered as a language primitive but rather as an emergent property, the product of phonological structure and phonetic realization. The shift from isochrony to variability led to the breakdown of the initial dichotomy in auditory perception, causing the disconnection of the syllable from its role as a fundamental unit of rhythm. The concept of syllable and foot regularity has been replaced by the degree of syllabic irregularity depending on the range of syllable-complexity with sub-syllabic durational measures as its acoustic basis. (for an additional perspective in neuroscience, consider Rimmele; Yuran Zhang et al.; Jerbi & Boulenger in this volume; also, for the language acquisition view, see Goswami; Mandke & Rocha in this volume).

Moving somewhat away from what had been defined in the past as syllabic isochrony in all languages, in the new approaches the measurements of durational variation seem to serve more successfully to separate rhythmic types (Ramus 1999; Ramus et al. 1999; Low 1998; Grabe & Low 2002; Barry et al. 2003; Russo & Barry 2004; 2008a/b; Dellwo 2006; Mok & Dellwo 2008).

Speech is rhythmically structured in time (Arvaniti 1994, 2009; Arvaniti & Jeon in this volume; Cutler 1994; Barry et al. 2003; Cummins 2009; Russo & Barry 2008a/b, 2010; Barry & Andreeva 2010; Wagner 2010). We expect a temporal regularity in the prominent syllables produced in natural, communicatively meaningful speech (see Goswami in this volume), given the multi-level nature of accentuation. This expectation holds true even if the acoustic basis of rhythm-carrying prominences does not show predictability in each production parameter (i.e., duration, fundamental frequency, intensity, and spectral definition). In a language without lexical stress, such as French (which features a phrase-final demarcative accent), supra-lexical information-based prominence is observed. It's notable that accentuation effects in French primarily occur at the ends of phrases or stress groups (see Barry & Andreeva 2010). This prominence disrupts the typical pattern of temporal regularity (isochrony) in syllable sequences. Moreover, in normal speech (with a lexical-stress system or without a lexical-stress system), the rhythm is still carried by the prominences within utterances, but we can rarely find a pattern of regular beats.

The rhythmic differences between language types should be audibly comprehensible, as well as quantitatively demonstrable. And if rhythm is part of a language, such differences should be related to phonology. This connection becomes effective and plausible because of the concepts of mora, syllable, and foot. Thus, the view has grown that the rhythmic character of a language is an emergent property, the product of phonological structure and post-lexical processes in speech production (Bolinger 1965; Dauer 1983; 1987; Low 1998; Ramus et al. 1999; Grabe & Low 2002; Russo & Barry 2003; Wagner 2007)². In the task of rhythm measurement, prosodic patterns serve as a guiding factor. This implies that maintaining some degree of isochrony is crucial for the effective functioning of the rhythmic predictor, as rhythmic measurements capture two levels of speech organization. These measurements encompass both segmental and prosodic syllabic structures embedded within phrasal prosody (Auer et al. 1999; Couper-Kuhlen 1993; Barry et al. 2009).

It is well-known that rhythm measures have been conceived to capture the rhythm typology of different languages. They aim to assign the rhythm of an utterance to either the syllable-timed pole, characterized by less variability in vocalic durations, or to the stress-timed pole, which features greater variability in vocalic durations along the rhythmic continuum. Structurally based measures, which

² Durational differences can be a consequence of phonological structures. This is true for vocalic length distinction or a complex syllable structure in German (frequent consonant clusters: e.g., *streckst*), durational stress properties in Wolof or Quechua, lexical schwa in French, or a relatively simple syllable structure, no vocalic length distinction, and no phonological vowel reduction in Italian, among many other phonological properties that differ across languages.

basically focus on different degrees of deviation from physical isochrony, appear to have been much more successful in differentiating languages rhythmicity (Ramus 1999; Ramus et al. 1999; Low et al. 2000; Grabe & Low 2002; Nolan & Asu 2009, Barry et al. 2003; Barry & Russo 2003; Kohler 2009; Fuchs in this volume): Ramus' delta values (ΔC and ΔV) (cf. Ramus 1999; Ramus et al. 1999), the standard deviation of the vocalic and consonantal intervals within an utterance, with in addition a measure of the vocalic proportion of the utterance (%V); the *Pairwise Variability Indices* (PVI) (Low et al. 2000)³. These measures capture separately the degree of variability in the vocalic intervals (vowel duration, PVI-V, ΔV , etc.) and the intervocalic (consonantal) intervals (PVI-C, ΔC , etc.). They represent a reflection of the structural properties of the syllables. This seems to contradict the common assumption of isochrony theory, according to which the syllabic unit is the important element at the basis of rhythmic impression, although, of course, there are measures also based on variability in syllable duration. Languages cannot be classified solely based on isochrony measures. However, they can be classified using both Ramus and Low structural measures (Ramus 1999; Grabe-Low 2002, etc.). These measures visualize the stress-timed – syllable-timed continuum, identifying vocalic and consonantal dimensions along which any language might deviate from the prototypical rhythm in a sort of rhythm space (see Barry & Andreeva 2010)⁴. These rhythmic measures capture various aspects of syllable complexity, making them suitable for comparing languages. These differences in complexity also impact the time required for articulating a syllable. These variability-based rhythm measures (PVI, ΔC , and ΔV) capture durational differences between consecutive vocalic and intervocalic intervals, which are correlated with differences in syllabic structure and the durational effects of degrees of prominence. However, rhythm measures, which focus on variability rather than isochrony (i.e., on the durational consequences of differences in syllable structure and phrasal modification), exhibit a less obvious connection between auditory impressions and physical measures. (Barry et al. 2003; Barry & Russo 2003; Dellwo 2006; Mok & Dellwo 2008; Asu & Nolan 2005). Arvaniti (2009) pointed out that there are no objective criteria for postulating a convincing degree of proximity or distance between measures to support a grouping or separation of languages. It's important to acknowledge the challenge of relating recent rhythmic measures to any auditory perception of rhythm. As a result, we must question whether languages can truly be reliably differentiated based on such measures (Arvaniti 2009; Barry et al. (2009); Barry & Russo 2010; Fuchs in this volume).

1.2. SPEECH RHYTHM AND THE SPEECH RHYTHM SPACE IN STUTTERING

Speech impairments can have an impact on rhythm. This is particularly evident in the case of stuttering, a motor control disorder that affects 1% of the global population (Yairi and Ambrose 2013). Stuttering speech is characterized by the presence of disfluencies, including repetitions of segments/sounds, syllables, words, prolongations of sounds, and interruptions (silent blocks), which can also manifest themselves as a glottal stop in the pre-phonatory posture (Monfrais-Pfauwadel 2014; Guitar 2019, Onslow 2020). Consequently, predictive rhythmic timing is malfunctioning in stuttering children, adolescents, and adults, since their ability in rhythmic speech production and timing is compromised. People who stutter know exactly what they want to say but are temporarily unable to articulate their speech due to muscle contractions. This sets them apart from non-stuttering individuals who also produce disfluencies, which are more reflective of lexical search or lexical planning time (Lickley 2018).

³ There exist derivative variants of these measures (see Barry et al. 2003; Russo & Barry 2008; Arvaniti 2009; Kohler 2009, Nolan & Asu 2009).

⁴ The term 'prototypical rhythm space' refers to the common categorization of languages into distinct rhythm classes, e.g. with English often considered the prototypical 'stress-timed' language and French as the prototypical 'syllable-timed' language.

Being hindered from producing their speech can lead to negative feelings in people who stutter, such as frustration, embarrassment, to the point where speakers may fear speaking up, avoid eye contact with their interlocutor, and/or isolate themselves.

There are two types of stuttering: developmental stuttering and acquired stuttering. Developmental stuttering typically begins between the ages of 2 and 7 and disappears in 80% of cases. Acquired stuttering is a generic term for all stutters that are not developmental. It can be caused by a stroke, tumour, head injury, side effects of certain medications, etc. (Guitar 2019).

In this chapter, we will focus solely on developmental stuttering. Its origin is multifactorial, involving neurological (Etchell et al. 2018) and genetic factors (Domingues & Drayna 2015; Riaz et al. 2005). Apart from these aspects, stuttering is also influenced by several linguistic and/or phonetic factors (Au-Yeung et al. 2003; Buhr & Zebrowski 2009; Howell et al. 1999).

This speech disorder primarily affects the first syllable at the beginning of a turn-taking (Monfrais-Pfauwadel 2014). Lexical words such as nouns and verbs tend to be more disfluent in adults who stutter compared to functional words. Similarly, stressed syllables are more difficult to pronounce for people who stutter compared to unstressed syllables.

Moreover, typical stuttering disfluencies have the peculiarity of being able to break within syllables instead of occurring between syllables or words. They are usually accompanied by tension that may be audible.

Stuttering has an impact on the timing and rhythmic flow of production since it affects timing mechanisms (Guitar 2013; Monfrais-Pfauwadel 2014; see DSM5 in Crocq et al. 2015; Didirková et al. 2021)⁵. We know, for example, that the oro-laryngeal timing of people who stutter has particular characteristics. In particular, VOT and VTT are longer in this category of speakers than in people who do not stutter (Agnello 1975). Other research, based on EMA data, has shown breaks in articulatory timing at the supraglottal level (Didirkova et al., 2021). Stuttering demands a temporal adaptation from speakers when synchronizing rhythmical movements to provide a structural grid of regularity and recurrence. Stuttering affects notably the background of regularity (i.e. the underlying rhythm of speech), the sequences of evenly spaced phonetic material, matched segments, and syllables. Thus, it can be defined as a neurodevelopmental disorder that disrupts the temporal organization of speech. Monfrais-Pfauwadel (2014: 2) speaks of audible and perceptible traces of motor and then psychic struggle in contrast with normal speakers. This is consistent with accounts of interruptions (*freezing*) in stuttering (see Alm 2021; Assaneo & Poeppel 2018; Orpella et al. 2024 on the interaction between auditory and speech-motor cortices, and the synchronization between auditory and speech-motor regions related to speech rates). Orpella et al. (2024) suggest that there is a reactive inhibitory control response from stutterers when they produce a word that will likely be stuttered. Technically, PWS show deactivation of left-hemisphere sensorimotor structures and over-activation of right-hemisphere parts. The problem is due to a lack of motor integration to regulate the movements of speech.

Studies on *Persons Who Stutter* (PWS) have evaluated the speech on rhythmic measures such as rate. They have already shown that people who stutter do not use a typical tempo in speech and do not have a rhythmic speech (Boecher et al. 2022). PWS have specific patterns in perceptually non-fluent speech, mainly characterized by a lack of coordination between supraglottal articulations and laryngeal gestures; they have a longer laryngeal movement reaction, compared with fluent speech produced by PWNS, Persons Who do Not Stutter (Zimmermann 1980; Van Lieshout et al. 1996; Heyde et al. 2016; Max & Gracco 2005; Didirková et al. 2021). This affects the temporal variability of oral articulations and the speech rate. A longer duration of onset movements than in PWNS, closing gestures, complex

⁵ DSM-5® = *Diagnostic and Statistical Manual of Mental Disorders*.

consonant clusters or vowel nuclei are encouraged by the steady position of the phonatory system, lips, or jaws. In PWS some articulatory movements show high velocity despite lower tempo, a negative correlation that could reflect defective speech-brain synchronization; in line with this proposal, the brains of PWS seem exhibit alterations, resulting in less stable speech motor planning and execution (Alario et al. 2006; Alm 2004, 2021). These factors lead PWS to asynchronous movements and a variable articulatory behaviour. One consequence of this are the stuttering-like disfluencies mentioned above and the difficulties for PWS to increase their speech rate (Howell et al. 1999). Thus, PWS show a poor temporal coordination, variable gestural movements, and a dysfunctional inter-articulatory coordination (Didirková et al. 2021).

The aim of this chapter is to study how rhythm is disrupted in stuttering speech by comparing adults who stutter (PWS) with typically developing adults (PWNS) (n = 14 per group). We assess simple and complex rhythmic chunks to achieve this. Speech rhythm has been quantified using rhythmic measures (the PVI from Grabe & Low 2002, ΔV and ΔC from Ramus 1999, Ramus et al. 1999). Both PVI and $\Delta V/\Delta C$ provide a diagnostic frame for identifying the two-dimensional presentation of the values (vocalic or consonantal) along which stuttered speech deviates from the prototypical normal speech.

1.3. METHODS

Among the types of stuttering described in the literature (Ward 2018; Yairi and Ambrose 2013), our study deals with persistent developmental stuttering, that generally starts between ages 3 to 7 and remains persistent from adolescence to adulthood (Didirková et al. 2021).

A corpus study was conducted, and the Raw and *Normalized Pairwise Variability Index* (nPVI) was computed for individual utterances to distinguish PWS from PWNS. We quantified rhythmicity in the speech of PWS using PWNS as control subjects.

1.4. CORPUS AND PARTICIPANTS

Our corpus investigation is based on audiovisual recordings coming from the French ANR project under grant no. ANR-18-CE36-0008 (BENEPHIDIRE: *Bégaiement: la Neurologie, la Phonétique, l'Informatique pour son Diagnostic et sa Rééducation*, PI: Fabrice Hirsch⁶). The main objective of this ANR project is to enhance our understanding of stuttering to facilitate diagnosis and treatment of this disorder by speech-language therapists. To collect the data, a multidisciplinary team, composed of researchers in linguistics, computer science, neurologists, along with therapists specialized in treating this disorder was assembled. This ANR project acquired morphological brain imaging data, articulatory data using dynamic MRI and acoustic data⁷. Our study is part of the work package aimed at studying the acoustic and motor characteristics of disfluencies.

We analysed data from 28 French native speakers (14 male and female *Who Stutter*; 14 male and female *Who do Not Stutter*⁸, PWNS), who participated in an interview task conducted by a speech

⁶ See <https://anr.fr/Projet-ANR-18-CE36-0008>. Place of acquisition: I2FH platform (CHU Gui de Chauliac, Montpellier) CHU Brabois, Nancy, France. ANR groups: 30 adults who stuttered vs. 30 control subjects-30 adults who stuttered at the start of rehabilitation vs. the same group 1 year later. In more detail, the ANR project involved processing brain morphological data and using the FSL software to extract and analyse various parameters, such as anisotropy fraction, mean, radial, and axial diffusivity. Additionally, it included measuring the displacements of each articulator, employing spatial landmark detection algorithms to identify facial contours, the mandible, and lips, and utilizing MRI and acoustic data. This research has been authorised by the French Data Protection Committee.

⁷ This project also aimed to investigate a neurological marker of stuttering, the Frontal Aslant (FAT). Its objective is to assess, whether the integrity and connectivity of this structure can serve as indicators of stuttering severity and its risk of chronicity.

