2022 ACTA UNIVERSITATIS CAROLINAE PAG. 65–81
PHILOLOGICA 1 / PHONETICA PRAGENSIA

VARIATION IN SPEECH TEMPO AND ITS RELATIONSHIP TO PROSODIC BOUNDARY OCCURRENCE IN TWO SPEECH GENRES

JAN VOLÍN

ABSTRACT
The present study focuses on two problems connected with speech tempo. First, earlier research has been prevalently concerned with central tendencies while variation was mostly perceived as an auxiliary result. We believe, however, that information about data dispersion is essential for proper modelling and experiment design in the field of temporal structure of speech. Therefore, the present study provides reference values for some of the tempo metrics of variation that pertain to (a) between-genre differences, (b) within-genre differences, (c) inter-speaker differences, and (d) intra-speaker differences. Second, we tested the claim that faster tempi lead to fewer prosodic breaks in spoken texts. This claim had been supported by studies where a respondent was asked to produce the same text at various rates. We, on the other hand, pose a question of the number of prosodic breaks in speakers who are fast or slow inherently. The material used in the study represents two genres: poetry reciting and news reading, and we obtained recordings from 24 speakers in each genre. Apart from providing the quantifications, the outcomes suggest, for example, that the predisposition of individual speakers to produce fast or slow tempi differs between the two genres. The fastest speakers in news reading were not necessarily the fastest in poetry reciting. This result points at specific behaviour in different situations and invites caution concerning the idea of hard-wired speaking stereotypes in individuals. Also, the correlation between speakers' rates and the number of phrases they produced was significant only in news reading, not in poetry reciting. This result was corroborated by greater variation in prosodic boundary placement in news reading. In addition, the results offer an insight into the relationship between articulation rate and speech rate, together with the comparison of measurements in syllables per second and phones per second. The latter can be of interest since Czech (the language of the material) belongs to languages with a complex syllabic structure.

Key words: articulation rate, news reading, poetry reciting, prosodic boundary, speech rate

1. Introduction

Research in speech tempo or durations of speech sounds has provided a rich pool of results during its relatively long tradition. Besides sheer scientific curiosity, the motiva-

https://doi.org/10.14712/24646830.2022.29

tion for past studies varied between, for instance, synthesis-by-rule concerns (O'Shaughnessy, 1984; Carlson & Granström, 1986; Campbell, 1992), forensic use (Johnson et al., 1984; Künzel, 1997; Jessen, 2007), or automated competence assessment (Lennon, 1990; Cucchiarini, Strik & Boves, 1997; Graham & Nolan, 2019). It has been clearly established, however, that despite certain universal tendencies, the temporal structure of each language has to be studied on its own (e.g., Barik, 1977; Grosjean, 1980; Trouvain & Möbius, 2014).

Temporal patterns in the Czech language were repeatedly examined in the past and individual studies offered quite significant insights, although from today's perspective, researchers usually worked with smaller samples of speakers or with stylistically limited material. Moreover, some of the studies were published in sources that are currently difficult to access. A thorough dedicated study dealing with Czech is still missing. An outstanding exception is the monograph by Dankovičová (2001) which comprises several meticulous studies and offers valuable quantitative descriptions.

The conceptual fixation of linguists on lexical contrast sometimes leads to small appreciation of the temporal dimension in the prosodic structure of languages. Occasionally, it is even viewed as some sort of an insubstantial variable. Port (1979: 46) uses a strikingly harsh phrase: "phonologically irrelevant factors such as speaking tempo" (sic!), but this is probably a reflection of the widely held view at that time that phonology is solely concerned with segmental phonemes. We, on the other hand, argue that if tempo is systematically used in conveying any component of the communicated meaning, then it must have its own phonology.

One of the reasons for underestimating the functions of tempo in speech is probably methodological: the research is relatively poorly equipped. Current analytical tools do not generate temporal tracks as readily as amplitude or F0 tracks. (Although for a simple but relatively crude method see Volín, 2009). An implicitly connected problem is the belief in the existence of the so-called 'personal tempo'. Palková (1994: 317) defines it as a mean speech-production rate typical of an individual speaker. Informal observations, indeed, lead to perceiving certain speakers as slow, while others as moderate or fast. This idea, again, has its roots in averaging across large speech materials and in disregard for local contextual variation.

We dare to assume that rather than a personal 'signature tempo', individuals display specific strategies when accelerating or decelerating their speech for specific communicative purposes. This was suggested, for instance, for English (Goldman-Eisler, 1961), for French (Fougeron & Jun, 1998), for German (Trouvain & Grice, 1999) or for Greek (Fourakis, 1986). All of these studies, however, follow the common experimental paradigm: various speech tempi are elicited on request. An individual speaker is asked to establish his/her 'normal' rate and relative to that produce a fast/slow or a very fast/very slow version of the same text. Therefore, the speakers' judgements put the productions into classes of rates, but their ideas of what is very fast or very slow might be quite disparate. Nevertheless, the change in an individual behaviour when switching between tempi provides important information about the production of various speech units.

One of the more recent examples of the above-presented paradigm is the study by Werner and colleagues who focused on silent pauses and their association with various tempi produced by a speaker (Werner et al., 2022). The relevance of this study to our

present goals is in that besides others, the authors also used recordings of Czech speakers. However, the authors were interested solely in silent pause modelling and they did not provide any exact quantification of the rates in their material.

In contrast with that, our present study targets two areas of interest: (1) providing exact variation values based on a larger sample of speakers, and (2) correlating the occurrence of prosodic boundaries for fast or slow speakers in their own comfortable modes. The latter means that our speakers did not modify their tempi upon request. Instead, as a group, they created a continuum from slow to fast through their unconscious planning of 'adequate' rate for the given genre. Two speech genres were examined (see below). With regard to variation, we aim at (a) between-genre differences, (b) within-genre differences, (c) inter-speaker differences, and (d) intra-speaker differences.

