

Journal Pre-proof

Rate vs. rhythm characteristics of cluttering with data from a
“syllable-timed” language

Judit Bóna, Anna Kohári



PII: S0094-730X(20)30056-5
DOI: <https://doi.org/10.1016/j.jfludis.2020.105801>
Reference: JFD 105801

To appear in: *Journal of Fluency Disorders*

Received Date: 13 March 2019
Revised Date: 14 July 2020
Accepted Date: 28 October 2020

Please cite this article as: Bóna J, Kohári A, Rate vs. rhythm characteristics of cluttering with data from a “syllable-timed” language, *Journal of Fluency Disorders* (2020), doi: <https://doi.org/10.1016/j.jfludis.2020.105801>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Rate vs. rhythm characteristics of cluttering with data from a “syllable-timed” language

Judit Bóna (Corresponding author)

Department of Applied Linguistics and Phonetics
ELTE Eötvös Loránd University

H-1088 Budapest, Múzeum krt 4/A
Hungary

bona.judit@btk.elte.hu
+36/1/ 411 6500

Anna Kohári

Research Institute for Linguistics

H-1068 Budapest, Benczúr u. 33.
Hungary

kohari.anna@nytud.hu
+36/1/3429372

Highlights

- Articulation rate and speech rhythm in cluttered and control speech were analyzed in a “syllable-timed” language.
- Results show only slight differences between the groups in certain parameters of speech rhythm.
- The timing differences between cluttered and control speech are manifested primarily in the articulation rate.
- Peculiarities in speech rhythm are almost negligible in cluttered speech.

Introduction

Cluttering is a type of fluency disorder characterized by a speech rate which is perceived to be fast and/or irregular as well as by an abnormal speech rhythm (St. Louis & Schulte 2011: 241-242). Speech rate and articulation rate (i.e. average speech tempo including and excluding pauses,

respectively) are of central importance for understanding cluttering. Previous research (St. Louis, 1992; St. Louis et al. 2007; van Zaalén-op't Hof et al. 2009; St. Louise–Shulte 2011; Bakker et al. 2011; Oliveira et al. 2013) has confirmed that speech rate and articulation rate are significantly faster in cluttering than in control speech. However, there are no differences between cluttering, exceptionally rapid speech and control speech if speakers have to speak as fast as possible (Baker et al. 2011).

Non-typical speech rhythm appears in the early definitions of cluttering and in its diagnostic criteria (Daly 2006). The Predictive Cluttering Inventory (Daly 2006) lists a number of symptoms related to speech rhythm such as *irregular speech rate; (the speaker) speaks in spurts or bursts; telescopes or condenses words; speech rate progressively increases (festinating); variable prosody; irregular melody or stress pattern*. The lowest common denominator definition of cluttering also states that cluttering can be accompanied by excessive collapsing or deletion of syllables (St. Louis & Schulte 2011: 241-242). In addition, perception of the rhythm of cluttered speech is affected by the occurrence of pauses different from that of control speech (Daly 2006; St. Louis et al. 2007; Myers & St. Louis 2007; Reichel 2007; Scott et al. 2007; Logan, 2010; Lanouette 2011). Although phonetic analyses show that the location and frequency of pauses in cluttered speech do not differ from those in control speech, their durations are significantly shorter (Bóna 2016). In the following subsections, first, the characteristics of articulation rate will be presented, followed by those of cluttering and speech rhythm. Finally, the relationship and difference between the two will be discussed.

1.1 Articulation rate

Articulation rate measures the speed of pronunciation. To acquire this parameter, the duration of pauses is not considered, meaning that the number of words or syllables in a given unit of time (generally, a minute or second) is calculated, but the duration of pauses is excluded from the analysis (Fletcher 2010). There are several factors affecting speech and articulation rate. One of these is the language itself; different languages are characterized by different speech and articulation rates. For example, the mean articulation rate of the Dutch in the Netherlands is 5.05 syllables/s (Verhoeven et al. 2004), while it is 4.3 syllables/s for American English native speakers (Morill et al. 2016), and 5.8 syllables/s for Hungarians (Bóna 2014). Although these data are not completely comparable between the different sources, it is clear from the examples above that each language has its own typical articulation rate. Hungarian, for example, is faster than American English. This can be observed by perception since noticeable differences in articulation are at about 5% (Quené 2007).

As for the articulation rate in cluttering, it is higher than in control speech. Van Zaalen-op't Hof et al. (2009) found a high articulation rate in 56% of the people observed who clutter (hereafter, PWC). The speech task performed also influenced the articulation rate: there was a significant difference between PWC and people who stutter (PWS) or control speakers in spontaneous speech, while there were no significant differences between the groups when reading or retelling a memorized story. According to the results of Oliveira et al. (2013), there were significant differences in speech rate between PWC and control speakers: speech rate was 295.28 syllables per minute and 179.46 words per minute in PWC, while 247.62 syllables per minute and 141.35 words per minute in controls. The variability of articulation rate is also larger in cluttering than in control speech. Out of the essential symptoms of cluttering, fast speech rate is the most important one (first on the list), while irregular speech rate is the third, as reported by both UK and USA

clinicians (St. Louise & Rustin 1992). St. Louise (1996) summarized data of 29 PWC of 12 papers, published in a special issue of the Journal of Fluency Disorders. He found that there were only four cases in which PWC were not characterized by irregular speech rate. Temporal and spatial variability in the speech of PWC and control speakers were analyzed by means of electromagnetic midsagittal articulography by Hartinger & Mooshammer (2008). They found that when producing loan words (long words with a complex syllable structure), the variation was higher in PWC both in the temporal and in the spatial domain than in control speakers. Results of van Zaalen-op't Hof et al. (2009) also showed that PWC produced more errors in accuracy in repeating multi-syllable word strings than PWS or control speakers.