⁸ In both cases 13 male speakers and 1 female speaker.

pathologist (a phoniatrist), accompanied by a speech therapist. The French treatment model is based on a relaxed style of interacting and motor rehabilitation in order (1) to reduce avoidance of speaking and (2) to develop gradually normal speech and to eliminate negative feelings. The therapy aims at restoring a flexible and spontaneous speech which allows patients to express themselves even when dysfluencies persist in conversational settings with smooth transitioning between listening and speaking.

The participants were asked by a phoniatrist to perform several tasks in the way of semi-directed speech and reading. The semi-spontaneous speech focuses on the description of a typical day, hobbies, Covid period and life, the emotional experience of the person with stuttering. Participants (PWS and PWNS) completed the interview task and a reading passage during the same interview under clinical test-taking conditions. Thus, PWS and PWNS participants were engaged in an in-person conversation. Control participants completed the same tasks and reading passage as PWS and matched the same questions.

We measured for each speaker 13 minutes of interview task, distributed over 9 minutes of semi-spontaneous speech and 4 minutes of the reading task.

The task also included a syllable-timed speech test as a training device aimed at enhancing speech in PWS. With this device, PWS were pushed to produce their speech with more isochronous intervals.

The stuttering of our PWS was evaluated as severe by their speech therapist on the *Riley's Stuttering Severity Instrument scale* (Riley 1994).

In the following sections we show the analysis conducted on 4 speakers (2 per group PWS and PWNS)⁹. We extracted 229 spontaneous speech samples from the recorded interviews. All samples were longer than 4 syllables (*ips* = inter-pause stretches > 4 syllables), matched for length and tasks in terms of PWNS. The *ips* are the utterances units used for calculating individual rhythm measures, which are then grouped and averaged over speaker. In addition to spontaneous speech, a total of 79 read speech samples were extracted.

2. ANALYSIS

The use of purely durational measures to capture the rhythmic effects might appear like an oversimplification, but in fact many structural properties of speech are linkable to duration (i.e. length properties). In PWS, there is a reduced amount of time available for articulatory gestures. The reduction in articulatory time leads to articulatory changes. These changes in articulation affect the quality of vowels and consonants. Additionally, alterations in speech timing occur as a result of these changes. This is why word length has also increased the rhythmic variability in PWS.

We used two approaches in our study: an automatic analysis and a manual analysis. Data analysis relied first on the auditory and acoustic identification of stuttered utterances. The manual method allowed us to better highlight the stuttered sequences.

Phrase boundaries were defined for utterances as syntactic boundaries or pauses (typical disfluencies) produced by the speaker. For our rhythmic computation, non-pathological disfluencies, revisions, or filler pauses, were removed from measures, so it was possible to identify prosodically uninterrupted 'inter-pause stretches' (*ips*), while stuttered (supra-glottal) disfluencies within phrase boundaries were considered as part of segmental stutterer's production¹⁰. Therefore, we assume that stuttering-like

⁹ Stutterers (PWS) are labelled in the ANR-18-CE36-0008 (BENEPHIDIRE) as Speaker B31 and Speaker 00001bis. Normal subjects (PWNS) are labelled in the ANR as C14, C15.

¹⁰ See Didirková et al. (2021); Peters et al. (2000), who investigated the underlying motor mechanisms of stuttering disfluencies.

disfluencies, segments/syllables repetitions, prolongations, stuttering blocks, count for rhythmic intervals¹¹.

This extraction process was performed using the speech editor Praat (software, v. 6.3.09)¹². We compared the duration of acoustic segments in the speech of PWS prior to and following the use of speech techniques, i.e. the automatic speech recognition and rhythm quantification based on the variability of vocalic and intervocalic intervals (see 3.).

The segmentation, aligned with the speech signal, was based on an acoustic-perceptive identification, as well as on automatic segmentation (see below).

This section also provides acoustical evidence for some basic differences in the syllable structure of PWS and PWNS which can be expected to result in systematically divergent rhythmic measures. We applied the following method.

The extracted speech samples were transcribed orthographically using an automatic speech recognition system called *Whisper*, an open-source project available at <https://github.com/openai/whisper>. The transcription was saved as a text file (.txt).

For the purpose of this work audio recordings were converted to .wav format and analysed in Praat. The entire corpus was first transcribed orthographically in automatic mode (see below) before being segmented and annotated semi-automatically using Praat.

After transcription, the speech samples were automatically segmented into phonetic and word segments using the WebMAUS Basic Service, a web-based tool provided by the Phonetics and Phonology Group at the Ludwig Maximilian University of Munich. Its interface can be accessed at <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>.

The transcription and segmentation results were then manually verified and corrected if necessary, ensuring the accuracy of the data. All segmentation boundaries were moved to their nearest zero-crossings using a Praat script called 'move-to-zero.praat' (See Figure 1: Tiers 1, 2, 3)¹³. A fourth tier was inserted to label all segmental events as a consonant or a vowel based on the phonetic transcription in the third tier (See Figure 1: Tier 4) using a Praat script called 'relabel_merge_interval.praat'¹⁴.

¹¹ Stuttering blocks are involuntary interruptions in the flow of speech, an interference with the brain's motor program, see section 2.

¹² Boersma, P. and Weenink, D. (2023) *PRAAT: Doing Phonetics by Computer* (Version 6.3.09) [Computer software]. <http://www.fon.hum.uva.nl/praat/>.

¹³ This script iterates through all TextGrid files and tiers within a directory, while also analysing zero-crossings in the corresponding .wav file (Channel 1). It preserves all original interval labels. Originally authored by Jessamyn Schertz (jessamyn.schertz@utoronto.ca, July 2020, <http://individual.utoronto.ca/jschertz/scripts.shtml>), with modifications made by Qianwen Guan in June 2023.

¹⁴ This script assigns replacement label 'y' to portions in the TextGrid labeled as 'x'. If you want to substitute an empty interval with text, use "" for the empty interval. Copyright Christian DiCanio, Haskins Laboratories, October 2011, with modifications made by Qianwen Guan in June 2023.

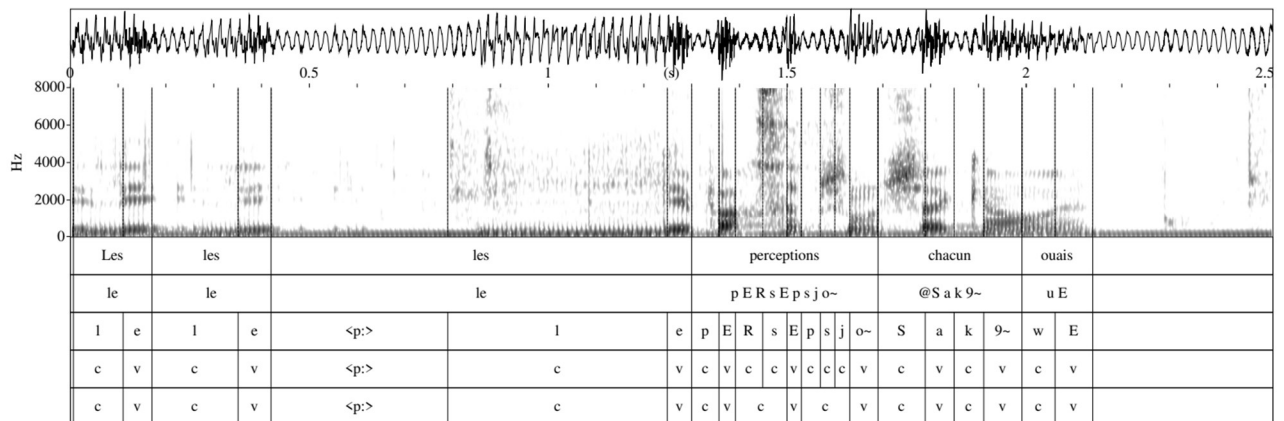


FIGURE 1. SPECTROGRAM: TIERS AND INTERVALS

Figure 1. Caption. Spectrogram illustrating characteristic disfluencies in the speech of a person who stutters, including prolonged intervals and atypical rhythm patterns. Visible disruptions in speech flow are evidenced by irregular spacing between phonetic elements, reflecting the temporal dynamics of stuttering. Vocalic and intervocalic intervals on tier 5.

Various X-SAMPA symbols were employed for annotation purposes in our study, such as in Figure 1 (see <p:> for stuttered pause)¹⁵. The tier 5 in Figure 1 indicates the interval, whether it is vocalic or intervocalic (= consonantal). For more details on the definition of vocalic and intervocalic intervals see Section 2.3.¹⁶

The specific annotation tier 5 was added to identify the stuttered PVI interval. This includes stuttered disfluencies, such as prolongations (consonantal and vocalic lengthening such as [kkkkkã] ‘when’, [ssssssu] ‘(I) am’, i.e., elongations of a sound) and repetitions (segments, syllables, words) annotated onto interval measures for PWS; blocks in utterance-internal phrase if any (silent intervals between two segments or syllables, combined with spasmodic tension and movements, see Didirková et al. 2021; Didirková & Hirsch 2020), see Figure 2 and 3:

¹⁵ Cf. <https://www.phon.ucl.ac.uk/home/sampa/french.htm> . French vowels in X-SAMPA: nasal vowels: e~, a~, o~, 9~, oral vowels: i, e, E, a, A, O, o, u, y, 2, 9, @; French consonants in X-SAMPA: plosives: p, b, t, k, g, fricatives: f, v, s, z, S, Z, j, nasals: m, n, J, N, liquids: l, R, w, H, j.

¹⁶ The intervocalic (i.e. consonantal interval) can include in duration c + a stuttering <p:> as in Figure 1, tier 5.

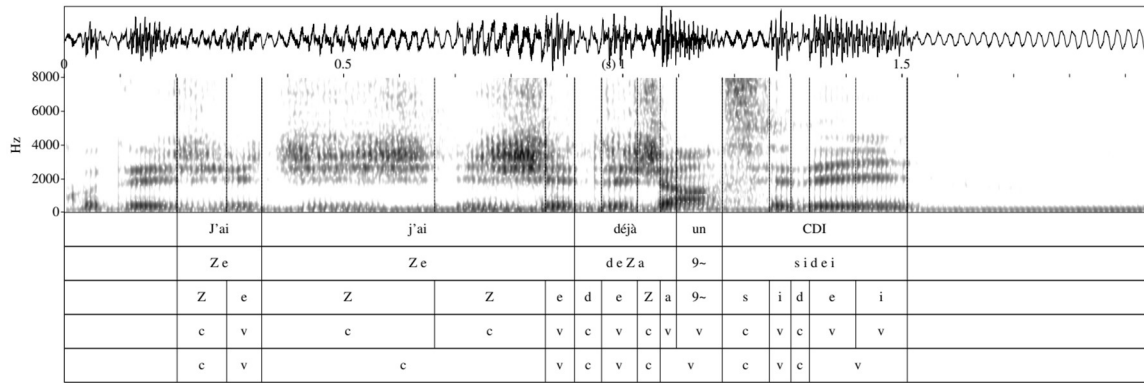


FIGURE 2. SPECTROGRAM: STUTTERING ON CONSONANTS (SPEAKER B31)

Figure 2. Caption. Spectrogram of stuttered speech. This visual representation captures the prolonged and repeated articulations typically seen in PWS, such as extended consonant intervals and the irregular vocalic segments, reflecting the disrupted timing and rhythm patterns that challenge the regular speech flow.

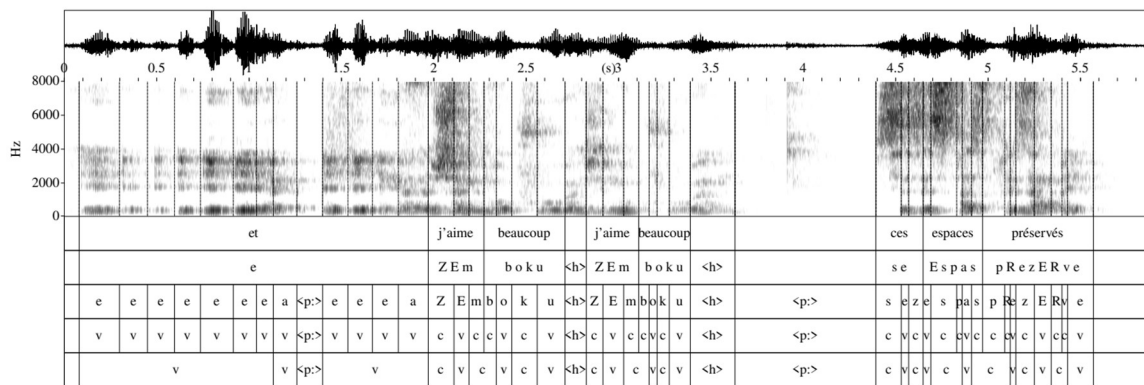


FIGURE 3. SPECTROGRAM: STUTTERING ON VOWELS AT THE BEGINNING OF THE UTTERANCE (SPEAKER 00001BIS)

Figure 2. Caption. Spectrogram highlighting the pattern of stuttering on vowels at the beginning of an utterance in a person who stutters. The image showcases the characteristic stuttering disfluencies including a high number of repetitions, prolongations, and blocks, illustrating the unique temporal dynamics and rhythm disruptions encountered in stuttered speech.

A high number of repetitions, prolongations, and blocks were identified in our analysis. By 'stuttered' block we mean a disfluency made up of a silent duration between two segments or syllables associated with tension, whereas the stuttered prolongations are identified as an identical structure of the formants (for vowels and sonorants) or spectral cues typical of a consonantal sound. Stuttered disfluencies are only produced by PWS.

2.1. RHYTHM MEASUREMENTS AND THE PAIRWISE VARIABILITY INDEX: A QUANTITATIVE ANALYSIS

We highlight that the rhythmic measures introduced in section 1, such as PVI-V and ΔV , reflect differences between languages with only single-slot syllabic nuclei and those with single and double-

slot nuclei. Additionally, PVI-C and ΔC measures are sensitive to differences in the onset and coda structure (see Russo 2010). Thus, a language with more variable onset- and coda-structure, long and short vowels, reduction of unstressed syllables will generate higher variability measures than a language without such features.

Furthermore, it's important to note that the same rhythmic measures, calculated for the same language, but from two different corpora, can result in radically different typological associations in terms of rhythmicity.

We calculated the variability of vowel and consonantal duration and computed rhythmicity of the PWS utterances, adopting the PVI methodology first proposed by Grabe and Low (2002) to measure rhythmic duration.