2. Method

2.1 Material

The two genres examined were *poetry reciting* (POR) and *news reading* (NWS). POR was represented by three Czech poems (P1, P2, P3) from the beginning of the 20th century. Each of them comprised 20 verse lines and in agreement with general conventions of that period they were rhyming. In these poems, consecutive pairs of verse lines were analysed as prosodic wholes (referred to as 'speech units' below) since the pairs also formed distinct semantic units. This was especially clear in the poem P2, which was published in two-line stanzas. The other two poems had four-line stanzas, but major punctuation marks were prevalently present at the end of the second and fourth line. There are indications that the speakers produced the poems with the reflection of this fact (whether conscious or unconscious). Each speaker produced 30 such verse pairs (3 × 10) comprising 584 syllables in total. The title and the pause after it were excluded. The titles were read in disparate ways and the first pause was manifestly longer than all other pauses within the text and reflected some sort of preparatory strategy of the speakers rather than the properties of the text. Quite a few poetry readers actually seemed to be 'bracing' themselves for the 'real' beginning after the title.

The genre of *news reading* (NWS) was represented by four paragraphs (news items) of a realistic news bulletin (NI1, NI2, NI3, NI4). The actual text originally comprised six paragraphs plus some introductory and concluding phrases, but these phrases together with the first and the last paragraphs were excluded from analyses in order to balance the extent of the material used. Even despite this measure, the NWS text still consisted of 700 syllables. In parallel to verse pairs in POR, the NWS was analysed in sentences. Each speaker produced 19 of those in the four paragraphs analysed. Given the disparate structuring of the POR and NWS material, the mean length of a verse pair in our material was 19.2 syllables while that of a sentence was 36.8 syllables.

All recordings were processed identically. Forced alignment for words and phones was performed with Prague Labeller (Pollák, Volín & Skarnitzl, 2007), manual corrections and further labelling were carried out in Praat (Boersma & Weenink, 2019). The data were extracted with dedicated Praat scripts.

Individual poems and news bulletin paragraphs will be referred to as *genre units*. These should not be confused with *speech units*, i.e., verse pairs in poems and sentences in news.

2.2 Speakers

There were 24 speakers (12 female + 12 male). All speakers were current or former university students majoring in philological programmes. Their mother tongue was Czech and their ages ranged from 20 to 32 years. They volunteered after they had read an advertisement calling for people with inclination to poetry and without speech disorders or hearing problems. Financial remuneration was offered. The recording procedure was almost the same for POR and NWS material (a single exception is described below). Speakers were given individual poems or news paragraphs (= genre units) on separate sheets of paper, and were asked to get familiar with the contents and form of each of them. They were allowed to practice individual parts of the texts for as long as they needed. Then they were asked to read out the poem or paragraph as if talking to audiences. To alleviate the situational stress, the speakers were reassured that any mistakes would be edited out and their performance would be strictly anonymous. They were also invited to self-correct, i.e., to read out any speech unit again if they were not satisfied with the outcome. All recordings were made in the sound-treated studio of the Institute of Phonetics in Prague. The only difference in the procedure was the fixed order of paragraphs in NWS (according to the original news bulletin) and random order with fillers in the case of poems in POR.

2.3 Measurements

There is an array of descriptive statistics that reflect central tendencies and variation in a data set. However, certain considerations limit their use in given cases. The current study deals predominantly with rates, hence, harmonic mean had to be used when averaging tempi across several units that belong together. Arithmetic mean, on the other hand, was used when tempo of a unit was its descriptor and variation among units needed to be captured. With regard to metrics of variation, we argue that given our current goals the most beneficial ones are the variation range and variation coefficient. Variation range (Rg_{var}) is the distance between the lowest (minimum) and the highest (maximum) value in the set. In literature, it is often presented just by these edge values, but we find it convenient to report the distance itself as well.

Variation coefficient (C_{var}) is the ratio between the arithmetic mean and the standard deviation from that mean expressed as a percentage. Unlike variation range above it does not depend on two values only, it is calculated with all the data points in a set. As a rule of thumb, coefficients below 30% are considered to represent concentrated data, while coefficients over 50% reflect high dispersion in the data (e.g., Skalská, 1992: 12).

The current presentation practice favours measurements both in syllables per second (syll/s) and phones per second (pho/s). Since the relationship between the two is not straightforward in languages with complex syllabic structures (Pfitzinger, 1998; Koreman, 2006), we will report both rate units.

Outcomes of statistical significance tests concerning differences found will be considered significant at the level of $\alpha = 0.05$, and so will the correlation coefficients. However,

approximate values of *p* will be provided, as customary in current empirical research reporting.

2.4 A terminological note

The term *speech tempo* will be used as a general term (hyperonym) covering other, more specific metrics. The plain term *tempo* will also refer to speech tempo in the present text. *Articulation rate* (AR) is conventionally calculated as number of speech units per unit of time with the exclusion of pauses, i.e., only articulation of lexical items is considered. *Speech rate* (SR), on the other hand, includes pauses into the calculation. It expresses a number of speech units produced per unit of time throughout all speech activity, that is with non-lexical items and pauses included. Logically, for the same stretch of spoken text speech rate cannot be higher than the articulation rate. If there are no pauses and other non-lexical elements, it must be equal, otherwise it is lower.

3. Results

The results concerning speech tempi and their variation will be presented in the following order: (1) the differences between genres, (2) differences among genre units, i.e., within-genre differences, (3) differences between speakers, i.e., inter-speaker differences, and (4) differences among speech units produced by a speaker, i.e., intra-speaker variation. Subsequently, Section 3.5 describes the relationship between the number of prosodic phrases produced and the speakers' tempi.