1.2 Speech rhythm from an acoustic phonetic aspect

Although speech rhythm has been researched for many decades, no general consensus on its exact definition has been reached (see Turk Shattuck-Hufnagel 2013). Speech rhythm can be connected with several acoustic characteristics of speech (changes in articulation rate, intensity, and fundamental frequency). It is highly probable that not only the appearance or the absence of one of these features create the rhythm, but they form a complex mechanism together (Cumming 2011). The alternation of stressed and non-stressed syllables, the changes in intensity (volume), and changes of fundamental frequency (intonation) are interconnected properties (Cumming 2011; Tilsen–Johnson 2008). In the field of experimental phonetics, the term ‘speech rhythm’ often refers to systematic surface timing patterns which means that the durations of speech sounds affect the perception of rhythm (e.g. Ramus et al. 1999, Grabe–Low 2002, Dellwo 2010).

According to a classical theory, languages can be divided into three rhythm classes based on perceived characteristics: ‘stress-timed’, ‘syllable-timed’, and ‘mora-timed’. For example,

English, German, and Dutch have been classified as stress-timed, Hungarian, Italian, and Spanish have been described as ‘syllable-timed’, while Japanese has been categorized as a mora-timed language (Pike 1945; Bloch 1950; Abercrombie 1967). Some early approaches assumed that isochronous or near-isochronous units of speech could be found in all spoken languages, although the types of units are different in the three classes of rhythm. It was hypothesized that the durations of syllables are roughly equal in syllable-timed languages, just as the durations of the interstress intervals in stress-timed languages (Abercrombie 1967). Moreover, morae were described to be near-equal in duration in languages classified as mora-timed ones (Ladefoged 1975). However, experimental works have failed to confirm the existence of such isochronous duration patterns in the three classes of rhythm (e.g. Roach 1982; Dauer 1983; Hoequist 1983).

According to Dauer (1983), differences in the speech rhythm of different languages are the result of different combinations of phonological, phonetic, lexical and syntactic factors. The basic principle of the theory is that the structure of syllables is one of the factors that determines the rhythm of speech. Greater variability of syllable types is characteristic of stress-timed languages, while fewer different syllable types are observed in syllable-timed languages. Another important factor is lexical stress. It is assumed that the duration of vowels is shorter in an unstressed position than in a stressed one, and sometimes their formant structure is reduced to a schwa. This phenomenon is more common in stress-timed languages than in syllable-timed languages. Speech rhythm might be seen as a continuum with syllable-timed languages at one end and stress-timed languages at the other (Dauer 1987).

In the past two decades, novel methods have been proposed to quantify the perceived and supposed differences between stress-timed, syllable-timed, and mora-timed languages. The acoustic measurements have verified that some of them can capture the overall rhythmic characteristics of

a language thus, the stress-timed vs. syllable-timed distinction (Ramus et al. 1999, Grabe–Low 2002). In this framework, a consonantal interval is defined as a block of successive consonants that is followed and preceded by vowels (and analogously for a vocalic interval). The archetypical stress-timed languages showed a higher degree of durational variability compared to the ones traditionally classified as syllable-timed languages. In other words, the differences in the duration of vocalic and consonantal intervals within the same utterance are greater in stress-timed languages than in syllable-timed languages. The large durational variability of vocalic intervals (quantified by ΔV , see Methods) implies that in the given language, stressed and unstressed positions influence vowel durations to a great extent. In addition, stress-timed languages typically have a high complexity in consonant clusters which can lead to a high variability in consonantal interval duration. Using other types of rhythm metrics revealed further rhythmic characteristics of different languages. The percentage of vocalic interval duration in an utterance (%V) tended to be higher in syllable-timed languages than in stressed-time languages (Ramus et al. 1999). Low values of %V imply that consonant clusters over-dominate the syllable structure. Grabe and Low (2002) proposed further metrics (PVI) which were able to capture the durational alternation among adjacent consonantal or vocalic intervals. The findings based on PVI measures were in agreement with the traditional classification, but the evidence from this study did not support the fact that languages could be classified into distinct rhythm categories. In order to solve this issue, it has been suggested that languages should be placed on a rhythmic continuum based on the applied metrics (Grabe–Low 2002). Based on rhythm metrics, Hungarian shows similar rhythmic features to languages which have been classified as syllable-timed languages as per the classical theory. The differences in the duration of vowels within the same utterance were measured to be lower

compared to other languages, such as English or Dutch. The same observation has been made about consonantal interval durations as well (White–Mattys–Wiget 2012, Kohári 2018).