The basic hypothesis to be tested is that the range of vowel variability, consonantal duration, possible syllable complexity, and other phonological differences between PWS and PWNS lead to an important difference in the rhythm measures between groups. Quantified rhythmicity thus depends on the intersection of multiple parameters, and it is defined in terms of degree rather than rhythmic dichotomy (see Barry & Andreeva 2010).

The rhythmic nature of the speech alterations in stuttering also leads us to some understanding of the cognitive phonological processes behind the behavioural PWS data. We conducted measurements on vocalic intervals and the intervals between vowels (excluding 'normal' disfluencies, such as pauses and hesitations) within a speech passage. We calculated the PVI, the mean difference in vocalic and intervocalic interval from one vowel or one consonantal interval to another (raw PVI and normalized PVI).

This index of variation quantifies the extent of variability observed in consecutive measurements. The equation (1) provides the raw *Pairwise Variability Index* (rPVI).

(1) rPVI

$$r PVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1) \right]$$

Where m is the number of intervals, vocalic or intervocalic, and d is the duration of the k th interval. In (1) rPVI (PVI-C) is not normalized for speech rate. However, a normalised PVI, which relates the difference between intervals to the mean duration of the two intervals, was introduced by Deterding (2001) (cf. Low et al. 2000) as an explicit correction for tempo change¹⁷. Thus, speech rhythm was quantified in our study between successive vowels also using the *Normalized Pairwise Variability Index* (nPVI).

This version, known as nPVI, is represented by the following equation in (2) (see Deterding 2001; Barry & Russo 2003; Russo & Barry 2008a/b, 2010 among others).

(2) nPVI

$$n PVI = 100 \times \left[\sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{(d_k + d_{k+1}) / 2} / (m-1) \right]$$

¹⁷ In their investigation on vowel durations, Grabe & Low (2002) used the normalized variant of the *Pairwise Variability Index* (nPVI).

In (2), the duration (d) of a vowel ($k + 1$) is subtracted from the duration of the preceding vowel (k) and divided by their average duration. The absolute values of the resulting sub-totals are summed up and divided by the number of vowels in the phrase (m) minus one, the result is multiplied by 100 to obtain a normalized score.

The normalisation method which is used for PVI-V (but not for PVI-C) shows that the nPVI-V reduces local inter-syllabic differences, such as stressed vs. unstressed or short vs. long vowel, which are essential cues of rhythmic impressions. Thus, the range of vowel variability is generally reduced in comparison to Ramus measures (ΔV and ΔC , see section 3) but, as it will be shown below, its sensitivity to tempo effects remains.

PVI-V captures the degree to which consecutive vowel durations vary: a) long vs. short vowels, b) phonetic variation due to differences in degree of aperture; the effects of phrasal accentuation. The PVI-C captures, on the other hand, the degree to which consecutive consonantal durations vary (e.g. single consonants or consonant clusters). PVI-V and PVI-C provide a measure of variation which takes the sequential nature of rhythmic impressions into consideration.

The PVI-C measure captures the degree to which consecutive consonantal durations vary (i.e., single consonants or clusters)

We performed all the calculations of rPVI and nPVI for vocalic and intervocalic intervals using R (Core Team 2021). Vocalic intervals were identified as the portion of the signal between the onset and offset of a vowel, characterized by vowel formants. This definition encompassed sections with varying numbers of vowels, including monophthongs or multiple vowels spanning across the transition between adjacent words.

Intervocalic intervals, on the other hand, were defined as the segment of the signal between the offset of one vowel and the onset of the subsequent vowel, regardless of the number of intervening consonants. To measure the duration of both vocalic and intervocalic intervals, we employed a left-to-right approach using wide-band spectrograms in Praat. Our first query was whether r/nPVI were different between PWS and PWNS. The higher values of rPVI-C (as for ΔC) indicate that PWS speech is sensitive to complex consonantal structures in the onset and coda of a syllabic structure, as commonly found in languages with a predominant C(C)VC(C) structure.

To offer a broader perspective and some possibility for comparison, we calculated (in section 3.) both the sequentially calculated pairwise variability measures (PVI) using the pairwise normalisation procedure for vowels (Grabe & Low 2002), the three global measures used by Ramus et al. (1999) applied to both stuttering and normal speech and the speech rate (the number of vowel intervals per second, including pauses):

(3) Ramus measures

%V (within ips = interpause stretches)

ΔV – standard deviation of vocalic intervals

ΔC – standard deviation of intervocalic intervals

The variability of vocalic and intervocalic (consonantal) intervals are taken by both Ramus and by Grabe and Low as correlates of the complex interaction of structural properties. However, their way of calculating the variability is different. The Ramus %V measure does not capture variation, and ΔV and ΔC are measures of overall vocalic and consonantal variation rather than an accumulative pairwise measure. The vocalic proportion of the utterance (%V), as a measure, is difficult to interpret in connection with any concept of rhythm.

3. RESULTS

In this section, we present the results of stuttered speech from two speakers, comparing them with two non-stuttering speakers. The sample size is insufficient to conduct inferential statistical analyses. Therefore, we provide only descriptive statistics, including percentages and standard deviations, following the approach outlined by Ramus et al. (1999). We also report observations of rPVI and nPVI across different speech conditions (stuttered vs. fluent) and speech styles (read vs. spontaneous). Moreover, we examine differences among individual speakers. We computed the average duration of speeches, including both fluent and stuttered speech, categorized according to various contextual factors. Error bars on the graph depict the standard error of the mean (SEM).

In Table 1 and Figure 4 we display the outcomes for our four speakers (2 per group, PWS and PWNS). These results include the standard deviation of durations for vowel and intervocalic consonantal intervals (ΔV and ΔC , respectively, shown in Figure 4), alongside the percentage of vowel intervals (%V) and consonantal intervals (%C) within *ips*. Additionally, we detail the proportion and standard deviation of computed pauses and hesitations.

TABLE 1. % OF VOCALIC INTERVALS, INTERVOCALIC INTERVALS, PAUSES AND HESITATIONS IN STUTTERING AND NORMAL SPEECH

	<i>Stuttering</i>	<i>Normal Speech</i>
<i>Intervocalic</i>	46.4%	48%
<i>Vocalic</i>	48.3%	49.9%
<i>Pauses</i>	2.5%	1%
<i>Hesitations</i>	2.7%	0.45%

Table 1. Caption. *Comparative Analysis of Vocalic and Intervocalic Intervals, Pauses, and Hesitations in Stuttered versus Normal Speech: This table presents the percentages of vocalic intervals (%V) and intervocalic intervals (%C), as well as the proportions of pauses and hesitations, for both PWS and PWNS.*

The calculations in Table 1 show that stuttered speech exhibits a higher percentage of pauses and hesitations compared to normal speech. Notably, the percentage of vocalic intervals (%V) remains consistently reliable across our speakers. Measures such as %C and %V do not effectively differentiate between PWS and PWNS. In contrast, stuttering disfluencies prove to be more distinguishing factors between the two groups. In Figure 4, the average and standard deviation of vocalic intervals, intervocalic intervals, duration of pauses and hesitations show differences for ΔV and ΔC . The differences observed in the average and standard deviation (ΔV and ΔC), between PWS and PWNS can be attributed to the distinct speech patterns associated with stuttering: disfluencies (such as repetitions, prolongations, and blocks) can lead to differences in the timing and duration of vocalic intervals and consonantal intervals between the two groups. PWS exhibit more variability (higher standard deviation) in these intervals due to the interruptions caused by disfluencies. Stuttering involves disruptions in the neural processes responsible for speech production and motor control. These disruptions can lead to inconsistencies and variability in the timing and duration of speech intervals, as well as in the deviation of these intervals from the expected norms.

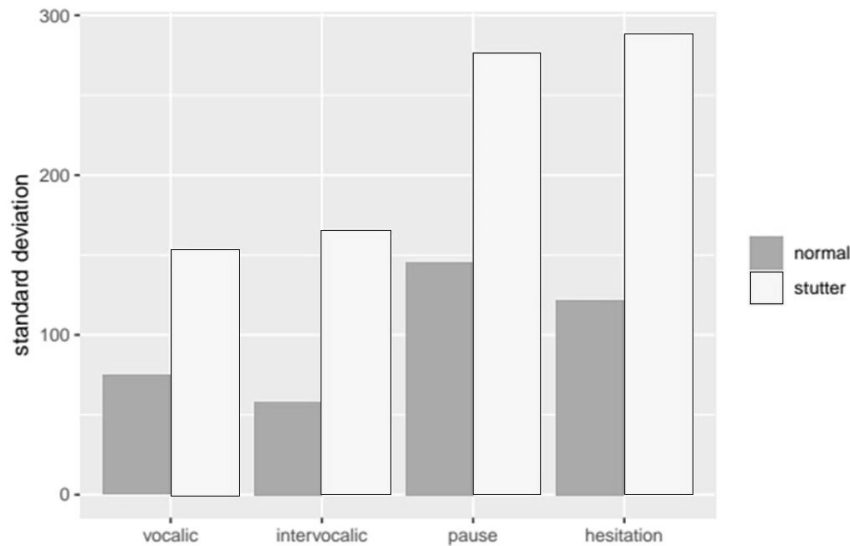


FIGURE 4. AVERAGE STANDARD DEVIATION OF VOCALIC INTERVALS, INTERVOCALIC INTERVALS, DURATION OF PAUSES AND DURATION OF HESITATIONS

Figure 4. Caption. *Variability in Speech Patterns: comparing Standard Deviation of Speech Components between PWS and PWNS: This bar graph quantifies the standard deviation for vocalic intervals, intervocalic intervals, pauses, and hesitations. The data show PWS experience more variability in the duration of pauses and hesitations, indicative of stuttering disfluencies. PWNS show lower standard deviation in both vocalic and intervocalic intervals, indicating more consistent timing. This increased variability in intervocalic intervals among PWS suggests a disruption in speech rhythm typical of stuttering patterns.*

In Figure 4, we see that PWNS have a lower standard deviation in vocalic intervals than PWS. This suggests that PWNS have less variability in the duration of their vocalic intervals. The standard deviation for intervocalic intervals is lower for PWNS compared to those who stutter, which again indicates less variability for the non-stuttering group in the timing between vowels across consonants. For the duration of pauses, the standard deviation is higher for PWS. This implies that the length of pauses among PWS varies more than PWNS. Furthermore, Figure 4 shows a higher standard deviation for PWS in the duration of hesitations, suggesting greater variability compared to PWNS. Thus, it appears that PWS have more variability in the duration of their pauses and hesitations but less variability in vocalic and intervocalic intervals compared to PWNS. This could reflect a compensation mechanism where PWS try to maintain a steadier rhythm in some speech components while experiencing more variability in others, as a response to the disruptions caused by stuttering.

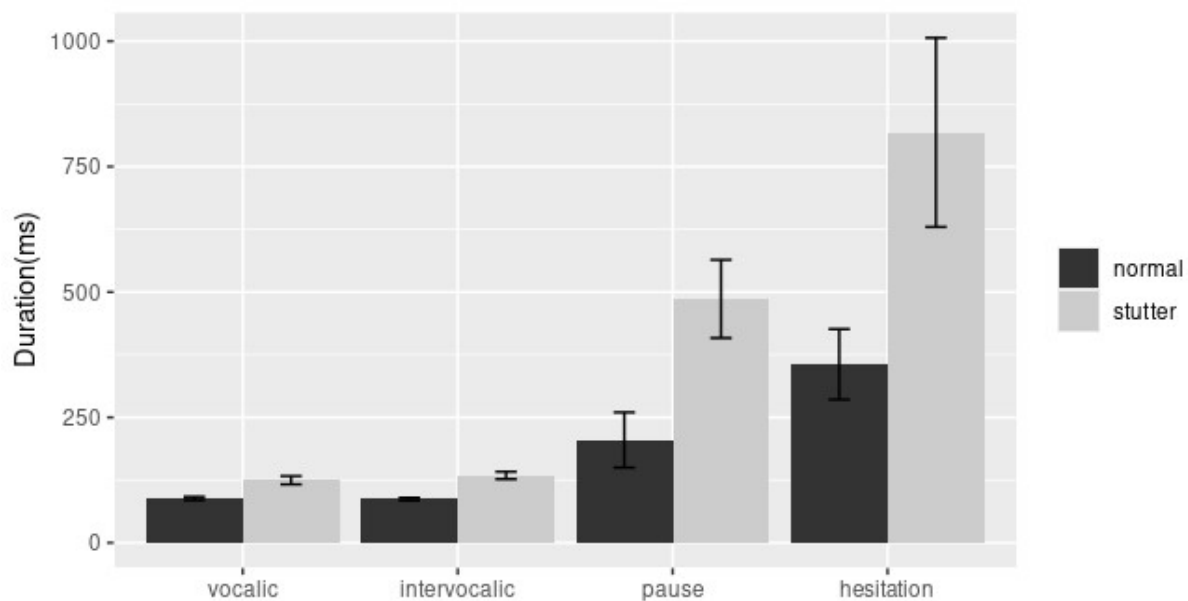


FIGURE 5. AVERAGE DURATION OF SPEECHES ACCORDING TO VOCALIC, INTERVOCALIC, PAUSE, AND HESITATION

Figure 5. Caption. *Comparative Duration of Speech Elements in Stuttering and Normal Speech: This bar chart presents the mean durations of vocalic and intervocalic intervals, pauses, and hesitations for both normal speakers and PWS. PWS demonstrate longer and more variable durations for pauses and hesitations, highlighting the temporal disruption characteristic of stuttered speech.*

The Figure 5 presents the average duration of various speech elements for PWS and PWNS, categorized according to different contextual factors such as vocalic, intervocalic interval, pause, and hesitation, with error bars indicating the standard error of the mean. We observe the following.

For the vocalic intervals both PWS and PWNS show relatively short average durations, with PWS having slightly longer vocalic intervals on average, as indicated by the error bars; for the intervocalic intervals: again PWS show longer average duration compared to PWNS. The average duration of pauses is notably longer for PWS. This difference is clearer, as shown by the lack of overlap in the error bars.

There is a significant difference in the average duration of hesitations, with PWS showing much longer durations. This is due to the individual's attempt to avoid or postpone stuttering events, which results in prolonged filler sounds or silent blocks.

The error bars indicate the variability of each measurement, and we can see that there is greater variability in the duration of hesitations for both groups, but it is most pronounced in PWS. This aligns with the variable nature of stuttering and how it can impact different aspects of speech timing. Overall, PWS tend to have longer durations of pauses and hesitations, which can be attributed to the speech disfluencies characteristic of stuttering.