3.1 Between-genre differences

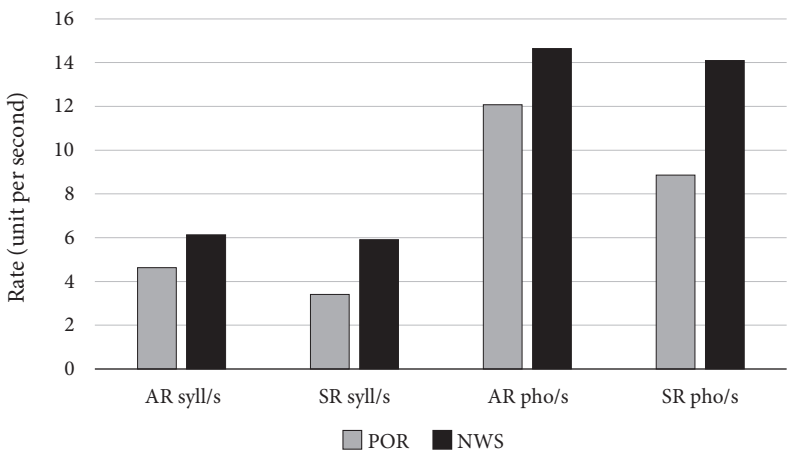

Figure 1. Mean articulation and speech rates (grand means) in two genres: poetry reciting (POR) and news reading (NWS). Values in syllables per second (syll/s) are on the left, phones per second (pho/s) on the right.

Mean articulation and speech rates between the two genres differed: the news reading (NWS) was on average always notably faster than poetry reciting (POR). Articulation rate was faster by 1.5 syll/s or 2.6 pho/s, while speech rate was faster by 2.5 syll/s or 5.24 pho/s. All the differences are displayed in Figure 1. They were tested by ANOVA for repeated measures, which returned highly significant results in all four cases: $F(1, 23) = 287.5$, $p < 0.001$; $F(1, 23) = 181.9$, $p < 0.001$; $F(1, 23) = 461.5$, $p < 0.001$; $F(1, 23) = 355.1$, $p < 0.001$ (arranged left to right after Fig. 1).

As to our key concern, variation, Table 1 summarizes the selected descriptors. It has to be pointed out that one data point in these calculations is a genre unit, i.e., one of the poems or one of the news bulletin paragraphs. The computation is then based on 72 + 96 data points (24 speakers × 3 poem or 4 news items). The coefficient of variation (C_{var}) in articulation rate was below 10%, which signals highly concentrated values. Speech rate C_{var} was somewhat higher but still did not exceed 15%. It is useful to note that while C_{var} in AR is roughly equal in both genres, the poetry reciting is more varied in terms of SR. Obviously, this is caused by unequal pausing strategies of individual speakers.

Interestingly, the variation range (Rg_{var}) exhibits an opposite pattern: the speech rate values are comparable, while articulation rate values are more dissimilar. It has to be pointed out, though, that Rg_{var} depends on two values only, which clearly disregards the situation in the rest of the data set. As a metric, Rg_{var} is often reported as a useful descriptor, but it has to be considered with caution.

Certain insight can be added by inspection of the minima and maxima themselves. There are two facts to be noted. First, it is apparent that the differences between the two genres are slightly greater in maxima than in minima. Second, the fact that NWS is on average faster is not caused solely by the maxima: both the lowest and the highest values are shifted upwards.

Table 1. Variation metrics across poetry reciting (POR) and news reading (NWS) given for articulation rate (AR) and speech rate (SR), both expressed in syllables per second (syll/s) and phones per second (pho/s).

	C_{var} (%)		Rg_{var}		*Max*		*Min*	
	POR	NWS	POR	NWS	POR	NWS	POR	NWS
AR-syll/s	8.1	9.0	1.8	2.5	5.5	7.5	3.8	5.0
SR-syll/s	12.3	9.3	2.1	2.5	4.5	7.3	2.4	4.8
AR-pho/s	7.0	7.9	4.1	4.8	14.2	17.6	10.1	12.8
SR-pho/s	11.3	8.1	5.2	5.1	11.7	17.2	6.5	12.0

3.2 Within-genre variation

Figure 2 shows that the mean tempi of the three POR genre units (i.e., three poems: P1, P2, and P3) were not equal. The strongest effect of Genre unit was returned by a one-way ANOVA for articulation rate in *syllables per second*: $F(2, 69) = 26.19$; $p < 0.001$, with post-hoc Tukey test confirming all three poems significantly different from each other.

The same effect for speech rate in syllables per second was weaker: $F(2, 69) = 13.79$; $p < 0.001$, with post-hoc Tukey test suggesting significant differences between P1 on the one hand, and P2 and P3 on the other. (The significance of the difference between P2 and P3 was no longer present.). The test criterion was slightly smaller when the unit of *phones per second* was used, but the result was still highly significant for articulation rate: $F(2, 69) = 11.62$; $p < 0.001$, with post-hoc inspection identifying P3 significantly different from P1 and P2. Finally, the weakest effect of Genre unit was produced for speech rate in phones per second: $F(2, 69) = 6.93$; $p \approx 0.001$. The post-hoc Tukey test found only the difference between P1 and P3 significant.

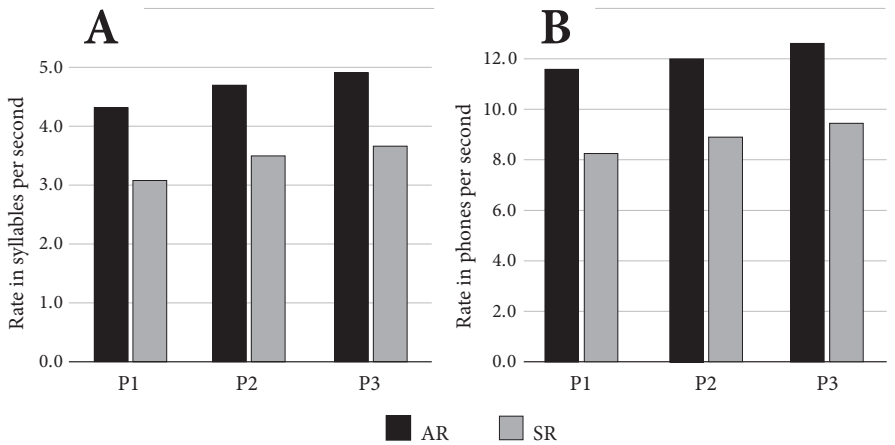

Figure 2. Mean tempi in the three investigated poems (P1, P2, P3). Panel A) captures the values in syllables per second, panel B) in phones per second. Darker columns represent articulation rate (AR), lighter columns pertain to speech rate (SR).