Like any new method, the rhythm metrics described above have received some criticism. Recent critical studies have highlighted that speech style (sentence reading, story reading, or spontaneous speech) and sample size have an impact on the values of the metrics (Arvaniti 2012). Taking these effects into consideration, the rhythm metrics can be useful tools to investigate the durational dimension of speech rhythm within a language (cf. White–Mattys–Wiget 2012, Nolan–Jeon 2014). Until now, this methodology has not been applied to investigating cluttered speech.

1.3 Relationship between articulation rate and speech rhythm

It is important to clarify that articulation rate and speech rhythm are two distinct facets of the temporal structure of speech, referred to as the so-called "surface duration" properties in the aforementioned theoretical framework. Both aspects investigate the durational dimension of speech as elementary units based on segments or intervals. Articulation rate is the measure of the typical, average tempo of speech acquired via calculating the total number of the investigated durational units (sounds, syllables, etc.) divided by the elapsed time (Fletcher 2010), whereas speech rhythm is related to the patterns of internal durational variability around the mean (Ramus et al. 1999, Grabe–Low 2002).

The original "classic" metrics of speech rhythm were therefore introduced to quantify the perceived speech rhythm e.g. by taking the standard deviations of unit durations (e.g. ΔV , ΔC). It has been found, however, that the calculated values of these metrics showed significant dependence on the articulation rate itself, due to easily understandable reasons (Dellwo 2010). For a didactical example, let us assume we have a 'slow' and 'fast'-rate speech sample with identical

rhythm (relative to the speech tempo). Then the one with slower articulation rate would yield larger standard deviations values. To avoid this kind of bias, speech rate-normalized speech rhythm metrics (e.g. VarcoV, VarcoC see details in Methods) have also been developed (Dellwo 2010), which are adjusted to (and thus do not trivially correlate with) articulation rate. Therefore, the temporal patterns, irregularities can be captured by speech rhythm metrics independently from articulation rate.

To the best of our knowledge, no research has been conducted so far using objective measurements within the framework described above and on the rhythm of cluttered speech. There are almost only subjective (perceptual) observations about the influence of the deletion of sounds or syllables on speech rhythm in cluttering.

The aim of this study is to show by objective measurements whether there are any differences between the rhythm of cluttering and control speech, and which parameters point to such differences.

The hypotheses were the following: 1) The rhythm of cluttering is different from that of control speech based on phonetic measurements, meaning that the objective measurements are able to confirm the subjective perception of the difference between the two groups in speech rhythm. 2) There are differences in the duration of both vocalic and consonantal intervals between PWC and control speakers.

2. Methods

2.1. Participants

To answer the questions, spontaneous narratives of PWC were recorded and compared to the speech of control speakers. 10 PWC (4 women and 6 men) and 10 control speakers (5 women and 5 men) participated in the experiment. All were aged between 20 and 32. (Mean age of PWC was 26.0, SD: 3.9. Mean age of control speakers was 24.7, SD: 3.5.) All were native Hungarian speakers with normal hearing, and they had at least 14 years of education. PWC were recruited by acquaintances and speech therapists for the study. Control speakers were participants in the recordings of BEA Hungarian Speech Database (Gósy 2012). Two experts in fluency disorders, a speech-language pathologist, and a linguist, specialized in fluency disorders, determined the diagnostic decisions independently of each other. The participants were ordered in two groups. The rate of agreement of the two experts was 100%. Like in Bakker et al. (2011), subjects were classified as PWC, if they had perceptually rapid and/or irregular speech rate, and their speech was characterized by at least one of the following (Bakker et al. 2011): (1) excessive disfluencies (in most cases non stuttering-like, but other disfluencies), and/or (2) specific articulation characteristics (which manifested in overly coarticulated speech or omissions of sounds and syllables; but it was not dyslalia, for example; van Zaalen- op 't Hof et al. 2009; Bakker et. al 2011).

In our analysis, PWC were pure clutterers, which means that they did not stutter or had any comorbid speech and language disorders. They also had no history of stuttering in the past. All of them considered themselves PWC because their speech problems were often pointed out by their speech partners.

2.2. Material

Spontaneous speech of each participant was recorded at their most convenient tempo. They spoke about familiar topics which did not require special knowledge or previous preparation: education, work, hobby, family. The interviewer let the participants speak freely. She only spoke to ask questions to help participants continue their narrative. In the other task for PWC, they had to speak about their university education and about the requirements of getting a degree as slowly as they could. Speech samples were recorded in a soundproof room by a digital recorder (44.1 kHz, 16 bit). We intended to analyze approximately two-minute speech samples from each recording. At the beginning of the interviews, speakers often thought about what to say and, as a result, their speech was slower and more disfluent. However, in the middle of the recordings, speakers had already “warmed up” and they were quite comfortable with speaking. For this reason, the speech samples for analysis were taken from the middle of the recordings.