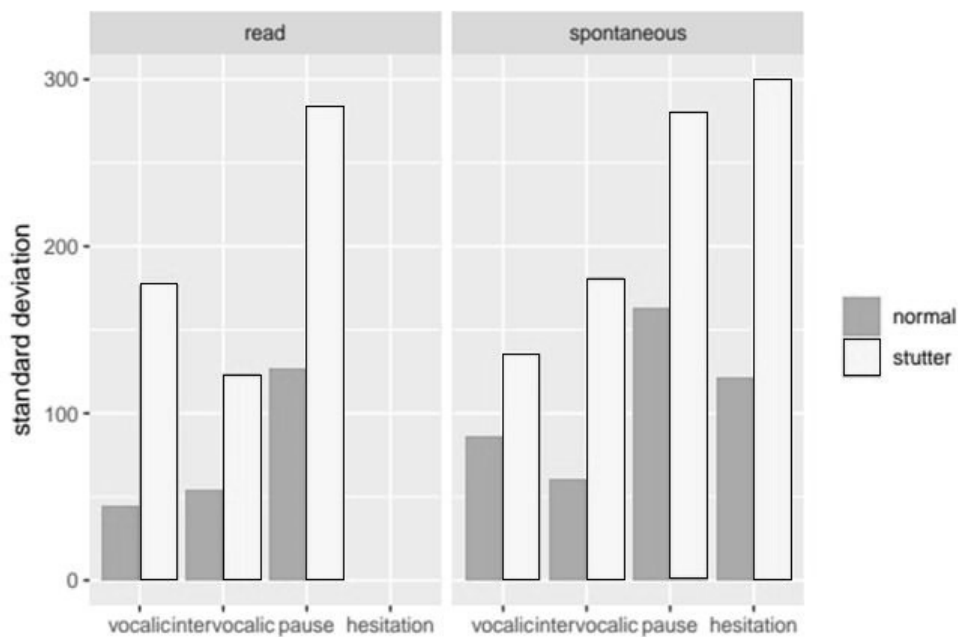


FIGURE 6. EFFECT OF SPEECH STYLE: READ AND SPONTANEOUS.

Figure 6. Caption. *Variability in Speech Components During Read and Spontaneous Speech: The bar graph compares the standard deviation of vocalic and intervocalic intervals, pauses, and hesitations for both PWS and PWNS across read and spontaneous speaking tasks. Notably, PWS exhibit a higher standard deviation in intervocalic intervals than PWNS in both speech contexts, which is indicative of greater timing irregularities during stuttered speech. This variability is more pronounced during spontaneous speech, suggesting that unplanned speaking poses additional challenges for PWS.*

In Figure 6, the standard deviation of vocalic and intervocalic intervals, as well as the duration of pauses and hesitations, are compared across read and spontaneous speech modalities for PWS and PWNS.

This Figure presents a clear distinction between read and spontaneous speech, showing that the rhythmic measures are indeed text-dependent.

In the context of read speech, PWS show higher standard deviation compared to PWNS, suggesting more variability in the duration when reading.

For intervocalic intervals, again, PWS display higher standard deviation during read speech, indicating more inconsistency in the timing between vowels across consonants.

For PWS, the standard deviation for pauses is higher compared to PWNS, indicating that even in a controlled reading environment, PWS exhibit more variability in their pausing. There is a significant increase in the standard deviation of pauses for PWS during reading tasks. For hesitations, PWS have a notably higher standard deviation, reflecting much more variability in the occurrence and duration of hesitations when compared to PWNS in read speech. In spontaneous speech, the variability in the duration of vocalic intervals for PWS increases further as evidenced by the standard deviation, surpassing that of PWNS. Similarly, the standard deviation for PWS in intervocalic intervals during spontaneous speech is greater than for PWNS, indicating a heightened level of variability. The variability in pauses for PWS in spontaneous speech also increases, maintaining a higher standard deviation compared to PWNS. The standard deviation of hesitations in spontaneous speech for PWS is significantly higher than for PWNS. We use here 'significantly higher' in a descriptive sense, observing that the Figure 6 indicates a noticeably larger standard deviation for PWS, which suggests more variability in their hesitations during spontaneous speech compared to PWNS.

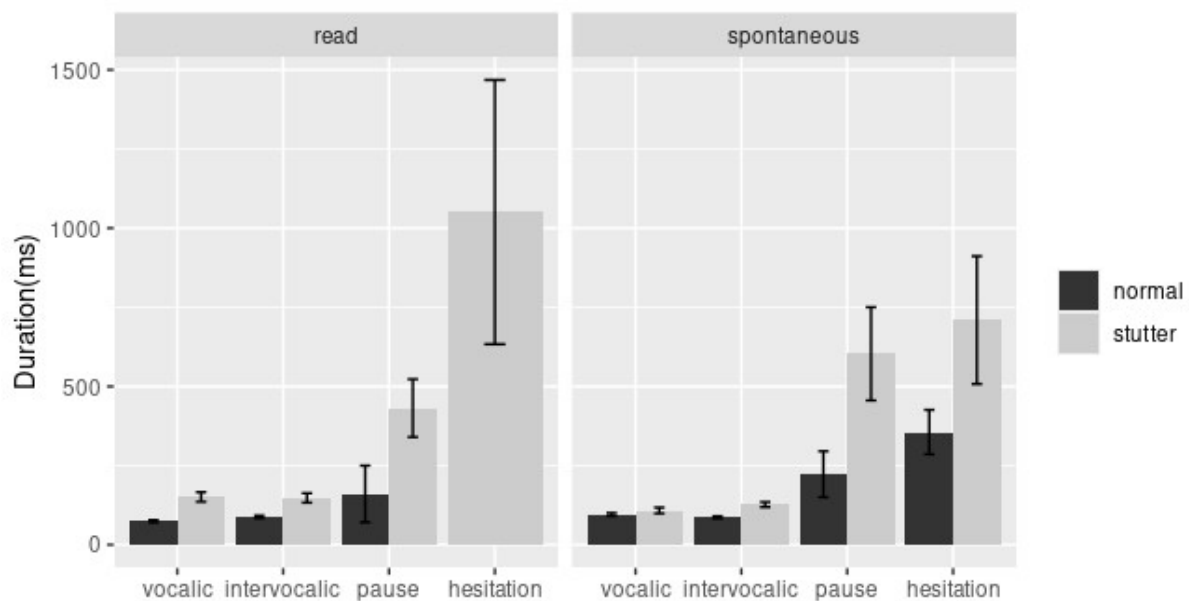


FIGURE 7. MEAN DURATION OF NORMAL AND STUTTERED SPEECH ACROSS READ AND SPONTANEOUS TASKS

Figure 7. Caption. *Duration of Speech Elements in Read and Spontaneous Speech: The bar graph illustrates the average duration of vocalic and intervocalic intervals, pauses, and hesitations for PWS compared to PWNS during read and spontaneous speech. The graph indicates that PWS experience longer and more variable durations of these speech components, especially during spontaneous speech, highlighting the increased challenges faced by PWS in real-time conversational contexts.*

Figure 7. illustrates the mean duration of various speeches (normal and stuttered) within the four contextual categories across distinct speech styles (read and spontaneous), with error bars representing the standard error of the mean.

For read speech, both PWS and PWNS have similar, relatively short mean durations for vocalic intervals, but PWS have slightly higher variability as indicated by longer error bars. For intervocalic intervals, PWS show longer mean durations with greater variability than PWNS. This may indicate more difficulty in transitioning between sounds. The mean duration of pauses for PWS is much longer than for PWNS, with significantly larger error bars. PWS demonstrate a longer mean duration for hesitations compared to PWNS, along with greater variability, as evidenced by the longer error bars. In spontaneous speech, the mean durations for vocalic intervals are relatively similar between PWS and PWNS; however, PWS exhibit more variability. Again, PWS have longer mean durations for intervocalic intervals than PWNS, with larger error bars indicating more variability. PWS have longer pauses on average than PWNS in spontaneous speech as well, and the variability is quite high. PWS have a notably longer mean duration of hesitations and also show considerable variability in these durations. The error bars suggest that the variability in speech patterns is contextually influenced. The larger error bars for PWS across all categories indicate that stuttered speech is more variable and less predictable than non-stuttered speech, particularly in spontaneous contexts. In Figure 7, the intervocalic intervals for PWS appear to have longer mean durations than those for PWNS in both read and spontaneous speech, which could suggest that these intervals are particularly affected by stuttering. The error bars for the PWS group are noticeably longer in the intervocalic category compared to the PWNS group, indicating greater variability around the mean. This greater variability for PWS suggests that intervocalic intervals might be a key area where stuttering manifests, potentially due to the motor

coordination needed to transition between sounds or the timing required to articulate consonants that come between vowels.

The results also show that persons who stutter severely speak with a lower speech rate than speakers who stutter less severely or who speak in a 'normal' way, i.e., normal subjects¹⁸, see Figures 8 and 9:

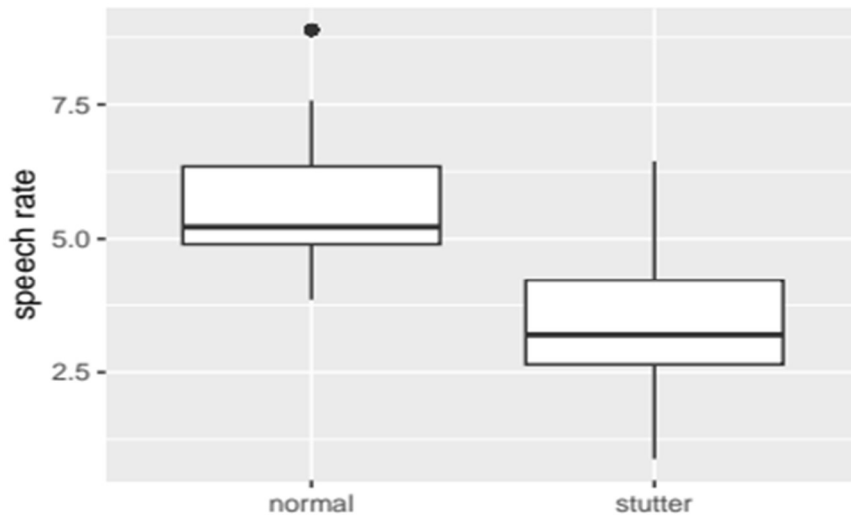


FIGURE 8. COMPARISON OF SPEECH RATES BETWEEN PWNS AND PWS

Figure 8. Caption. *Speech Rate Distribution for PWS vs. PWNS: This box plot reveals that PWS generally have a slower speech rate than PWNS, as evidenced by the lower median value. The plot also shows a wider range of speech rates among PWS, indicating greater variability within this group.*

We quantified speech rate by dividing the number of vocalic intervals by the total time in seconds of each utterance. The median value, indicated by the line in the middle of each box, is lower for PWS compared to normal speakers. This suggests that, on average, PWS have a slower speech rate.

The boxes represent the Interquartile Range (IQR), which is the middle 50% of the data. The IQR for PWS is narrower and shifted lower on the scale, indicating that most PWS have a lower speech rate compared to most normal speakers. The dot above the 'normal' category indicates an outlier, a speech rate that is unusually high compared to the rest of the data for normal speakers.

The figure underscores the impact of stuttering on speech rate, showing that stuttering tends to slow down speech and reduce variability in speech rate among PWS. This information is crucial for understanding and developing therapeutic strategies to help PWS manage their speech rate and improve communication effectiveness.

¹⁸ We calculated the speech rate on speaker B27, who stutters severely, and on B29, who relaxes the vocal cords using buzzing to avoid stuttering.

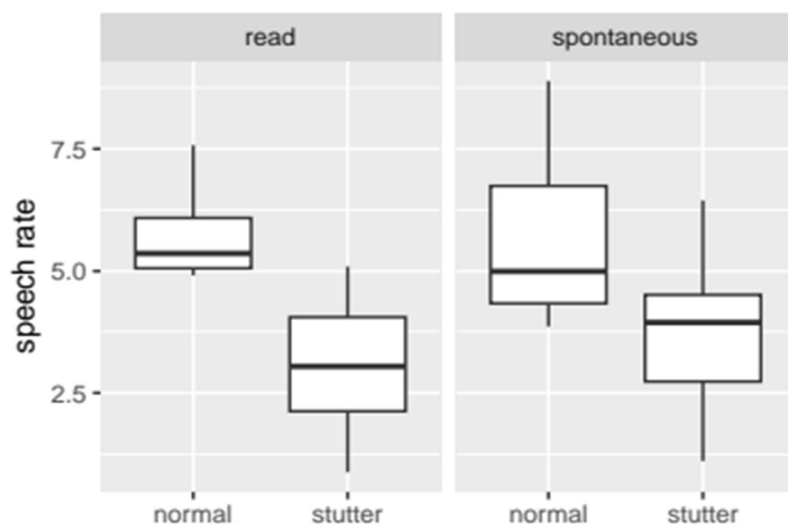


FIGURE 9. SPEECH RATE: EFFECT OF SPEECH STYLE.

Figure 9. Caption. *Speech Rate in Read and Spontaneous Speech for PWS and PWNS: The box plots compare speech rates, showing that PWS have a consistently slower speech rate than normal speakers in both read and spontaneous speech modes. The wider spread of rates for PWS during spontaneous speech suggests greater variability in speech production when speaking without a script.*

In Figure 9 we compare the speech rate of PWNS and PWS during read and spontaneous speech tasks. In read speech, the median speech rate for PWNS is higher than for PWS, indicated by the median line within each box. This suggests that during read tasks, PWS generally speak more slowly. The median speech rate for PWS in spontaneous speech appears to be lower than for PWNS, consistent with the pattern seen in read speech. The IQR for PWS in spontaneous speech is narrower than for PWNS, and the median is lower, reinforcing that PWS have a more confined range of speech rates and generally slower speech. PWS show longer whiskers in spontaneous speech compared to read speech, indicating a broader range of speech rates in spontaneous conditions. However, the median is still lower than that of PWNS, underlining a slower speech rate for PWS.

The sequential Grabe and Low variation measures also showed notable effects for intervocalic (consonantal) variation.