The same analysis was carried out for the NWS genre. Similarly to POR, Fig. 3 indicates that there were differences in the mean tempi of the individual genre units (i.e., the four bulletin paragraphs). It has to pointed out, that while poems were read in a random order with quite a lot of fillers in between, the news were read in a constant order dictated by the original broadcast. Thus, NI1 was always before NI2, etc. The figure shows how the mean tempo decelerates from the first domestic news through the second one and the foreign news down to the sports news with the lowest means.

The strongest effect of Genre unit was returned by a one-way ANOVA for AR in *syllables per second*: $F(3, 92) = 10.61$; $p < 0.001$. This is consistent with the test in POR reported above. The post-hoc Tukey test indicated NI1 significantly different from NI3 and NI4, and NI2 significantly different from NI4. The same effect for SR in syll/s was slightly weaker: $F(3, 92) = 10.06$; $p < 0.001$, but still highly significant. The post-hoc Tukey test suggested significant differences between NI1 and NI2 on the one hand, and NI3 plus NI4 on the other hand. The test criteria were smaller when the unit of pho/s was used, but the results were still significant both for AR and SR: $F(3, 92) = 4.79$; $p < 0.01$, and

$F(3, 92) = 3.74$; $p \approx .014$, respectively. The post-hoc test for the former found NI1 different from all the other NIs, whereas in the latter case significance was reached only for NI1 against NI3 and NI4.

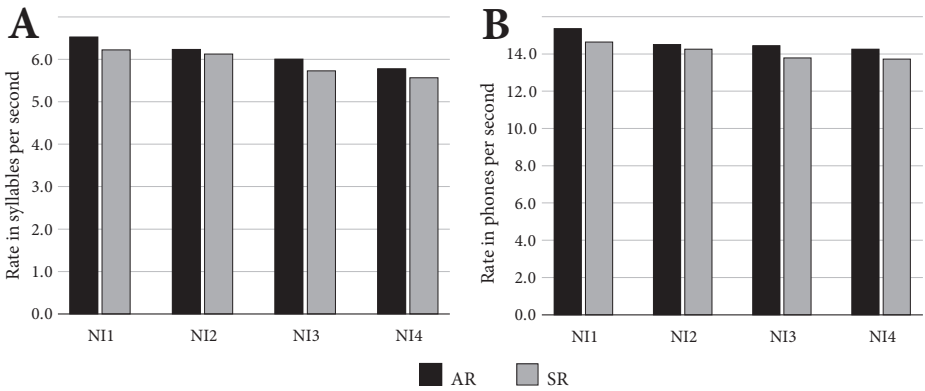

Figure 3. Mean tempi in the four investigated news items (NI1, NI2, NI3, NI4). Panel A captures the values in syllables per second, panel B in phones per second. Darker columns represent articulation rate (AR), the lighter ones represent speech rate (SR).

3.3 Interspeaker variation

The grand means across genres from Section 3.1 need to be broken into contributions by individual speakers. These are captured in Figures 4 and 5. The former displays the POR personal means, the latter the NWS means. The comparison of the figures confirms that the difference between AR and SR is smaller in news reading – a fact already noted in Section 3.1 above. It is also clear at first sight that the values produced by individual speakers are quite evenly distributed. There are no visible categorical breaks. Furthermore, it should be noted that the SR values are not exactly parallel to the AR values. This, again, indicates various pausing strategies among individuals. Also, the ordering individual rates by magnitude leads to roughly the same order in syll/s and pho/s – only small changes are observable.

The opposite is true when POR and NWS orderings are compared. Although in our current sample the slowest reciter is the slowest newsreader as well (speaker F10), the order of the other speakers by their tempi is not the same in POR as in NWS. This suggests that individuals have their specific inner concepts of each of the genres. In fact, only three speakers have the same position in the ordered set of POR and NWS. Seven speakers moved in the ordered data by one or two positions, four speakers moved by three or four positions. The remaining ten speakers moved by 5 or more positions, while four of those by even more than 10 positions.

Another way of looking at the same problem might be computation of Pearson's correlation coefficient between POR and NWS performances. This step returned $r = 0.44$ for