2.3. Procedures

Firstly, the speech samples were annotated at the level of speech units i.e. from pause to pause using Praat 5.1 (Boersma & Weenink 2009). Based on this annotation, the recordings were annotated at the level of sounds by the Hungarian adapted version of MAUS, a software for automatic segmentation (Schiel 1999). The identity and boundaries of speech sounds as well as the silent and filled pauses were annotated. Finally, the boundaries of speech sounds were checked and corrected manually. The boundaries of vowels were adjusted to the beginning and end of the formant structure on the basis of oscillogram, spectrogram and auditory information, in line with the conventions (Peterson–Lehiste 1960, Grabe–Low 2002). The duration of voiceless plosives occurring after a pause was determined in a way that the closure duration of the next plosive with

similar articulation in the same speech sample was measured. This duration was then added to the closure duration. In Hungarian, the realization of vowels with non-periodic voices is quite frequent (Markó 2014). These units were annotated as parts of vowels. The occurrences of disfluency were not considered for further analysis (cf. Arvaniti 2012). Each boundary, set by MAUS, was checked by the first author. The second author then selected 10% of the whole sample randomly and rechecked it. The similarity of the annotation of the two authors was 98%. In the remaining 2%, sound boundaries were set by additional acoustic and perceptual analysis. We analyzed 10 utterances from each participant which were considered syntactically correct, complete and not interrupted by pauses. Each chosen utterance contained at least 10 syllables. First, the articulation rate was analyzed since it also has an effect on speech rhythm. We analyzed 200, at least 10-syllable-long utterances which contained 2835 syllables and 6472 speech sounds.

Following the annotation of the recordings, we calculated speech rhythm metrics, in line with the widely used methodology (see e.g. Grabe–Low 2002, Dellwo 2010, Arvaniti 2012). Instead of individual phonemes, these metrics were based on the durations of consonantal and vocalic intervals. A consonantal interval was defined as an uninterrupted sequence of consonants (either a single consonant or a cluster of successive consonants). Analogously, a sequence of vowels (i.e. one or more successive vowels) was labelled a vocalic interval (Ramus et al., 1999, Grabe–Low 2002). Thus, speech was considered to be an alternation of vocalic and consonantal intervals.

The following values were calculated (Grabe–Low 2002, Dellwo 2010, Arvaniti 2012, White–Mattys–Wiget 2012):

- ΔV : the standard deviation of vocalic interval duration
- ΔC : the standard deviation of consonantal interval duration

- %V: sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals
- VarcoV: the standard deviation of vocalic interval duration divided by the mean vocalic duration and multiplied by 100
- VarcoC: the standard deviation of consonantal interval duration divided by the mean consonantal duration and multiplied by 100
- rPVI-C: raw Pairwise Variability Index for consonantal intervals
- nPVI-V: normalized Pairwise Variability Index for vocalic intervals
- nPVI-C: normalized Pairwise Variability Index for consonantal intervals
- nPVI-CV: normalized Pairwise Variability Index for vocalic and consonantal intervals

Some of the metrics above (ΔV , ΔC , rPVI-C) correlate markedly with articulation,- or speech rate (Grabe–Low 2002, Dellwo 2010), therefore, we also used normalized versions of the speech rhythm metrics (VarcoV, VarcoC, nPVI-V, nPVI-C, nPVI-CV, %V) adjusted to be independent of articulation or speech rate (cf. Dellwo 2010).

The various PVI metrics are useful additional tools for exploring the temporal patterns within utterances as they calculate the average differences between successive intervals and, hence, are sensitive to the order of the intervals. All rhythm measures were calculated for such utterances using a program written in C++ environment.

Statistical analysis was carried out by R 3.4.3 (R Core Team 2017). The normality of the data was assessed using the Shapiro-Wilk normality test with a significance level of 0.05 and graphical techniques. Results revealed that most of the variables were not normally distributed except for

three measurements: %V for PWC ($W(100) = 0.98381, p = 0.260$) and control group ($W(100) = 0.97864, p = 0.104$), nPVI-V for PWC ($W(100) = 0.97755, p = 0.085$) and control group ($W(100) = 0.99257, p = 0.861$), and nPVI-CV for PWC ($W(100) = 0.99128, p = 0.767$) and control group ($W(100) = 0.98204, p = 0.191$). In these cases, we used two-sample t-tests as the variance of the two populations were equal based on F-test with a significance level of 0.05. We also calculated r^2 for evaluating the effect size (Rosenthal 1996, Rosenthal–Rosnow 2008), where 0.01 is considered a small effect size, 0.10 represents a moderate effect size, 0.25 indicates large effect size and 0.50 means a very large effect size. When the speech rhythm measures were not found to be normally distributed, we utilized nonparametric tests, namely Mann-Whitney-Wilcoxon tests to reveal the articulation rate and rhythm measure characteristics of cluttered and non-cluttered speech. When the speech rhythm measures were not found to be normally distributed, the effect size r was calculated using the 'rcompanion' package version 2.3.7 in R (Mangiafico 2016). r^2 between 0.01 and 0.09 indicates a small effect size, r^2 between 0.10 and 0.24 a medium effect size, r^2 between 0.25 and 0.49 a large, and $r^2 \geq 0.50$ can be considered a very large effect size.