In Figure 10, we present the PVI results for all utterances, including both raw for consonantal intervals and normalized for vocalic intervals, along with a comparison of rhythm patterns between stuttering and normal speech in both read and spontaneous conditions. Each point in the figure represents the results of individual utterances, with vocalic nPVI values plotted on the vertical axis and intervocalic rPVI values on the horizontal axis:

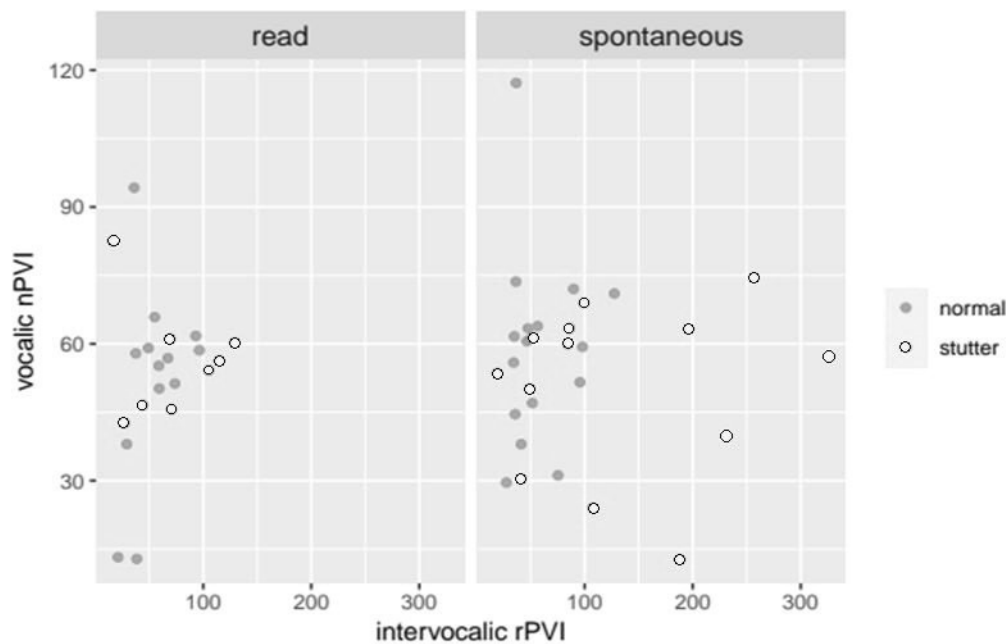


FIGURE 10. PVI (RAW AND NORMALIZED) RESULTS FOR ALL UTTERANCES (PWS AND PWNS).

Figure 10. Caption. *Pairwise Variability Index (PVI) in Read and Spontaneous Speech for PWNS and PWS: This scatter plot displays individual utterances, comparing vocalic nPVI and intervocalic rPVI. The data points illustrate that PWS, particularly in spontaneous speech, tend to have higher intervocalic rPVI values, signifying greater variability in the timing of their speech and reflecting the rhythmic irregularities associated with stuttering.*

In read speech, PWS tend to have greater variability in both vocalic and intervocalic intervals. In the spontaneous section, while there is some overlap between PWS and PWNS, PWS still tend to show higher vocalic nPVI and intervocalic rPVI values. This suggests that in spontaneous speech, the pattern continues, with PWS displaying greater rhythm variability.

In spontaneous speech, points representing PWS are further right on the horizontal axis than those for PWNS, this indicates higher rPVI values for PWS, suggesting greater variability in their intervocalic intervals.

The result from Figure 10 regarding rPVI-C is consistent with the acoustic analysis we conducted, and it appears to be correlated with frequent repetitions and prolongations of consonants, especially in onset positions of syllables. Additionally, it is associated with tense pauses (blocks), during which a word sequence fails to initiate.

On the vertical axis, higher points indicate greater variability in vocalic intervals. PWS who have higher nPVI values, show more variability in the rhythm of their vocalic intervals. The high nPVI-V values for PWS also reflect prolonged vocalization associated with stuttering, as well as prolonged pauses between vocalization chunks. The higher rPVI-C for PWS is consistent with the presence of speech disfluencies showed above. These disruptions in the speech flow contribute to the irregularity of consonantal intervals, which is captured by a higher rPVI measure.

The figure provides a visual confirmation of these descriptions, showing that stuttered speech can be differentiated from 'normal' speech in terms of temporal variability, with PWS generally showing increased variability in rhythm.

We subsequently calculated intra-speaker across-utterance variability. This variability was more pronounced in PWS compared to PWNS. Individuals who stutter exhibited a wide range of

rhythmic measurements, both high and low. This is also because rhythm values are a reflection of the language material that occurs in the corpus, and of the style in which the utterances are produced (see Barry & Russo 2003; Russo & Barry 2010). Two speakers reading the same two texts can vary significantly, even in different dimensions for different texts, one on the consonant axis, the other on the vowel axis. With spontaneous speech, and speaker-variation it can be expected to be a much higher variability.

In Figure 11, we give the average of the PVI results. PWS has high variability in both vocalic and intervocalic intervals, as measured by the nPVI and rPVI respectively. The Figure provides a visual comparison between PWS and PWNS across different speaking conditions, allowing for an analysis of how stuttering may affect speech rhythm in controlled (read) versus natural (spontaneous) settings. Vocalic nPVI values are plotted on the vertical axis against intervocalic rPVI values on the horizontal axis:

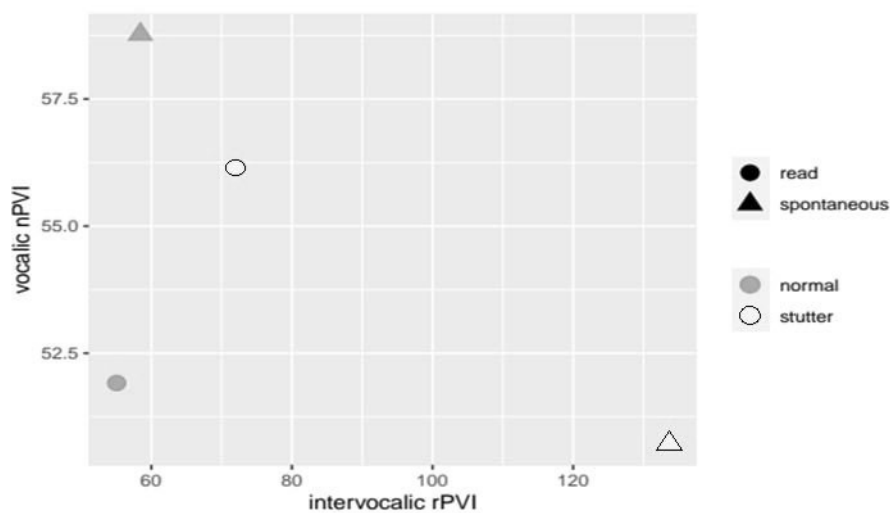


FIGURE 11. PVI RESULTS FOR PWS AND PWNS: AVERAGE (READ AND SPONTANEOUS SPEECH).

Figure 11. Caption. *Rhythmic Variability in Speech: The plot illustrates the relationship between vocalic nPVI and intervocalic rPVI for both PWNS and PWS in read and spontaneous speech contexts. Each symbol represents a distinct speech sample, with higher intervocalic rPVI values observed among PWS, especially in spontaneous speech. This indicates a more variable and disrupted speech rhythm in PWS, which contrasts with the more uniform rhythmic pattern seen in normal speakers.*

PWS and PWNS differ in terms of average PVI in reading and in interview condition. Read speech has allowed us to identify text-dependent and speaker-dependent differences in rhythm measures. The reading condition revealed less rhythm differences between PWS and PWNS than spontaneous speech.

The spread of the point along the horizontal axis suggests that there is some variability in the timing between consonants (intervocalic rPVI) for stuttered speech in read and spontaneous conditions. In this Figure, the highest intervocalic rPVI value for PWS is indicated by a triangle in the spontaneous speech section, positioned further right on the horizontal axis. This high rPVI value suggests that there is considerable variability in the timing between the consonants for this PWS speech sample during spontaneous speech.

PWS produced a range of rPVI-C values. As in Figure 10, in our PWS corpus we found a production variability between consonantal and vocalic segments greater in consonantal targets and complex

sequences than in vowels. This leads to more rhythmic speech in those intervals (i.e., relatively low nPVI scores for PWS, and high rPVI).

3.1. SPEECH RHYTHMICITY OF INDIVIDUAL SPEAKERS

We found systematic differences between two speakers who stutter, thus clearly, personal speech-production strategies of PWS affect ‘rhythm’ measures.

The range of PVI values produced by the same speaker was greater in the PWS group compared to the PWNS group, whereas PWNS produced a more regular nPVI level.

We show in Figure 12 the average duration of speech components at the individual speaker level, incorporating both normal and stuttered speech, categorized according to different contextual factors such as vocalic, intervocalic, pause, and hesitation, with error bars indicating the standard error of the mean:

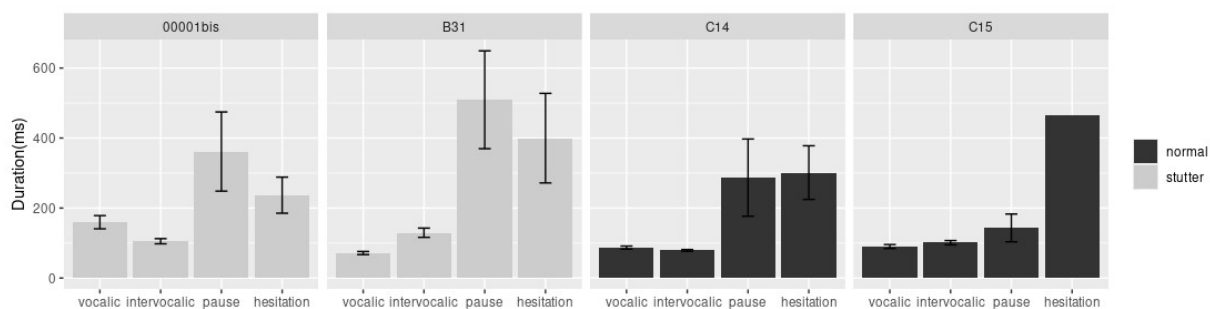


FIGURE 12 AVERAGE DURATION METRICS FOR VOCALIC AND INTERVOCALIC INTERVALS, PAUSES, AND HESITATIONS AMONG INDIVIDUAL SPEAKERS

Figure 12. Caption. *Duration of Speech Components Among Individual Speakers: The bar graph compares the average duration of vocalic and intervocalic intervals, pauses, and hesitations for individual speakers identified as 00001bis, B31, C14, and C15. It highlights the contrast in speech durations between normal speakers and people who stutter (PWS), with PWS generally showing longer durations in pauses and hesitations, indicative of the speech disfluencies commonly associated with stuttering. The differences in duration and variability between speakers underscore the individual nature of stuttering manifestations.*

The Figure 12 suggests that for some speakers, the duration of vocalic intervals is longer in stuttered speech compared to normal speech, although this is not consistent across all speakers. However, the Figure shows that, the intervocalic intervals (rPVI) can be longer for stuttered speech than for normal speech. The duration of pauses appears to be longer in stuttered speech for all individual speakers represented in the figure, suggesting a common trait among people who stutter. The average duration for hesitations, is longer for stuttered speech, while in others, it is comparable between stuttered and normal speech. The error bars across all categories indicate variability in the durations for both normal and stuttered speech. A larger error bar indicates for PWS more variability in that particular speech component. Overall, the Figure reveals that the average duration of both intervocalic intervals and pauses is longer in stuttered speech compared to normal speech, with individual variations across the speakers. Hesitations and vocalic intervals show a less consistent pattern and may vary more on an individual basis.

The Figure 13 illustrates the mean duration of various speeches (normal and stuttered) within the four contextual categories across distinct speech styles (read and spontaneous), with error bars representing the standard error of the mean, which reflects the amount for each speaker:

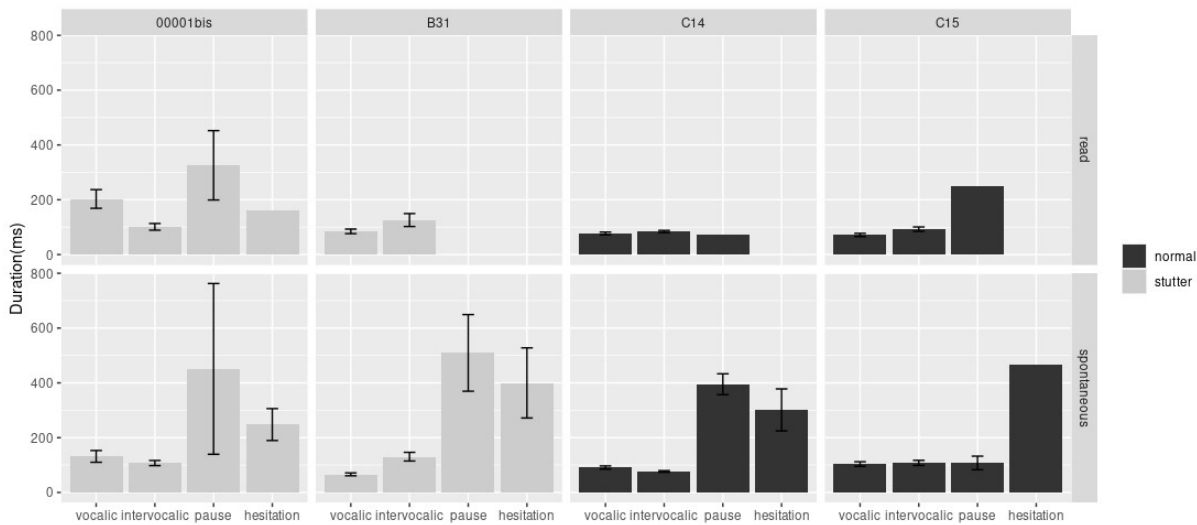


FIGURE 13 COMPARATIVE ANALYSIS OF SPEECH FOR INDIVIDUAL SPEAKERS. MEAN DURATIONS FOR VOCALIC INTERVALS, INTERVOCALIC INTERVALS, PAUSES, AND HESITATIONS

Figure 13. Caption. Detailed Speech Component Durations by Speaker: Presenting both mean values and variability, this dual-bar graph compares the duration of vocalic and intervocalic intervals, pauses, and hesitations for each of the four speakers distinguished by 'normal' and 'stutter' speech patterns. Inter-speaker variability is marked, with PWS often showing extended durations and higher variability in pauses and hesitations. The graph also illustrates significant inter-speaker variability, particularly in the stuttering group, underscoring the personalized nature of speech disruptions experienced by PWS.

PWS often have increased mean durations for pauses and hesitations, as this is a common feature of stuttered speech. This is typically seen in the higher dark gray bars within these categories. The standard error represented by the error bars is larger for PWS in certain speech elements, such as pauses and hesitations. This is indicative of greater variability in how these elements are expressed by PWS compared to PWNS.

By comparing read and spontaneous speech, PWS show a larger discrepancy in mean durations between these two types of speech, reflecting the increased challenges PWS face in spontaneous speech scenarios.