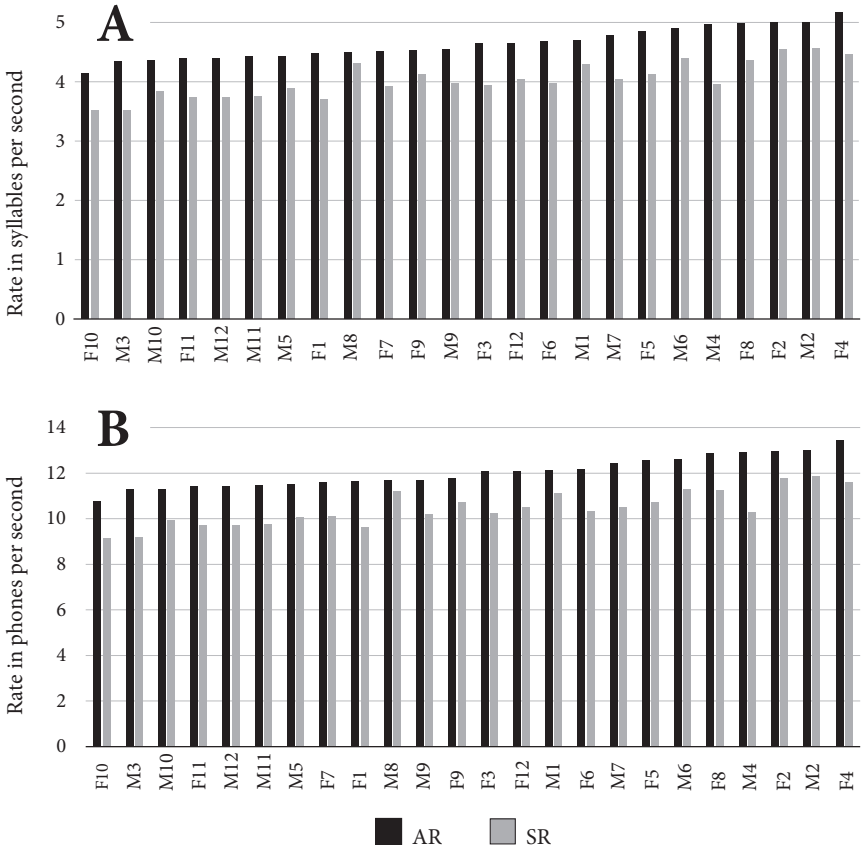

Figure 4. Mean tempi produced by individual speakers in poetry reciting (ordered by the AR values). Panel A) captures the values in syllables per second, panel B) in phones per second. Darker columns represent articulation rate (AR), lighter columns represent speech rate (SR).

both AR and SR in syll/s, and $r = 0.5$ for both AR and SR in pho/s (significant at the level of $\alpha = 0.05$). This suggests only moderate correspondence between the performances of a speaker in the two different speech genres.

Table 2 displays variation metrics across the sample of speakers. Unlike Table 1 above, Table 2 builds on individual people. Thus, for instance, *Min* refers to the slowest speaker, while Rg_{var} refers to the difference between the means of the fastest and slowest speaker under the given measurement condition.

It can be noted that the coefficient of variation (C_{var}) is below 8%, which means very low dispersion of the individual tempi. This is lower than the corresponding values in Table 1. The outcome is not surprising – Table 2 builds on mean tempi of individual speakers, while Table 1 reflects variation in mean tempi of individual genre units (poems or news paragraphs). The same holds for variation range (Rg_{var}): individual speakers differ less than individual genre units. For instance, the slowest and the fastest speakers in

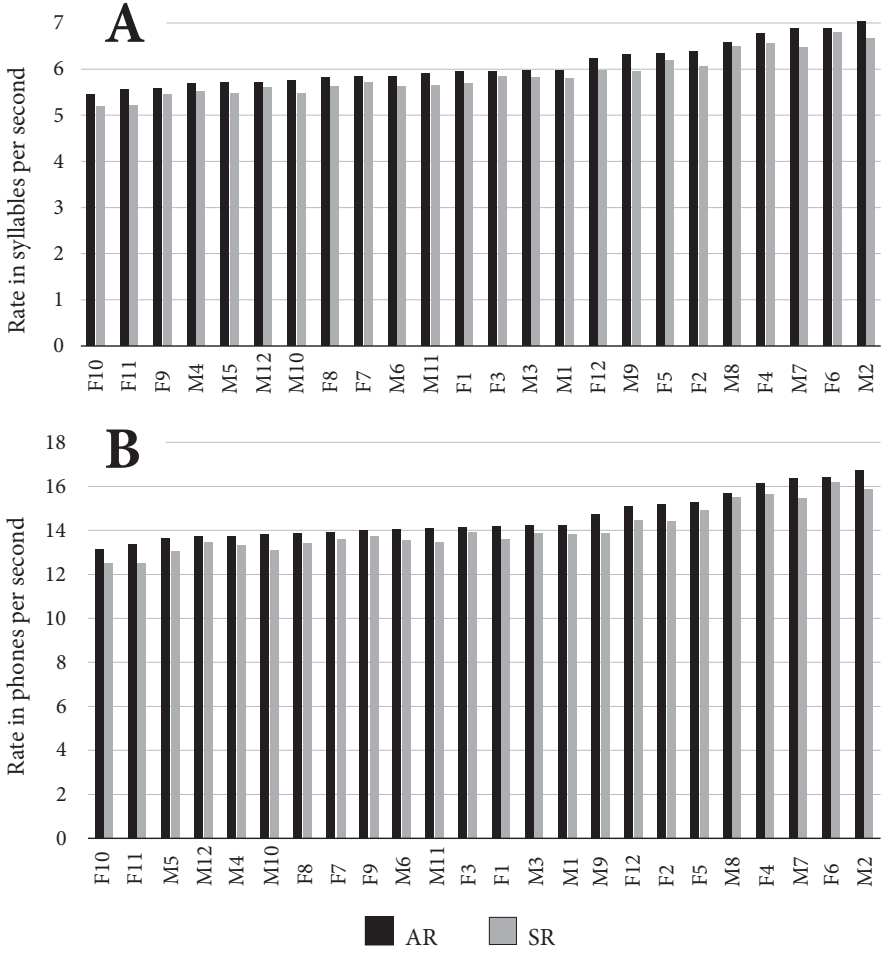

Figure 5. Mean tempi produced by individual speakers in news reading (ordered by the AR values). Panel A) captures the values in syllables per second, panel B) in phones per second. Darker columns represent articulation rate (AR), lighter columns represent speech rate (SR).

Table 2. Variation metrics across speakers in poetry reciting (POR) and news reading (NWS) given in articulation rate (AR) and speech rate (SR), both expressed in syllables per second (syll/s) and phones per second (pho/s).

	C_{var} (%)		Rg_{var}		*Max*		*Min*	
	POR	NWS	POR	NWS	POR	NWS	POR	NWS
AR-syll	5.7	7.5	1.0	1.6	5.1	7.0	4.1	5.4
SR-syll	7.6	7.7	1.1	1.6	4.6	6.8	3.5	5.2
AR-pho	5.7	7.2	2.7	3.7	13.5	16.8	10.8	13.1
SR-pho	7.5	7.4	2.7	3.7	11.9	16.2	9.2	12.5

POR differ by 1 syll/s, given that the fastest reciter spoke at AR of 5.1 syll/s while the slowest spoke at AR of 4.1 syll/s. Similarly, the fastest newsreader produced AR of 7.0 syll/s, while the slowest one 5.4 syll/s – hence the Rg_{var} of 1.6 syll/s.