3. Results

The average articulation rate was 7.94 syll/s (SD: 1.05) in cluttering and 6.25 syll/s (SD: 0.80) in control speech. The articulation rate was significantly faster in cluttering than in control speech ($W(198) = 9244, p < 0.001$) (Figure 1). The effect size was found to be very large ($r^2 = 0.510$).

Insert Figure 1 about here

The speech rhythm values referring to consonants were analyzed (Table 1). Data showed that standard deviation of consonantal interval duration (ΔC) was higher in control speech than in cluttering. There were significant differences between the two speech types ($W(198) = 3291, p < 0.001$). However, only a small effect size was found ($r^2 = 0.088$). This difference between cluttering and control speech remains even when the durational variability of consecutive consonantal intervals (rPVI-C) is compared between cluttering and control speech ($W(198) = 3406, p < 0.001$), with a small effect size ($r^2 = 0.076$).

Since the articulation rate of cluttered speech differed from that of non-cluttered speech, and as ΔC and rPVI-C are dependent on articulation rate (cf. Dellwo 2010), an analysis of normalized values was required. So, the metrics VarcoC and nPVI-C, which are independent of the articulation rate and only typical for rhythm, were also compared. The value of VarcoC was significantly higher in cluttering than in control speech ($W(198) = 5886, p = 0.030$), with a small effect size ($r^2 = 0.023$). However, there was no significant difference between the types of speech in nPVI-C.

Insert Table 1 about here

Next, the data of vocalic intervals were analyzed (Table 2). There was no significant difference between the types of speech in ΔV and nPVI-V. However, there were significant differences between the two groups of speech in VarcoV, which is independent of the articulation rate and typical for rhythm ($W(198) = 5953, p = 0.020$). In this case the effect size was also small ($r^2 = 0.027$).

Insert Table 2 about here

Finally, normalized data of the durations of consecutive CV-sequences (independent of the articulation rate) were examined (nPVI-CV). Results show that there are significant differences between control speech and cluttering ($t(198) = -2.165, p = 0.032$), with a small effect size $r^2 = 0.023$). This value is significantly higher in control speech (Table 3). There was no significant difference between the types of speech in %V.

Insert Table 3 about here

It can be seen from the data that a slightly different speech rhythm characterizes PWC than control speakers as shown in a traditional speech-rhythm graph (Figure 2). Independent of the articulation rate, there are differences in speech rhythm between cluttered and control speech.

Insert Figure 2 about here

4. Discussion

This study analyzed the speech rhythm of cluttering with an objective acoustic method for the first time, to the best of our knowledge. Spontaneous speech of PWC was compared to that of the control group. Our results show that the speech types slightly differed in their speech rhythm independently of the articulation rate. In what follows, we discuss the results obtained for the articulation rate-independent metrics, and briefly address the connections between the articulation rate and speech rhythm metrics.

The speech of PWC and the control group had speech rhythm differences with a small effect size in terms of VarcoC, a measure independent of the articulation rate. The relative duration of consonantal intervals normalized with the articulation rate has larger variability in cluttered speech than in control speech. The consonantal intervals consist of either a single consonant or consonant clusters. The duration of these units show higher variance in cluttering. PWC typically delete consonants in their natural speech (van Zaalen-op't Hof et al. 2009), therefore the consonant clusters might be shorter which can also contribute to differences in speech rhythm.

Besides consonantal intervals, the vocalic intervals also exhibited durational differences between the speech types. The values of VarcoV in cluttering slightly differed from those in the control speech. This indicates that a larger variability of vocalic interval durations is also a characteristic feature of cluttering. Based on the literature on cluttering, we have no reason to assume that the effect of stress position on the duration of segments become more pronounced in cluttering compared to control speech (as the differences in this measure are traditionally interpreted when comparing speech in different languages (White–Mattys–Wiget 2012)). It is more plausible to assume that it is the increased level of irregularity, i.e. the more frequent occurrence of extreme vowel durations that differentiates cluttering from control speech in this respect. Consonantal and vocalic intervals deviate in a similar manner from control speech in both cases, therefore, it may

not be surprising that the percentage of vowel (or consonant) interval durations relative to the total duration (i.e. %V) remains the same, regardless of speech type.

The PVI (pairwise variability index) values measure the temporal variability of successive intervals of a given type. Their articulation rate-normalized versions nPVI-V and nPVI-C based on vocalic and consonantal intervals, respectively, did not show significant differences between the speech types investigated. This suggests that the larger variability of interval duration in cluttering does not necessarily appear successively. Nevertheless, nPVI-CV, the variability index based on successive intervals (consonantal intervals are followed by vocalic intervals and vice versa, hence the notation ‘CV’) is significantly higher in control speech compared to cluttering. However, the effect size was found to be small.