3.2. SPEECH RHYTHMICITY IN PWS

We calculated also the % of vocalic intervals, intervocalic intervals, pauses and hesitations during stuttering speech and normal speech for individual speakers, see Table 2:

TABLE 2. % OF VOCALIC INTERVALS, INTERVOCALIC INTERVALS, PAUSES AND HESITATIONS - STUTTERING AND NORMAL SPEECH FOR INDIVIDUAL SPEAKERS

	Stuttering		Normal Speech	
	B31	00001bis	C14	C15
Intervocalic	49%	44%	49.2%	47.7%
Vocalic	48.4%	48%	49.5%	50%

<i>Pauses</i>	1.3%	3.5%	0.8%	1.4%
<i>Hesitations</i>	1.3%	4%	0.5%	0.4%

Table 2. Caption. This table delineates the percentage of intervocalic and vocalic intervals, along with the frequency of pauses and hesitations, for both PWS and PWNS. It highlights that PWS tend to have a higher percentage of hesitations and a variable distribution of pauses, reflecting the stuttering characteristics. Meanwhile, the percentage of intervocalic and vocalic intervals does not significantly differ between PWS and PWNS.

In Table 2, ΔC and ΔV do not separate PWS from PWNS, however, stuttering disfluencies are numerically important. This result is not surprising since Ramus’s ΔV measure captures global variation within ips, whereas Grabe & Low PVI’s measure captures ‘pairwise’ sequential variation.

We calculate the average and the standard deviation of vocalic intervals, intervocalic intervals, duration of pauses and duration of hesitations. The numerical results in Table 2 are also visible in Figure 14:

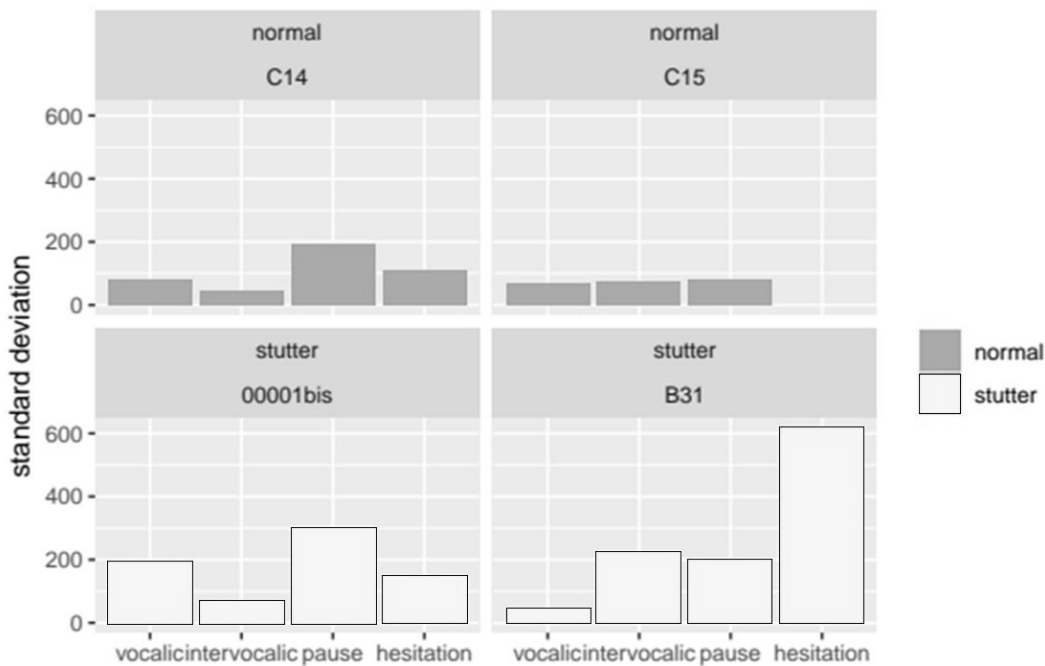


FIGURE 14 STANDARD DEVIATION OF VOCALIC INTERVALS, INTERVOCALIC INTERVALS, DURATION OF PAUSES AND DURATION OF HESITATIONS

Figure 14. Caption. Variability in Speech Components Among Individual Speakers: The graph displays the standard deviation of vocalic and intervocalic intervals, pauses, and hesitations for two PWNS (C14 and C15) and two PWS (00001bis and B31). The pronounced variability in the stuttering speakers’ intervocalic intervals and hesitations, particularly for B31, indicates the degree to which stuttering can affect speech rhythm and flow.

In PWNS the standard deviation values across all speech elements are relatively low, for the PWS, the standard deviation values are also relatively low (but higher than PWNS) and comparable to those of

the normal speakers (but higher than PWNS for intervocalic interval). A PWS speaker has a noticeably higher standard deviation in the hesitation category, which suggests a significant variability in the duration of hesitations. This could indicate moments where the speaker is experiencing blocks or is attempting to avoid disfluent moments. The standard deviation for intervocalic intervals for PWS speaker B31 shows a much larger standard deviation compared to both the normal speakers and PWS speaker 00001bis. This suggests that there is a greater variability in the timing of consonantal intervals for this speaker, which may reflect the irregular speech rhythm commonly associated with stuttering. This could be related to the stuttering disfluencies that affect the flow of speech, potentially leading to more pronounced and irregular spacing between consonantal sounds. This can be a characteristic of the speech patterns in PWS.

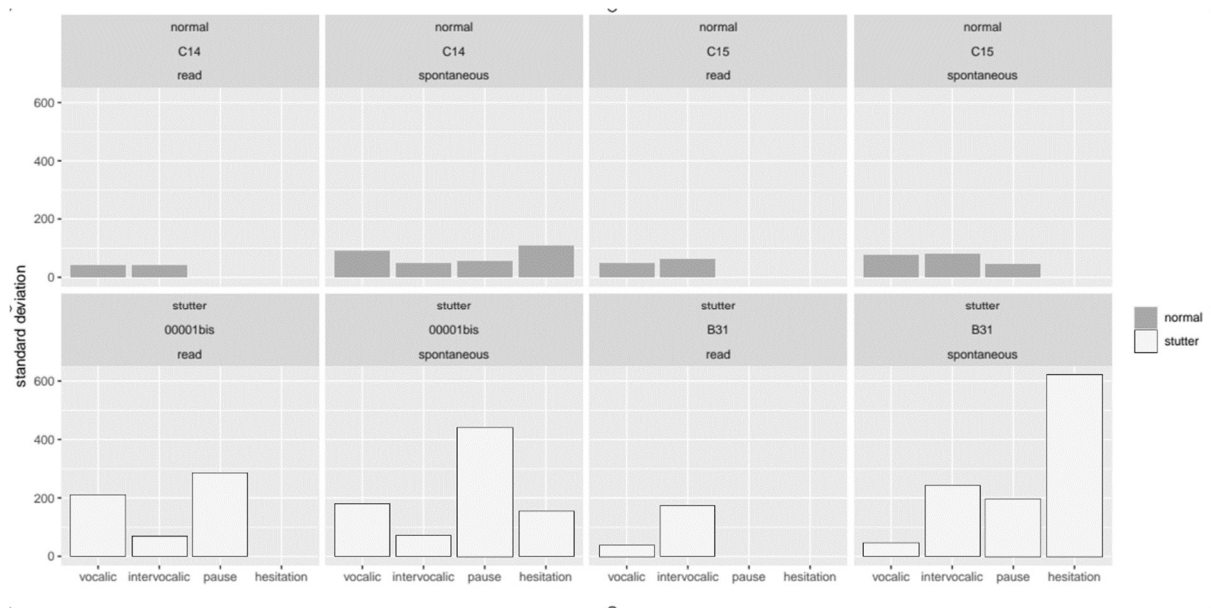


FIGURE 15 STANDARD DEVIATION OF VOCALIC INTERVALS, INTERVOCALIC INTERVALS, DURATION OF PAUSES AND DURATION OF HESITATIONS. EFFECT OF SPEECH STYLE: READ AND SPONTANEOUS

Figure 15. Caption. *Speech Component Variability Across Speaking Conditions for Individuals: In this graph, intervocalic intervals demonstrate notable variability for people who stutter (PWS), especially in spontaneous speech, with speaker B31 a heightened standard deviation in intervocalic intervals during spontaneous speech. This suggests that the timing between spoken sounds is a critical indicator of stuttering, highlighting the irregular speech rhythm and flow for PWS. Additionally, the variability in pause and hesitation durations for PWS further emphasizes the rhythmic disruptions characteristic of stuttering.*

Figure 15 displays the standard deviation of speech elements for both normal speakers and those who stutter (PWS), broken down by read and spontaneous speech styles. Both speakers exhibit low standard deviation in read and spontaneous speech styles across all speech elements. This suggests that their speech timing is fairly consistent, whether reading a text or speaking spontaneously. For PWS, during spontaneous speech, the standard deviation increases, especially for intervocalic intervals and hesitations. In a speaker who stutters (B31) during spontaneous speech, there's a notable increase in the standard deviation for hesitations. This large variability could indicate significant disruptions in speech flow due to stuttering, affecting the speaker's ability to maintain consistent hesitations. Overall, Figure 15 illustrates that while normal speakers maintain a consistent rhythm across both speech

styles, PWS exhibit more variability, particularly in spontaneous speech. This is most pronounced in the duration of their hesitations, suggesting that spontaneous speech poses more significant challenges for individuals who stutter.

The comparison of stuttering and normal speech rhythm in read and spontaneous conditions for the PVI measures is shown in the Figure 16, where the points represent the results of all utterances. Vocalic nPVI values are plotted on the vertical axis against intervocalic rPVI values on the horizontal axis:

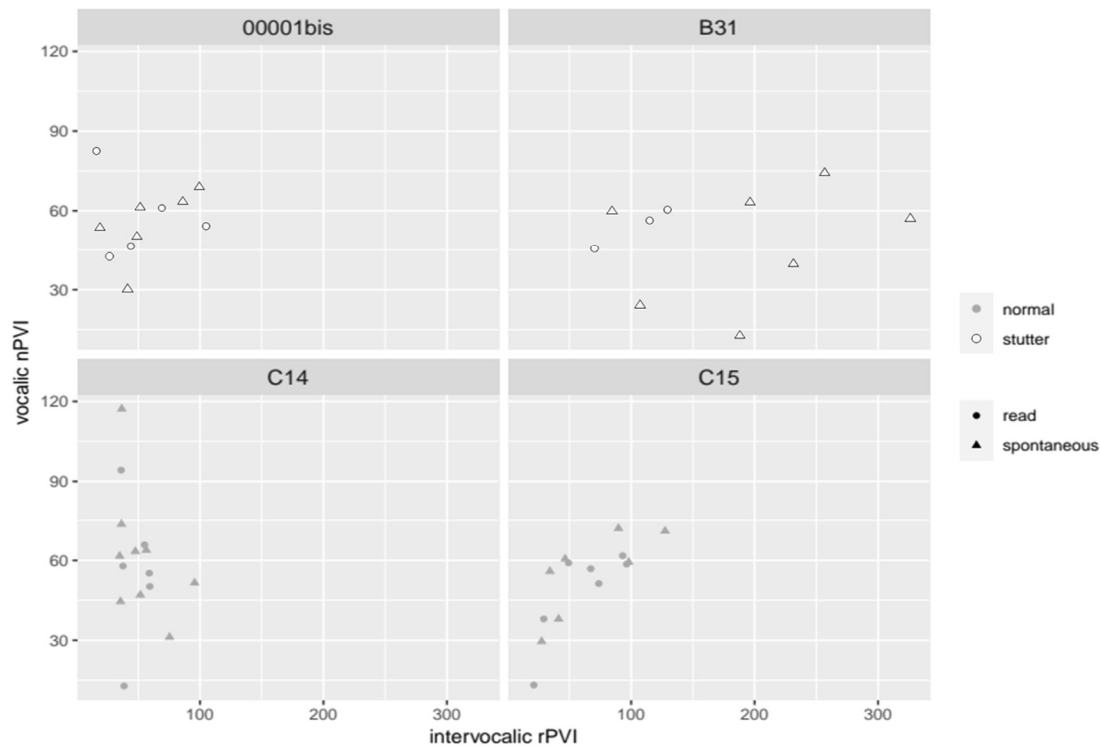


FIGURE 16. PVI RESULTS FOR ALL UTTERANCES.

Figure 16. Caption. *Pairwise Variability Index (PVI) Values by Individual and Speech Context: This scatterplot maps the vocalic nPVI against the intervocalic rPVI for four speakers, distinguishing between 'normal' and 'stutter' speech patterns during read and spontaneous speaking tasks. For both 00001bis and B31, who stutter, there is a noticeable spread in intervocalic rPVI values, particularly in spontaneous speech, indicating substantial rhythmic variability.*

In Figure 16, individual utterance measures are plotted for both normal speakers and people PWS across two speech conditions: read and spontaneous. The PVI is used to analyse the rhythmic characteristics of speech, with vocalic nPVI on the vertical axis indicating variability between vowel durations, and intervocalic rPVI on the horizontal axis indicating variability between consonant durations. It is well-known that the selection of words, and the prosodic structure of the utterances at phrasal level can result in considerable shifts in values. The tempo and style of speech (e.g., the difference between read and spontaneous speech, or the type of read text or the type of natural discourse) influenced the values that have been obtained. However, there are key elements to identify PWS and understand the rhythmic measures of nPVI and especially rPVI. The spread of the points along the horizontal axis (rPVI) shows variability in the timing of intervocalic intervals. We can observe variance of PVI values at intra-speaker level across utterances in the spontaneous speech interview. PWS have points that are spread further right, and this suggests higher variability in their intervocalic

rhythmic measure compared to normal speakers. One can observe how the speech rhythm changes from a read to a spontaneous condition. A significant shift in position, especially horizontally, indicates how stuttering impacts speech in a less controlled environment.