All the minima in Table 2 (i.e., the slowest individuals) are unsurprisingly higher than the lowest values in Table 1 (i.e., the slowest genre units). It could be expected that, correspondingly, the maxima in Table 2 (i.e., the fastest individuals) would be lower than the maxima in genre units. However, this is only true for NWS and AR in POR. The speech rate in POR marginally diverges from this trend.

3.4 Intraspeaker variation

The variation of tempi produced by a single speaker (the within-speaker variation) can be illustrated by a histogram of values representing his or her speech units. Speaker M12 was identified as a typical individual with modal Rg_{var} since his Rg_{var} lay in the middle of the data set ordered by magnitude. His values are displayed in Figure 6.

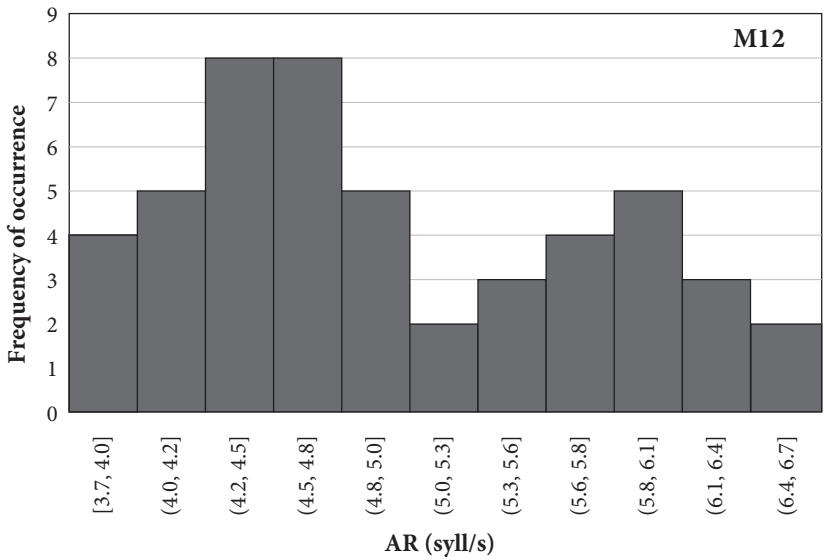

Figure 6. Histogram of AR values in speech units (n = 49) produced by speaker M12 (see text for selection reasons).

The first important fact to note is the bimodality of the histogram. Indeed, the articulation rates of POR were clearly lower than those of NWS, as signalled by highly significant effect of genre (Section 3.1). Thus, when collapsing data from two genres into one set, researchers map certain communicative potentials of a given speaker, but they should not necessarily expect normal (Gaussian) distribution of values within such a combined set.

The second detail to point out is the scale of intraspeaker variation, which is clearly larger than variation among means of individuals (analysed in the previous section). The difference between the slowest and fastest speech unit of this particular speaker was 3 syll/s.

Rather than mean values as in previous sections, we will present a few individual examples at this point to expose intraspeaker variation. (This is because the approach analogous to Sections 3.1 and 3.2 would require 24 tables of the Tab. 1 and Tab. 2 design, which would impair the lucidity of the presentation). The examples in Tables 3 and 4 were selected to represent the most monotonous, the most balanced, and the most varying speaker in each genre.

Table 3. Articulation rate metrics representing intraspeaker variation in three speakers of monotonous, balanced and changeable type in poetry reciting (POR) and news reading (NWS).

Genre	*Speaker*	*Min* (syll/s)	*Max* (syll/s)	C_{var} (%)	Rg_{var} (syll/s)
POR	monotonous	3.87	5.14	6.91	1.27
	balanced	3.68	5.30	9.30	1.62
	varying	3.40	5.69	10.95	2.29
NWS	monotonous	4.59	6.26	8.55	1.67
	balanced	4.21	6.65	9.69	2.44
	varying	5.49	9.87	15.27	4.37

Apart from the fact that all variation parameters are lower in POR than in NWS, it can be observed that the varying speaker in POR not only raises the maximum, but also lowers the minimum. This does not happen in NWS, although it is the same person. We might speculate that temporal strategies of an individual differ across speech genres. As to the other metrics, their values increase from monotonous to varying type. Analogous data for speech rate (SR) are displayed in Table 4.

Table 4. Speech rate metrics representing intraspeaker variation in three speakers of monotonous, balanced and changeable type in poetry reciting (POR) and news reading (NWS).

Genre	*Speaker*	*Min* (syll/s)	*Max* (syll/s)	C_{var} (%)	Rg_{var} (syll/s)
POR	monotonous	3.20	4.65	8.06	1.45
	balanced	3.08	5.21	11.42	2.13
	varying	2.92	5.69	14.82	2.77
NWS	monotonous	4.59	6.26	9.28	1.67
	balanced	4.98	7.65	10.48	2.67
	varying	4.34	9.46	17.60	5.12

Comparison between Tables 3 and 4 reveals that both C_{var} and Rg_{var} are higher in speech rate than in articulation rate. A similar trend was already reported in previous sec-

tions. Greater variation is obviously caused by the use of pauses, which lower the minima more than the maxima. For instance, the slowest speech unit of the monotonous speaker has AR that is 75.3% of her fastest unit. In terms of SR, it is only 68.8%.

For the sake of brevity, we will not report analogous results for measurements in pho/s. They were inspected and established as patterning consistently with the measurements in syll/s displayed in Tables 3 and 4.

A final observation presented in this section concerns an interesting difference in distribution of the variation metrics of C_{var} and Rg_{var}. Figure 7 documents that while the C_{var} values are spread more or less symmetrically and peaking at about the middle, the Rg_{var} values have massively skewed distribution with most data points in the low values and progressively fewer in high values.