For a more comprehensive view of the timing patterns in cluttering, we investigated the articulation rate and its relationship with speech rhythm. The rate measurements showed that PWC spoke faster than control speakers. Thus, the cluttering was typically characterized by a high articulation rate in line with previous studies (e.g. Bóna 2012). It should also be noted that some of the investigated speech rhythm values are not independent of the articulation rate. The result of ΔC and rPVI-C typically exhibiting higher values in control speech and lower ones in cluttering is mainly a consequence of the fact that lower speech rate yields larger interval durations, hence their absolute differences and variances will also be larger in magnitude. Fast tempo allows a lower variation of consonantal intervals, whereas lower tempo facilitates a degree of differentiation in the durations of consonants. Thus, in these cases, the articulation rate had masked the real differences of speech rhythm. It might be the tempo that affect these values dominantly rather than the internal variability.

To summarize, cluttering deviates significantly from control speech in terms of various metrics. At the beginning of the analysis, two main hypotheses were formulated. Based on our findings, we could not reject our first hypothesis according to which the rhythm of cluttering was somewhat different from that of control speech, despite the fact that the analyzed values scattered in largely overlapping intervals in the two groups. According to the second hypothesis, differences occur in the duration of both vocalic and consonantal intervals between cluttered and control speech. The results were consistent with this hypothesis: we could indeed identify certain values, based on either consonantal or vocalic interval durations that exhibited differences between the two speech types. Since several of the parameters measured are independent of the articulation rate, we can conclude that regardless of the apparent difference in tempo, there is also a detectable deviation in the rhythmic patterns of the two groups. However, it is important in terms of the magnitude of the measured differences, i.e. the calculated effect size was found to be large for the articulation rate and rather miniscule for speech rhythm metrics. These findings are consistent with perceptual experiments of timing and speech rhythm (White–Mattys–Wiget 2012). They found that listeners primarily rely on differences in the articulation rate as a cue for identifying languages and dialects where applicable. Yet, when the rate was set to be the same for the perceived speech samples, the listeners were also able to exploit speech rhythm as a useful cue. Even in the absence of perceptual timing experiments on cluttering, it is generally possible to determine a so-called “just noticeable difference”, a detectability threshold in the perception of the articulation rate, reported to be 5% (Quené, 2007). Higher articulation rate is a clearly detectable signature of cluttering, therefore, differences in speech rhythm are supposedly of smaller relevance for listeners.

It must be admitted that this analysis has certain limitations. On the one hand, the number of participants could have been higher. Their relatively low number can be justified by the fact that

pure cluttering is quite rare (St. Louis et al. 2003). However, literature on speech rhythm (Arvaniti 2012) shows that having 10 participants in one group as speakers is enough to make statistically significant comparisons. Another limitation is that the recordings in a soundproof room are not natural situations for PWC, which again might modify their speech production. Although according to the literature (van Zaalen-op't Hof et al. 2009), if they pay special attention, the quality of their speech might improve. For this reason, in real situations they would probably speak faster and with less typical speech rhythm. In addition, we investigated one dimension of the speech rhythm (surface timing pattern) while other acoustic features (e.g. intensity and fundamental frequency) may contribute to the perception of speech rhythm. Finally, it should be noted that the analysis was performed in a language traditionally considered syllable-timed (Kohári 2018). It would be worth examining speech rhythm values in PWC speaking a stress-timed language.

The results have several implications for clinical practices. Since high speech and articulation rate are the most salient features of cluttering, they need to be in the focus of therapy. If PWC are able to decelerate their articulation rate and, as a result, their articulation becomes more accurate, it may have a positive effect on their perceived speech rhythm. In slow speech, realizations of speech sounds are more likely to achieve the target configuration, and fewer omissions and deletions of speech sounds might occur. Hence, from a therapeutic point of view, slowing down the tempo should be the primary task.

5. Conclusion

The results of this study provide us with numerical data about speech rhythm in cluttering for the first time. They objectively confirm the subjective impressions, namely that one of the main

symptoms of cluttering is the increased articulation rate, whereas differences in speech rhythm between the two speech types are less detectable.

Declaration of interest

There is no conflict to report.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors wish to thank Zsófia Koren-Dienes, Benedek Gergő Vincze, Miklós Vincze, and Monika Ford for their help in preparing this paper.

References

- Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351-373.
- Bakker, K., Myers, F. L., Raphael, L. J., & St. Louis, K. O. (2011). A preliminary comparison of speech rate, self-evaluation, and disfluency of people who speak exceptionally fast, clutter, or speak normally. In D. Ward, & K. S. Scott (Eds.), *Cluttering. A Handbook of Research, Intervention and Education* (pp. 45–65). Hove and New York: Psychology Press.
- Bloch, B. (1950). Studies in colloquial Japanese IV phonemics. *Language*, 26(1), 86-125.
- Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.1). http://www.fon.hum.uva.nl/praat/download_win.html
- Bóna, J. (2012). Linguistic-phonetic characteristics of cluttering across different speaking styles: a pilot study from Hungarian. *Poznań Studies in Contemporary Linguistics PSiCL*, 48, 203-222.
- Bóna, J. (2016). Characteristics of pausing in normal, fast and cluttered speech. *Clinical Linguistics & Phonetics*, 30(11), 888-898.
- Cumming, R. E. (2011). Perceptually informed quantification of speech rhythm in pairwise variability indices. *Phonetica*, 68(4), 256-277.
- Daly, D. A. (2006). *Predictive Cluttering Inventory (PCI)*. Retrieved on 5 March 2010 from <http://associations.missouristate.edu/ICA/>
- Dauer, R. M. (1983). Stress-timing and syllable-timing re-analysed. *Journal of Phonetics*, 11, 51-62.