In Figure 17 we show the average of the PVI results for the comparison of stuttering and normal speech rhythm in read and spontaneous conditions. Vocalic nPVI values are plotted on the vertical axis against intervocalic rPVI values on the horizontal axis:

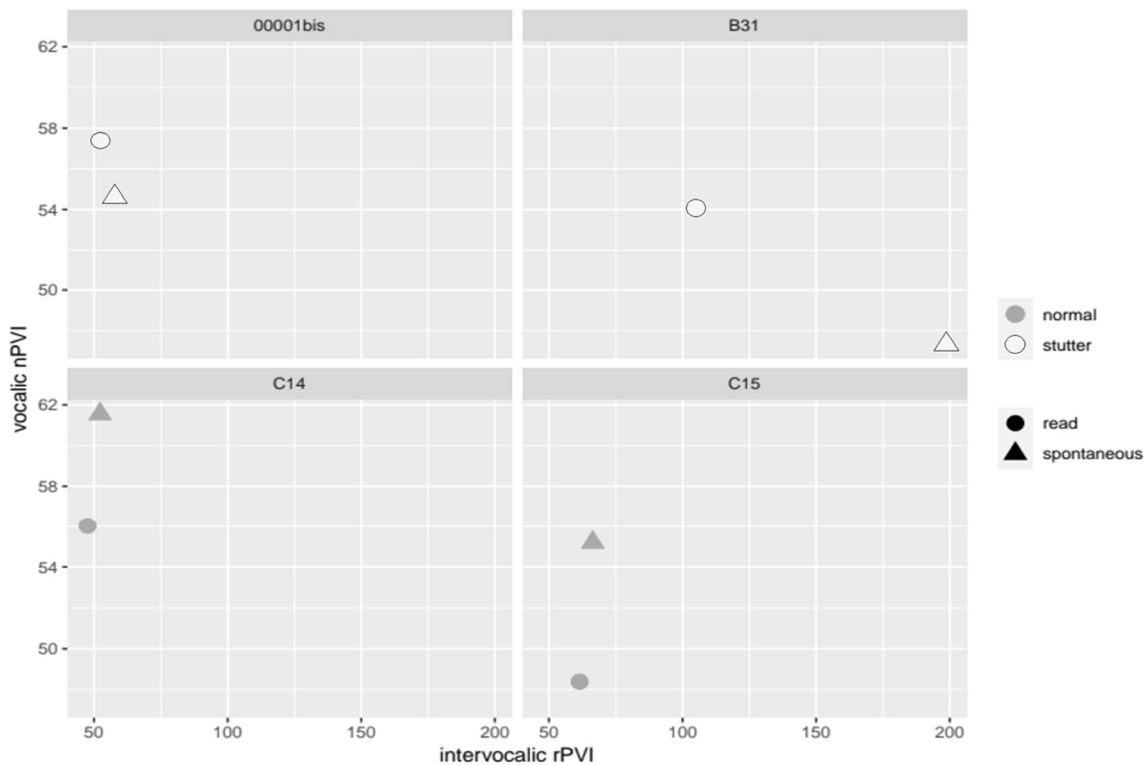


FIGURE 17. PVI RESULTS: AVERAGE

Figure 17. Caption. *Vocalic and Intervocalic Temporal Variability in Speech: This scatter plot contrasts vocalic nPVI and intervocalic rPVI across read and spontaneous speech tasks for each speaker, showcasing the temporal dynamics of speech. The plot points to differences between normal speakers (C14 and C15) and those who stutter (00001bis and B31), particularly in spontaneous speech. The individuals who stutter demonstrate wider scatter in intervocalic rPVI values, indicative of the variability in speech rhythm and timing that is characteristic of stuttering.*

In Figure 17, the average PVI results are plotted for individuals with normal speech and PWS across two different speech conditions: read and spontaneous. Higher nPVI values on the vertical axis indicate greater variability in vocalic intervals. In this figure, the nPVI values for all speakers seem to fall within a relatively close range, suggesting similar variability in vocalic interval among the speakers. Higher values on the horizontal axis indicate greater variability in intervocalic timing. PWS have higher rPVI values than normal speakers, this indicates more pronounced variability in their intervocalic interval. Normal speakers exhibit less variability in both nPVI and rPVI, indicating more regular speech rhythm, while PWS show more variability, particularly in rPVI, indicating a less regular rhythm. The PVI values, particularly the intervocalic rPVI, could be used to identify and characterize the rhythmic differences in speech related to stuttering. However, in spontaneous speech, we expect a considerably higher

degree of variation in measures due to both the material being spoken and the individual characteristics of the speaker. We have observed that personal speech-production strategies during stuttering affect rhythm measures. There will always be differences between PWS with different phonotactics and differences in durational oppositions. However, PWS demonstrate considerable variability in their speech rhythms, particularly exhibiting greater variability in the intervocalic intervals (consonants) measured by PVI-C (and ΔC), compared to people who do not stutter (PWNS).

4. CONCLUSIONS

This study offered a comprehensive analysis of speech rhythm in people who stutter (PWS) compared to those who do not stutter (PWNS), highlighting the pivotal role of rPVI (raw *Pairwise Variability Index*) as an identifier of rhythmic patterns specific to PWS (along with the measures of pauses and hesitations). This measure particularly stands out, denoting higher variability within PWS speech, thereby disrupting the expected rhythm patterns typically found in PWNS. This variability is especially pronounced in spontaneous speech conditions, suggesting that the dynamic speech production is impacted by the natural variability of spoken material and the speaking strategies of individuals. Our study underscores the critical role of the rPVI in delineating the rhythmic deviations in PWS, offering insights into the mechanisms underlying stuttering and its effects on speech production dynamics.

Our findings highlight the significant impact of stuttering on speech rhythm, particularly in spontaneous speech scenarios, which exhibit a heightened degree of variability due to both the linguistic material and the speaker's unique characteristics.

The analysis draws on the rhythm measures developed by Ramus (1999), Ramus et al. (1999) and Grabe and Low (2002), applying these to our corpus to dissect the nuanced rhythmic profiles of our participants. The data are extracted from the French ANR project (BENEPHIDIRE), as detailed in the methods section. The analysis on the corpus, encompassing interviews and read tasks performed by 14 French native speakers, underscores the variability in PWS's speech rhythm. PVI-measures (PVI-V and PVI-C) and Ramus et al. 1999 (ΔV and ΔC) both reflect the amount of variation in syllabic structure and prosodic factors which affect duration. Syllable complexity in PWS, repetitions, blocks or lengthening of sounds and syllables during speech were contributory factors to the rhythm of an utterance (expected high rPVI), and to the general rhythmic impression of stuttering speech.

In our sample of four speakers (two PWS, two PWNS), the rPVI metric emerged as a crucial element for distinguishing rhythm between groups, showcasing how stuttering introduces significant rhythmic variability, especially in intervocalic intervals. This variability reflects the inherent challenges PWS face in speech production, attributed to disruptions in neural processes governing speech and motor control. These disruptions result in variability across speech intervals, emphasizing the marked impact stuttering has on speech rhythm. This study demonstrates that these measures are adept at identifying the rhythmic deviations characteristic of stuttered speech, providing insights into the complex interplay between phonological structure, phonetic realization, and the emergent properties of rhythm.

Based on the detailed analysis presented in Section 3 of the article, it's clear that the raw Pairwise Variability Index (rPVI) notably shifts to the right for PWS when compared to people PWNS. This rightward movement on the horizontal axis, which represents intervocalic variability, underscores a key finding: PWS exhibit significantly greater variability in the timing between consonants.

PWS displayed pronounced intra-speaker variability in rhythmic values across utterances. This variability was evident in both the average values of the PVI and in the distribution of the normalized PVI mean values. The variability was markedly more in spontaneous speech than in reading conditions, underscoring the substantial influence of speech style on rhythmic measures. Particularly, intervocalic interval variability (PVI-C and ΔC) was more pronounced, likely due to variations in the onset and coda structures of syllables.

Interestingly, certain instances of speech from PWS were less variable, potentially due to more controlled motor planning processes during clinical tests. This resulted in a decreased variability in vowel duration (nPVI), which served as an indicator of fluency improvements.

The rPVI outcomes delineate a rhythm plot where PWS align closer to the 'stress-timed' end of the rhythm spectrum, diverging from the more 'syllable-timed' rhythm of PWNS. This distinction is rooted in the less isochronous nature of PWS utterances, with syllable complexity—marked by repetitions, blocks, and sound or syllable lengthening—contributing to this rhythmic deviation. The observed high variability in both vocalic and intervocalic intervals among PWS, as contrasted with PWNS, show the impact of stuttering on speech rhythm, making rPVI as a robust measure for differentiating between PWS and PWNS across various speaking tasks and styles.

Our analysis reaffirms the importance of rPVI in identifying the rhythmic patterns inherent in PWS and highlights the broader implications of rhythm measures in understanding the dynamic motor system's instability in stuttering. This contributes to our comprehension of stuttering's multifaceted impact on speech rhythm, paving the way for future research into its underlying mechanisms and potential therapeutic interventions.

The results from our analysis underscore the complexity of speech production in stuttering, which goes beyond simple durational measures. The study's findings reiterate that rhythm is a composite, emergent property rather than a simple binary classification of speech patterns.

Our chapter significantly advances our understanding of the causal mechanisms of neurodevelopmental stuttering by providing a detailed examination of speech rhythm variability in PWS compared to PWNS. By highlighting the increased variability in speech rhythm among PWS—especially in spontaneous speech—the study suggests that the neurological underpinnings of stuttering may involve disruptions or inefficiencies in the neural circuits responsible for timing and coordinating speech production. This aligns with current theories that posit stuttering as a motor timing disorder, providing empirical evidence to further this hypothesis (Ludlow & Loucks 2003; Alario et al. 2006; Kell et al. 2009; Watkins 2016; Assaneo & Poeppel 2018; Chang et al. 2019).

The observed intra-speaker variability in rhythmic measures among PWS, particularly in the context of rPVI, underscores the complexity of stuttering as a neurodevelopmental condition. It suggests that the stuttering mechanism may not solely be a result of static neural disruption but could also involve dynamic factors such as neural plasticity, attentional control, and the processing of linguistic information (Kell et al. 2009; Watkins 2016; Neumann 2017; Chang et al. 2019).

Understanding the precise nature of rhythm variability in stuttering can guide the development of targeted interventions and therapies, potentially focusing on improving rhythm perception and production as a means to mitigate stuttering symptoms.

Summary

This study provided an analysis of speech rhythm in both people who stutter (PWS) and people who do not stutter (PWNS). The analysis reveals that while rhythm measures fluctuate, the impact of stuttering on speech rhythm is substantial. Our findings indicate that rPVI measure (Pairwise Variability Index) serves as a key marker for identifying the rhythmic patterns inherent in PWS. This crucially illustrates the dynamic motor system's instability in stuttered speech.

Implications

The PVI measures reflect structural differences, which affect the emergent perceived rhythm of an utterance. Thus, we are not measuring the rhythm of PWS, but the properties which affect the potential for certain rhythmic patterns in the utterances of stuttering speech. These rhythmic

conditions can provide timing cues to overcome vulnerable pathways that affect PWS impaired neurological system.

Gains

This analysis has been also a tool to indirectly assess the severity of stuttering. It improves our comprehension of the causal mechanisms of neurodevelopmental stuttering, thus suggesting new and future pathways for research on this speech/motor disturbance.

Acknowledgments

We thank Qianwen Gwan (IR SFL 7023 CNRS) for her help with automatic data transcription and annotation.

Index terms

Index: Rhythm measures, PVI, Standard Deviation, Production, Perception, Stuttering, Speech disorders, Disfluency, Neuroscience.

5. REFERENCES

- Abercrombie, D. (1965). *Studies in Phonetics and Linguistics*. Oxford: Oxford University Press.
- Agnello, J. (1975). Voice Onset and Voice Termination features of stutterers. In L. M. Webster & L.C. Furst, eds., *Vocal tract dynamics and dysfluency*. New York: Speech and Hearing Institute, pp. 940-954.
- Alario, F. X., Chainay, H., Lehericy, S. & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, 1076(1), 129–143. DOI: [10.1016/j.brainres.2005.11.104](https://doi.org/10.1016/j.brainres.2005.11.104).
- Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*, 37(4), 325–369. DOI: [10.1016/j.jcomdis.2004.03.001](https://doi.org/10.1016/j.jcomdis.2004.03.001)
- Alm, P. A. (2021). Stuttering: A Disorder of Energy Supply to Neurons? *Frontiers in Human Neuroscience*, 15, art. 662204, 250-269. DOI: [10.3389/FNHUM.2021.662204](https://doi.org/10.3389/FNHUM.2021.662204)
- Arvaniti, A. & Hae-Sung J. (in this volume). *Individual and language differences in rhythm grouping preferences: the Iambic-Trochaic Law revisited*.
- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66(1–2), 46-63.
- Arvaniti, A. (1994). Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, 22, 239-268.
- Assaneo, M. F. & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4(2), eaao3842. DOI: [10.1126/sciadv.aao3842](https://doi.org/10.1126/sciadv.aao3842).
- Asu, E. L. & Nolan, F. (2005). Estonian rhythm and the Pairwise Variability Index. In *Proceedings FONETIK 2005*. Göteborg: Department of Linguistics, Göteborg University, pp. 29–32.
- Au-Yeung, J., Vallejo Gomez, I. & Howell, P. (2003). Exchange of Disfluency With Age From Function Words to Content Words in Spanish Speakers Who Stutter. *Journal of Speech, Language, and Hearing Research*, 46(3), 754-765. DOI: [10.1044/1092-4388\(2003/060\)](https://doi.org/10.1044/1092-4388(2003/060)).

- Auer, P., Couper-Kuhlen, E. & Müller, F. (1999). *Language in Time: The Rhythm and Tempo of Spoken Interaction*. New York: OUP.
- Barry, W. J. & Andreeva, B. (2010). Losing the trees in the wood? Reflections on the measurement of spoken-language rhythm. In M. Russo, ed., *Prosodic Universals: Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne, pp. 27–42.
- Barry, W. J., Andreeva, B. & Koreman, J. (2009). Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica*, 66(1/2), 78-94.
- Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S. & Kostadinova, T. (2003). Do Rhythm Measures Tell us Anything about Language Type? In M.-J. Solé, D. Recasens and J. Romero, eds., *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 3-9 August 2003. Barcelona: Causal Productions Pty Ltd, pp. 2693–2696.
- Barry, W. J. & Russo, M. (2003). Measuring rhythm. Is it separable from speech rate? In A. Mettouchi and G. Ferré, eds., *Proceedings AAI (acoustics, acquisition, interpretation): Prosodic Interfaces, Nantes*, 27–29 March 2003. Nantes: UFR Lettres et Langage, Université de Nantes, a.a.i. (JE2220), pp. 15-20.
- Barry, W. J. & Russo, M. (2004). Isocronia oggettiva o soggettiva? Relazioni tra tempo articolatorio e quantificazione ritmica. In F. Albano Leoni, F. Cutugno, M. Pettorino and R. Savy, eds., *Proceedings of the congress Il parlato Italiano (Naples, 13–15 February 2003)*. Napoli: D'Auria, Cdrom A02.
- Boecher, J., Franich, K. & Usler, E. (2022). Rhythm of Speech in People Who Do and Do Not Stutter - A Quantitative Analysis Using the Normalized Pairwise Variability Index. *Speech Prosody*, 2022, 23–26 May 2022, 57–61, Lisbon, Portugal. DOI: 10.21437/SpeechProsody.2022-12.
- Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R. & Fox, P. T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum Brain Mapp* 25(1), 105-17. DOI: 10.1002/hbm.20140.
- Bolinger, D. (1965). Pitch accent and sentence rhythm. In I. Abe and T. Tanekiyo, eds., *Forms of English: Accent, morpheme, order*. Cambridge, Massachusetts: Harvard University Press, pp. 139-180.
- Buhr, A. & Zebrowski, P. (2009). Sentence position and syntactic complexity of stuttering in early childhood: a longitudinal study. *J Fluency Disord*, 34(3), 155-72. DOI: 10.1016/j.jfludis.2009.08.001.
- Busan, P., Del Ben, F., Russo, L. R., Bernardini, S., Natarelli, G., Arcara, G. et al. (2019). Stuttering as a matter of delay in neural activation: A combined TMS/EEG study. *Clinical Neurophysiology*, 61-76. DOI: 10.1016/j.clinph.2018.10.995.
- Chang, S. E., Garnett, E. O., Etchell A. & Chow, H. M. (2019). Functional and Neuroanatomical Bases of Developmental Stuttering: Current Insights. *Neuroscientist*, 25(6), 566-582. DOI: 10.1177/1073858418803594.
- Crocq, M.-A., Guelfi, J.-D., Boyer, P., Pull, C.-B. & Pull-Erpelding, M.-C. (2015). *DSM-5® : Manuel diagnostique et statistique des troubles mentaux (BU Lyon Sud WM 141 DSM; 5th edn)*. Issy-les-Moulineaux: Elsevier Masson.
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1), 16-28.
- Cutler, A. (1994). Segmentation problems rhythmic solutions. *Lingua*, 92, 81-104.