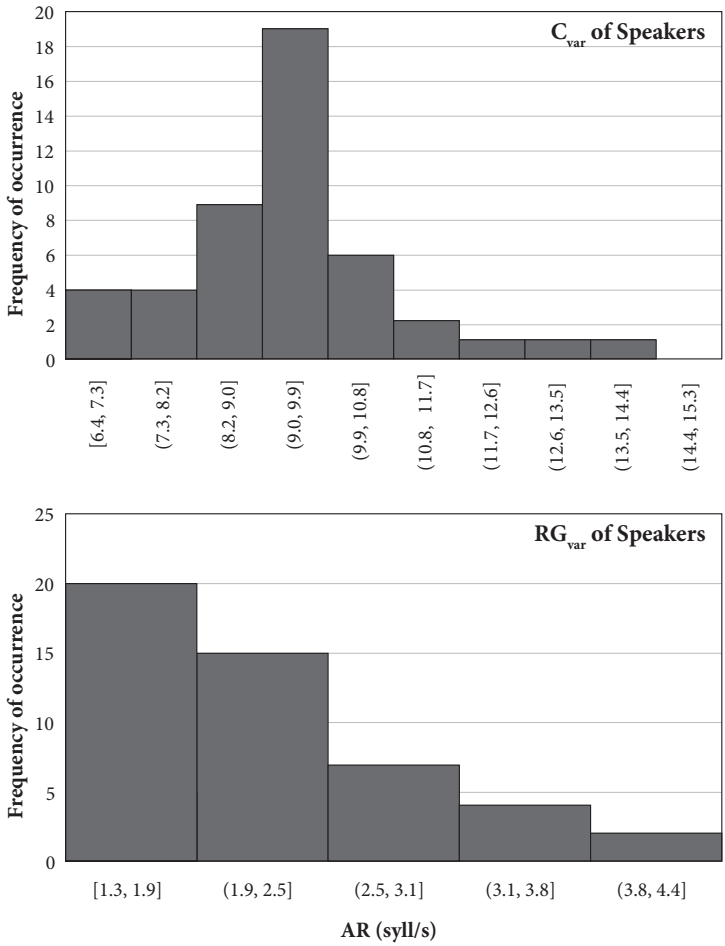

Figure 7. Histograms of within-speaker C_{var} and Rg_{var} values produced in individual performances. Measurement condition: AR in syll/s.

3.5 Division into prosodic phrases

The major question answered in this section concerns the frequency of occurrence of prosodic phrases in relation to AR or SR. Only full prosodic phrases were considered (i.e., intonation phrases in ToBI terminology). On average, the speakers produced 158 prosodic phrases each, of which 96 were in POR and 62 in NWS. The lowest number of prosodic phrases produced by one speaker was 131, while the highest number was 175. These extremes delimit the variation range and they were both produced by male speakers. (However, since male/female opposition was not examined in this study, this fact will not be elaborated on).

The declared focus of the present study is variation. The speakers produced exactly the same texts in two genres, but their production could differ by 44 prosodic boundaries. This span seems impressive, however, in terms of C_{var} it is only 7.6%, which indicates highly concentrated data. An overview for the sample of present genres is provided in Table 5. Interestingly, when the variation metrics are calculated for each genre separately, C_{var} emerges markedly higher for NWS than for POR (Table 5). This suggests that poem structuring guides the speakers more firmly, whereas the news texts provide greater freedom for prosodic boundary placement. Nevertheless, C_{var} of 12.2% still reflects concentrated data.

Table 5. Variation metrics for the number of prosodic phrases in the examined texts in poetry reciting (POR) and news reading (NWS). The metrics Rg_{var}, *Max*, *Min* are given in number of phrases.

	C_{var} (%)	Rg_{var}	*Max*	*Min*
POR	7.1	26	108	82
NSW	12.2	28	75	47
Both	7.6	44	131	175

When correlating speakers' speech rates with the number of prosodic phrases they produced (Pearson's formula), the coefficients were $r = -0.51$ for AR both in syll/s and pho/s, and $r = -0.64$ for SR both in syll/s and pho/s. This result applies to data undifferentiated for genres. When the numbers of prosodic phrases were split by genre, the significant correlation disappeared for POR, but strengthened for NWS, where the correlation coefficients were: $r = -0.58$ for AR in syll/s, $r = -0.67$ for SR in syll/s, $r = -0.57$ for AR in pho/s, and $r = -0.66$ for SR in pho/s.

4. Discussion

The two objectives set for the current study were met: (1) the variation of tempo in two speech genres was quantified, and (2) the relationship between the articulation rate/speech rate on the one hand, and the number of prosodic phrases produced in a text on the other hand, was examined.

As to the latter, our expectations were based on older laboratory experiments where the same speakers were asked to pronounce identical sentences in slow, moderate and fast rates, and their fast speech contained fewer phrases. In our study, we modified the research question and asked whether the speakers who use habitually faster or slower speech tempi would follow such a pattern as well. The results of correlation analyses showed that to some extent they do so. The returned coefficients were, indeed, negative, which means fewer prosodic breaks with faster rates. However, the relationship between the two variables does not seem to be very strong: only about 30% of variance could be explained when all our speech material was combined ($r^2 \approx 0.30$). What is even more interesting, though, is the difference between articulation rate and speech rate. The correlation coefficients were clearly higher for SR, suggesting that there is some systematicity in pausing, and that pure articulation is less flexible. Moreover, the statistical significance of the correlation coefficients was confirmed only for news reading.

This fact supports the increasingly prevalent claims that speech styles and genres matter in phonetic research (Wagner et al., 2015). The two genres examined in the present study differed in other aspects as well. Articulation rate in NWS was by 1.5 syll/s faster than in POR, and in terms of speech rate the difference was even larger: 2.5 syll/s. This implies that pauses in poetry reciting occupy greater space. This fact also caused greater C_{var} in speech rate in POR. On the other hand, with respect to the occurrence of prosodic phrase boundaries, greater variation was ascertained in NWS than in POR (as expressed by C_{var}). This indicates stronger demand on certain prosodic structuring in poetry and greater space to manoeuvre in news reading.