- Dauer, R. M. (1987, August). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th international congress of phonetic sciences* (Vol. 5, pp. 447-450).
- Dellwo, V. (2009). Choosing the right rate normalization methods for measurements of speech rhythm. In *Proceedings of AISV*, 13–32.
- Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence. Dissertation.* Bonn.
- Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds.): *The handbook of phonetic sciences*. Second edition. Blackwell Publishing. 521-602.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, 7(515-546).
- Hartinger, M., & Mooshammer, C. (2008). Articulatory variability in cluttering. *Folia Phoniatica et Logopaedica*, 60(2), 64-72.
- Hoequist Jr, C. (1983). Syllable duration in stress-, syllable-and mora-timed languages. *Phonetica*, 40(3), 203-237.
- Kohári, A. (2018). *Időzítési mintázatok a magyar beszédben*. [Temporal patterns in Hungarian speech.] ELTE Eötvös Kiadó, Budapest.
- Ladefoged, Peter (1975). *A Course in Phonetics*. Harcourt Brace Jovanovich, New York.
- Lanouette, E. B. (2011). Intervention strategies for cluttering disorders. In D. Ward, & K. S. Scott (Eds.), *Cluttering. A Handbook of Research, Intervention and Education* (pp. 175–197). Hove and New York: Psychology Press.

- Logan, K. (2010). *Helping Clients Who Clutter Regulate Speaking Rate*. Paper presented on the First International Cluttering Online Conference, <http://www.mnsu.edu/comdis/ica1/papers/loganc.html> (06.08.2014.)
- Mangiafico, S.S. (2016). Summary and Analysis of Extension Program Evaluation in R, version 1.18.1, rcompanion.org/handbook/ (15.11.2019.)
- Myers, F. L., & St. Louis, K. O. (2007). *Cluttering*. (DVD) Memphis, TN: Stuttering Foundation.
- Nolan, F., & Jeon, H. S. (2014). Speech rhythm: a metaphor?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130396.
- Oliveira, C. M. C. D., Broglio, G. A. F., Bernardes, A. P. L., & Capellini, S. A. (2013). Relationship between speech rate and speech disruption in cluttering. In *CoDAS* (Vol. 25, No. 1, pp. 59–63). Sociedade Brasileira de Fonoaudiologia.
- Pike, Kenneth (1945). *The intonation of American English*. University of Michigan Press, Ann-Arbor.
- Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 35(3), 353-362.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. *Linguistic controversies*, 73, 79.

- Rosenthal, R., & Rosnow RL (2008). *Essential of Behavioral Research: Methods and Data Analysis* (3rd ed.) McGraw Hill, Boston.
- Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford University Press.
- St. Louis, K. O. (1992). On defining cluttering. In F. L. Myers, & K. O. St. Louis (Eds.), *Cluttering: A clinical perspective* (pp. 37–53). Kibworth, Great Britain: Far Communications, Ltd.
- St. Louis, K. O. (1996). A tabular summary of cluttering subjects in the special edition. *Journal of Fluency Disorders*, 21(3-4), 337-343.
- St. Louis, K. O., Raphael, L. J., Myers, F. L., & Bakker, K. (2003). Cluttering updated. *The ASHA Leader*, 8(21), 4-22.
- St. Louis, K. O., Myers, F. L., Bakker, K., & Raphael, L. J. (2007). Understanding and treating cluttering. In E. G. Conture, & R. Curlee (Eds.), *Stuttering and Related Disorders of Fluency, Third Edition* (pp. 297–325). New York: Thieme.
- St. Louis, Kenneth O., & Schulte, Katrin (2011). Defining cluttering: the lowest common denominator. In Ward, David – Scott, Kathleen Scaler (eds.): *Cluttering. A Handbook of Research, Intervention and Education*. Psychology Press, Hove and New York. 233–253.
- St. Louis, K. O. & Rustin, L. Professional awareness of cluttering. In F. L. Myers, & K. O. St. Louis, (Eds.) (1992). *Cluttering: A Clinical Perspective*. Leicester: FAR Communications, 1992. Reissues: San Diego, CA: Singular. Chapter 2, pp. 23-35.
- Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*, 124(2), EL34-EL39.

Van Zaalen-op't Hof, Y., Wijnen, F., & De Jonckere, P. H. (2009). Differential diagnostic characteristics between cluttering and stuttering—Part one. *Journal of fluency disorders*, 34(3), 137–154.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, 66(4), 665-679.

Bio-note

Judit Bóna, Ph.D., is a phonetician. She earned her Ph.D. in 2007 from ELTE Eötvös Loránd University. She is an associate professor at the Applied Linguistics and Phonetics Department of ELTE Eötvös Loránd University in Budapest (Hungary). One of her main research areas is cluttering. She has published several papers on fluency disorders.