- Couper-Kuhlen, E. (1993). *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Amsterdam: Benjamins.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm, *11th International Congress of Phonetic Sciences*. Tallinn, Estonia: U.S.S.R. / Academy of Science of the Estonian S.S.R., vol. 5, 447-450.
- Dellwo, V. (2008). The role of speech rate in perceiving speech rhythm. In P. A. Barbosa, S. Madureira and C. Reis, eds., *Proceedings of Speech Prosody 2008, 4th Conference on Speech Prosody*, Campinas 6-9 May 2008, Campinas (Brazil). São Paulo: Capes, Fapesp, CNPq - Campinas: IEL, Unicamp, Motorola, Editora RG, cdrom, pp. 375-378.
- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski and I. Sziget, eds., *Language and language-processing*. Frankfurt/Main: Peter Lang, pp. 231-241.
- Deterding, D. (2001). The Measurement of Rhythm: A Comparison of Singapore and British English. *Journal of Phonetics*, 29(2), 217-230.
- Didirková, I., Le Maguer, S. & Hirsch, F. (2021). An articulatory study of differences and similarities between stuttered disfluencies and non-pathological disfluencies. *Clinical Linguistics & Phonetics*, 35(3), 201-221. DOI: 10.1080/02699206.2020.1752803.
- Dommelen, W. A. van (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, 33, 259-272.
- Domingues, C. E. F. & Drayna, D. (2015). The Genetics of Stuttering. In *Encyclopedia of Life Sciences*, Chichester: John Wiley & Sons Ltd, pp. 1-8. DOI: 10.1002/9780470015902.a0025254.
- Etchell, A. C., Civier, O., Ballard, K. J. & Sowman, P. F. (2018). A systematic literature review of neuroimaging research on developmental stuttering between 1995 and 2016, *Journal of Fluency Disorders*, 55, 6-45. DOI: [10.1016/j.jfludis.2017.03.007](https://doi.org/10.1016/j.jfludis.2017.03.007).
- Fuchs, R. (in this volume). *Applying duration-based and acoustic speech rhythm metrics*.
- Gibbon, D. & Gut, U. (2001). Measuring speech rhythm. In P. Dalsgaard, B. Lindberg and H. Benner, eds., *Proceedings of Eurospeech 2001-Scandinavia*, Aalborg, Denmark 3-7 September 2001, Aalborg: Kommunik Grafiske Losninger A/S, vol. 1, pp. 95-98.
- Didirková, I. & Hirsch F. (2020). A two-case study of coarticulation in stuttered speech. An articulatory approach. *Clinical Linguistics & Phonetics*, 34(20), 517-535. DOI: 10.1080/02699206.2019.1660913.
- Eriksson, A. (1992). *Aspects of Swedish Speech Rhythm*. PhD dissertation: University of Göteborg.
- Falk, S., Müller, T. & Dalla Bella, S. (2015). Non - verbal sensorimotor timing deficits in children and adolescents who stutter. *Frontiers in Psychology*, 6, 847. DOI: 10.3389/fpsyg.2015.00847.
- Goswami, U. (in this volume). *Rhythm and Language Acquisition: A Temporal Sampling Perspective*.
- Grabe, E. & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven and N. Warner, eds., *Laboratory Phonology 7*. The Hague: Mouton de Gruyter, pp. 515-546.
- Grabe, E., Post, B. & Watson, I. (2000). Acoustic correlates of rhythm in English and French four-year-olds. *Oxford University Working Papers in Linguistics, Philology & Phonetics, Papers from the Phonetics Laboratory*, 5, 7-17.

- Grabe, E., Post, B. & Watson, I. (1999). The acquisition of rhythm in English and French. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville and A. C. Bailey, eds., *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco 1-7 agosto 1999*. San Francisco: University of California and Berkeley, vol. 2, pp. 1201-1204.
- Grabe, E. (2002). Variation adds to prosodic typology. In B. Bel and I. Marlien, eds., *Proceedings of the 1st International Conference on Speech Prosody (Aix-en-Provence 11-13 April 2002)*. Aix-en-Provence: Laboratoire Parole et Langage, pp. 127-131.
- Guitar, B. (2013). *Stuttering: An integrated approach to its nature and treatment*. Baltimore: Williams & Wilkins.
- Harrington, J. (1987). Coarticulation and Stuttering: An Acoustic and Electropalatographic Study. In H. F. M. Peters and W. Hulstijn, eds., *Speech motor dynamics in stuttering*. Vienna: Springer, pp. 381–392.
- Heyde, C. J., Scobbie, J. M., Lickley, R. & Drake, E. K. (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. DOI: [10.3109/02699206.2015.1100684](https://doi.org/10.3109/02699206.2015.1100684).
- Hesse, T. (2023). Developmental stuttering may be caused by insufficient processing of auditory feedback. *Medical Hypotheses*, vol. 180. DOI: 10.1016/j.mehy.2023.111166.
- Hoeqvist, Ch. Jr. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, 40, 203-237.
- Howell, P., Au-Yeung, J. & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language, and Hearing Research*, 42, 345-54.
- Jerbi, K. & Boulenger, V. (in this volume). Neural synchronization to the syllabic rate in speech.
- Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H. & Giraud, A. L. (2009). How the brain repairs stuttering. *Brain*, 132(Pt 10), 2747-60. DOI: 10.1093/brain/awp185.
- Kohler, K. J. (2009). Rhythm in Speech and Language. A New Research Paradigm. *Phonetica*, 66(1/2), 29-45. DOI: 10.1159/000208929.
- Koreman, J., Van Dommelen, W., Sikveland, R., Andreeva, B. & Barry, W.J. (2009). Cross-language differences in the production of phrasal prominence in Norwegian and German. In M. Vainio, R. Aulanko and O. Aaltonen, eds., *Proceedings of 10th Nordic Prosody conference (Helsinki, 2008)*, Frankfurt: Peter Lang, pp. 139-150.
- Lee, C. S. & McAngus Todd, N. (2004) Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition*, 93(3), 225-254.
- Lickley, R. (2018). Disfluency in typical and stuttered speech. In C. Bertini, C. Celata, G. Lenoci, C. Meluzzi and I. Ricci, eds., *Fattori sociali e biologici nella variazione fonetica / Social and Biological Factors in Speech Variation*. Milano: Officinaventuno, pp. 373–387. DOI: 10.17469/O2103AISV000019.
- Lloyd, J. A. (1940). *Speech Signals in Telephony*. London: Pitman & Sons.
- Low, E. L. (1998). *Prosodic prominence in Singapore English*. PhD thesis: University of Cambridge.
- Low, E. L. & Grabe, E. (1999). A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech*, 42(1), 39-56.
- Low, E. L., Grabe, E. & Nolan, F. (2001). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, 43(4), 377-401.

- Ludlow, C. L. & Loucks, T. (2003). Stuttering: a dynamic motor control disorder. *Journal of Fluency Disorders*, 28, 273–295.
- Manrique, A. M. B. & Signorini, A. (1983). Segmental reduction in Spanish. *Journal of Phonetics* 11: 117–128.
- 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(3): 524–542. [https://doi.org/10.1044/1092-4388\(2005/036\)](https://doi.org/10.1044/1092-4388(2005/036)).
- Mandke, K. & Rocha, S. (in this volume). *Neural and behavioural rhythmic tracking during language acquisition: The story so far*.
- Mok, P. P. K. & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In P. A. Barbosa, S. Madureira and C. Reis (eds.). *Proceedings of Speech Prosody 2008, 4th Conference on Speech Prosody, Campinas 6-9 May 2008, Campinas (Brazil), São Paulo: Capes, Fapesp, CNPq - Campinas: IEL, 423–426, Unicamp, Motorola, Editora RG, cdrom*.
- Monfrais-Pfauwadel M.-Cl. (2014). *Begaiement, begaiements*. Louvain: De Boeck Sup.
- Neumann, K., Euler, H. A., Bosshardt, H. G., Cook, S., Sandrieser, P., Sommer, M. (2017). The Pathogenesis, Assessment and Treatment of Speech Fluency Disorders. *Deutsches Ärzteblatt International*, 114(22-23), 383-390. DOI: 10.3238/arztebl.2017.0383.
- Nolan, F. & Asu, E. L. (2009). The Pairwise Variability Index and Coexisting Rhythms in Languages. *Phonetica*, 66(1/2): 64–77.
- Onslow, M. (2020). *Stuttering and its' treatment*. Sidney: Eleven lectures.
- Orpella, J., Flick, G., Assaneo, M. F., Shroff, R., Pykkänen, L., Poeppel, D. & Jackson, E. S. (2024). Reactive Inhibitory Control Precedes Overt Stuttering Events. *Neurobiology of Language*. Advance publication. DOI: 10.1162/nol_a_00138.
- Peters, H. F. M., Hulstijn, W. & Van Lieshout, P. (2000). Recent Developments in Speech Motor Research into Stuttering. *Folia Foniatrica et Logopedia*, 52, 1–3.
- Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- R Core Team (2021). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Riaz, N., Steinberg, S., Ahmad, J., Pluzhnikov, A., Riazuddin, S., Cox, N. J. & Drayna, D. (2005). Genomewide significant linkage to stuttering on chromosome 12. *The American Journal of Human Genetics*, 76(4), 647-51. DOI: 10.1086/429226.
- Ramus, F. (1999). *Rythme des langues et acquisition du langage*. Doctoral Thesis: EHESS (Paris).
- Ramus, F., Nespors, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Riley, G. (1994). *The stuttering severity instrument for adults and children (SSI-3)* (3rd ed.). Austin, TX: PRO-ED.
- Rimmele, J. (in this volume). *Brain Spectral Profiles and their Role for Language*.

- Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal, ed., *Linguistic controversies*. London: Edward Arnold, pp. 73-79.
- Russo, M. & Barry, W. J. (2004). In che misura l'italiano è 'isosillabico'? Una comparazione quantitativa tra l'italiano e il tedesco. In P. D'Achille, ed., *Generi, architetture e forme testuali*, Proceedings of the 7th Congress of the *Società Internazionale di Linguistica e Filologia Italiana (SILFI)*, Roma, 1-5 October 2002. Firenze: Cesati, pp. 387-401.
- Russo, M. & Barry, W. J. (2008a). Calcul rythmique et quantification de la vitesse d'élocution, 27th *Journées d'Études sur la Parole (JEP'08)*, Avignon from 9th to 13th June 2008, AFCP (Association Francophone de la Communication Parlée) editor/ISCA, cdrom n°1562.
- Russo, M. & Barry, W. J. (2008b). Isochrony reconsidered. Objectifying relations between Rhythm Measures and Speech Tempo. In P. A. Barbosa, S. Madureira and C. Reis, eds., *Proceeding of Speech Prosody 2008*, Fourth Conference on Speech Prosody, May 6-9 2008, Campinas (Brazil), São Paulo: Capes, Fapesp, CNPq – Campinas: IEL, Unicamp, Motorola, Editora RG, cdrom.
- Russo, M. & Barry, W. J. (2010). Il pairwise Variability Index (PVI e PVIs): valori ritmici per i dialetti italiani e per l'italiano regionale. Implicazioni tipologiche. In M. Russo (ed.) *Prosodic Universals: Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne, pp. 185–226 (Biblioteca di linguistica, resp. Massimo Arcangeli).
- Russo, M. (2010). *Introduzione*. In Ead., ed., *Prosodic Universals: Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne, pp. 9–26.
- Salmelin, R., Schnitzler, A., Schmitz, F., Jäncke, L., Witte, O. & Freund, H. J. (1998). Functional organization of the auditory cortex is different in stutterers and fluent speakers. *Neuroreport*, 9: 2225-2229.
- Van Lieshout, P. H., Hulstijn, W. & Peters, H. F. M. (1996). From planning to articulation in speech production: What differentiates a person who stutters from a person who does not stutter? *Journal of Speech and Hearing Research*, 39(3), 546–564. DOI: 10.1044/jshr.3903.546.
- Wagner, P. (2010). A time-delay approach to Speech Rhythm Visualisation Modeling and Measurement. In M. Russo, ed., *Prosodic Universals: Comparative Studies in Rhythmic Modeling and Rhythm Typology*. Rome: Aracne, pp. 117-146 (Biblioteca di linguistica, resp. M. Arcangeli).
- Watkins, K. E., Chesters, J. & Connally, E. L. (2016). Chapter 79 - The Neurobiology of Developmental Stuttering. In G. Hickok and S. L. Small, *Neurobiology of Language*. Amsterdam: Academic Press, pp. 995-1004, ISBN 9780124077942, DOI: 10.1016/B978-0-12-407794-2.00079-1.
- Wenk, B. & Wiolland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.
- Yairi, E. & Ambrose, N. G. (2013). Epidemiology of stuttering: 21st century advances. *Journal of Fluency Disorders*, 38(2), 66–87. DOI: [10.1016/j.jfludis.2012.11.002](https://doi.org/10.1016/j.jfludis.2012.11.002).
- Zhang, Y., Zou, J. & Ding, N. (in this volume). *Analyses of the rhythms of phonemes and syllables*.
- Zimmermann, G. (1980). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech, Language and Hearing Research*, 23(1), 95–107. DOI: 10.1044/jshr.2301.95.