In future, however, not only the number, but also the actual placement of prosodic boundaries should be examined. Clearly, the linguistic specification of the positions with high or low concord among speakers would be of interest.

The analysis of inter-individual differences suggested that the relative tempo in the group is not the same across the two genres. The correlation coefficient between the performances of the speakers in POR and NWS was only moderate (cf. Section 3.3). It follows that individual temporal inclinations should not be over-estimated. Although informal experience points at the existence of habitually slow or fast speakers, generalizations across speaking genres might be injudicious. While a few speakers might not differentiate between the genres by tempo, the majority seem to exhibit specific personal concepts of the genre temporal form.

On the other hand, the differences between speakers within a genre were surprisingly low. The coefficient of variation was below 8% in all four measurement modes. This might suggest that just as we share the lexicon and syntax of a language, we also share the *prosodic grammar* for various communicative purposes.

Unsurprisingly, the within-speaker variation turned out to be greater than variation based on large averaging. There were speakers whose performance could be classified as varied, while others could be labelled as monotonous. The varied performance meant C_{var} up to 17%, whereas the monotonous one would produce C_{var} below 10%. Again, the coefficients of variation in individual speakers were lower for AR than for SR, suggesting that individuals are more stable in their speed of articulation than in pausing. This fact invites a more thorough research into pausing strategies (as pauses were the only difference in calculations of AR and SR). On the whole, however, the results are in line with the

findings of Dankovičová (2001), who focused on changes in AR within prosodic phrases. She reported variation of about 10% and only exceptionally, mainly in phrase final positions, slightly over 15%. Similar results were implied by Goldman-Eisler (1961), even if the methodology does not allow for direct comparison.

Finally, it has to be stressed that the reference values which we have provided in the present study do not speak for the Czech population as a whole. The sample comprised young university-educated and philology-oriented people, who represent a sector of population with high level of literacy and relatively advanced language competences. For future research, expansion to other social groups of Czech-speaking population would be desirable. Likewise, various other speech genres should be mapped and contrasted with the present results. We believe that the topic of tempo variation should be pursued further with the aim to provide a solid basis for 'temporal phonology'.

5. Acknowledgement

The study was carried out with the support of GAČR (Czech Science Foundation), Project 21-14758S. The author also wishes to express his thanks to doc. Radek Skarnitzl for valuable comments on the draft of the study, and to reviewers for their thorough work.

REFERENCES

Barik, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech, 20*, 116–126.

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. Computer programme downloaded at http://www.praat. org/.

Campbell, W.N. (1992). Syllable-based segmental duration. In: G. Bailly, C. Benoit, T. Sawallis (Eds.) *Talking Machines. Theories, Models, and Designs*. Amsterdam: Elsevier Science Publishers. 211–224.

Carlson, R., & Granström, B. (1986). A search for durational rules in a real-speech database, *Phonetica 43*, 140–154.

Cucchiarini, C., Strik, H., Boves, L. (1997). Automatic evaluation of Dutch pronunciation by using speech recognition technology. In: S. Furui, B.H. Juang, W. Chou, (Eds.), *Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding*, Santa Barbara, pp. 622–629.

Dankovičová, J. (2001). *The Linguistic Basis of Articulation Rate Variation in Czech*. (Forum Phoneticum 71). Frankfurt am Main: Hector.

Fougeron, C. & Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics 26*, 45–69.

Fourakis M. (1986). An acoustic study of the effects of tempo and stress on segmental intervals in Modern Greek. *Phonetica, 43(4),* 172–188. https://doi.org/10.1159/000261769

Goldman-Eisler, F. (1961), The significance of changes in the rate of articulation. *Language and Speech, 4*, 171–174

Graham, C., & Nolan, F. (2019). Articulation rate as a metric in spoken language assessment. In: *Proceedings of INTERSPEECH*, Graz: ISCA pp. 3564–3568.

Grosjean, F. (1980). Comparative studies of temporal variables in spoken and sign languages: A short review. In: W. Dechert & M. Raupach (Eds.) *Temporal variables in speech*, pp. 307–312. Berlin: De Gruyter Mouton.

Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science & Justice, 47*, 50–67.

Johnson, C. C., Hollien, H. & Hicks, J. W. (1984). Speaker identification utilizing selected temporal speech features. *Journal of Phonetics, 12*, 319–326.

Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics, 4*, 48–83.
Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America, 119*, 582–596.
Lennon, P. A. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387–417.
O'Shaughnessy, D. (1984). A multispeaker analysis of durations in read French paragraphs. *Journal of the Acoustical Society of America, 76*, 1664–1672.
Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.
Pfitzinger, H.R. (1998). Local speech rate as a combination of syllable and phone rate. In: *Proceedings of ICSLP '98 – Sydney*, vol. 3, pp. 1087–1090.
Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In: *Proceedings of XII*th *Speech and Computer* – SPECOM 2007, pp. 537–541.
Port, R.F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7*, 45–56.
Skalská, H. (1992). *Úvod do biostatistiky*. Hradec Králové: LF UK.
Trouvain, J. & Grice, M. (1999). The effect of tempo on prosodic structure. *Proc. of the 14th International Congress of Phonetic Sciences*. San Francisco: IPA, pp. 1067–1070.
Trouvain, J. & Möbius, B. (2014). Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. *Proceedings of 7th Speech Prosody,* Dublin, pp. 275–279.
Volín, J. (2009). Metric warping in Czech newsreading. In: R. Vích (Ed.) *Speech Processing – 19*th *Czech-German Workshop*, pp. 52–55. Praha: AVČR.
Wagner, P. Trouvain, J. & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics 48*, 1–12.
Werner, R. Trouvain, J., & Möbius, B. (2022). Optionality and variability of speech pauses in read speech across languages and rates. In: S. Frota, M. Cruz, & M. Vigário (Eds.) *Proceedings of 11th Speech Prosody*, Lisbon, pp. 312–316.

Jan Volín
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
E-mail: jan.volin@ff.cuni.cz