Anna Kohári, PhD is a junior research fellow of the Research Institute for Linguistics. She earned her Phd at ELTE Eötvös Loránd University in 2017. Her research interest is phonetics and applied linguistics. More specifically, her work examines speech rhythm and timing, particularly in conversation.

Figure 1: Articulation rate (median and standard deviation) in the two types of speech

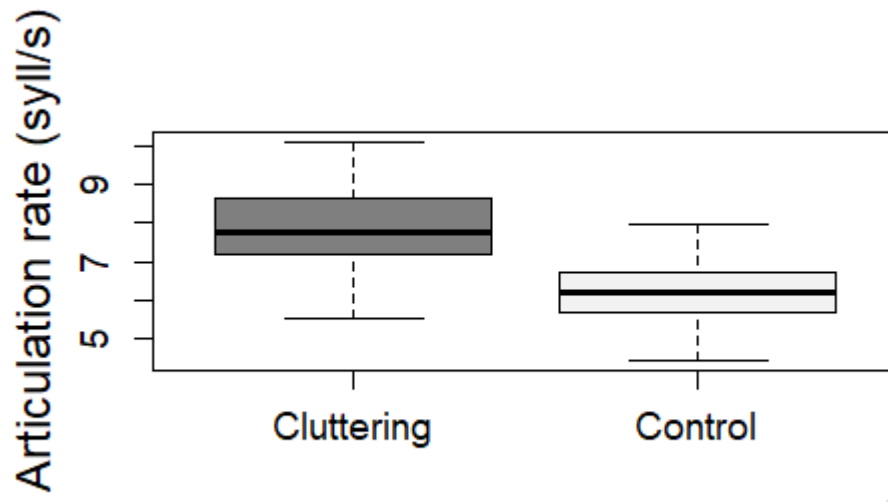


Figure 2: A traditional speech-rhythm graph (Co: control speech, Cl: cluttered speech)

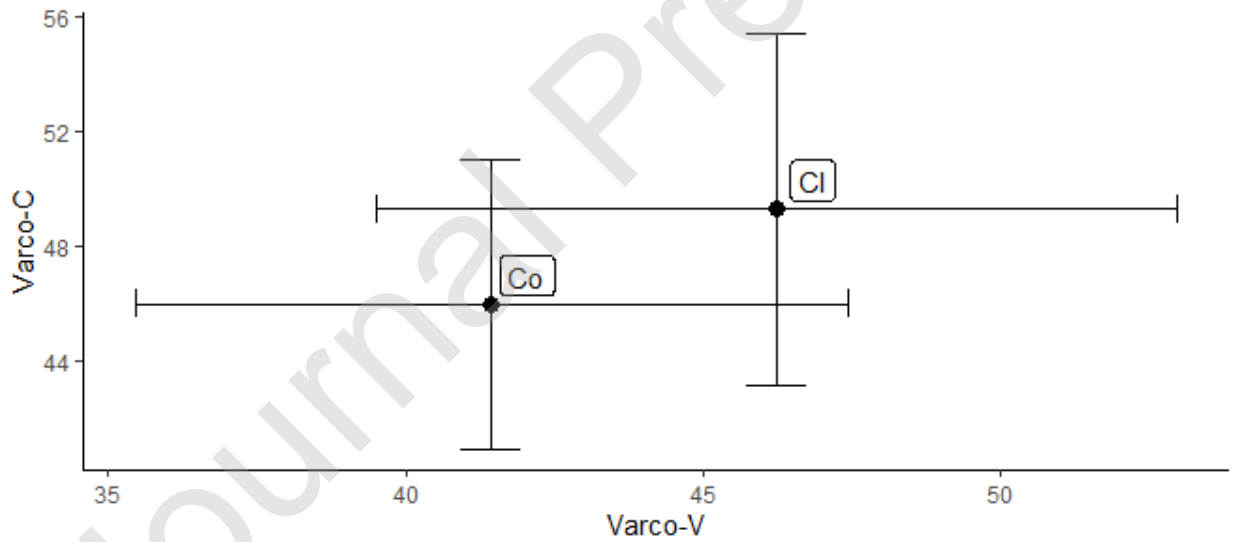


Table 1: Data referring to consonants in cluttering and control speech (mean and standard deviation)

	Cluttering N = 100	Control N = 100
ΔC	36.1 (11.6)	41.7 (11.9)
VarcoC	49.3 (12.3)	46.0 (10.1)
rPVI-C	40.6 (13.2)	47.3 (12.9)
nPVI-C	54.1 (12.7)	53.8 (12.6)

Table 2: Data referring to vowels in cluttering control speech (mean and standard deviation)

	Cluttering N = 100	Control N = 100
ΔV	28.5 (10.9)	30.8 (11.9)
VarcoV	46.2 (13.5)	41.4 (12.0)
nPVI-V	41.1 (10.7)	40.3 (11.1)

Table 3: %V and nPVI-CV in cluttering and control speech (mean and standard deviation)

	Cluttering N = 100	Control N = 100
%V	45.1 (4.3)	44.2 (4.5)
nPVI-CV	46.5 (7.2)	48.8 (7.9)