A CONTRIBUTION TO THE STUDY OF SPEECH TEMPO AND PAUSE VARIABILITY IN TWO DIFFERENT SPEAKING STYLES

JITKA VEROŇKOVÁ

ABSTRACT
This study analyzes speech tempo and pause variability in two speaking styles: read-aloud news and semi-spontaneous self-introductions. Recordings of 10 non-professional Czech female speakers were examined for speech rate, articulation rate, and pauses. Results show that the read-aloud news were significantly faster, with a higher tempo and lower pause volume. While within-genre variability in the news was observed, the introductions exhibited greater inter-speaker variation, particularly in pausing. A speaker's tempo in one style did not strongly predict their tempo in the other. These findings underscore the role of speaking style in shaping the temporal characteristics of speech.

Keywords: Czech; speech rate; articulation rate; speech tempo; pauses; news reading; speech elocution

1. Introduction

Speech tempo has long been a subject of interest to researchers. The sound features that cause listeners to perceive speakers as fast or slow, along with the factors influencing our perception of speech tempo[1], have been examined. Several factors, such as articulation rate, may have a greater impact; however, they do not act independently. The perception of speech tempo is affected by multiple interacting factors with considerable overlaps (cf. Kohler, 1986; Koreman, 2006; Plug et al., 2022).

Factors that have been examined for their influence on speech tempo production include, for instance, age, gender, and dialect region (e.g., Verhoeven et al., 2004; Quené, 2005; Yuan et al., 2006; Jacewicz & Fox, 2010; Bóna, 2014; Huszár & Krepsz, 2021; Ferguson et al., 2024). The results of research are not always consistent, and it is evident that the factors interact in complex ways.

The type of speech task is another factor that has been monitored. For example, Barik (1977) found that articulation rate, speech rate, and pausing were influenced by the speaker's degree of readiness during a given performance. Along with the role of the speech task or speaking style, inter-speaker and intra-speaker variability (or, conversely,

[1] In this paper, the term *speech tempo* (or simply *tempo*) is used as a general term.

https://doi.org/10.14712/24646830.2025.23

stability) has been monitored (e.g., Mixdorff et al., 2005; Jacewicz & Fox, 2010; Bóna, 2014; Huszár & Krepsz, 2021; Ferguson et al., 2024), usually by comparing read and (semi)spontaneous speech.

From an alternative perspective, Koopmans-van Beinum and van Donzel (1996) proposed speech rate to be one of the two main cues listeners use to differentiate between spontaneous and read speech. Their research on spontaneous speech supported the belief that variation in speech rate is somehow related to the information structure in the discourse.

In the first decade of the 21st century, a couple of contributions to the topic of speech tempo in Czech were published (Dankovičová, 2001; Veroňková-Janíková, 2004; Balkó, 2005). They all used recordings of non-professional speakers in various speech tasks. Dankovičová, who focused on articulation rate, suggests that recurring temporal patterns appear within a prosodic unit (Dankovičová, 2001). The speech material analyzed by Balkó encompassed several speaking styles, and the results demonstrated that both articulation and speech rates are influenced by the type of task and the degree of difficulty in planning the speech; at the same time, individual differences exist among speakers (Balkó, 2005). Similar results were reported by Veroňková (2004) based on six different tasks, including read-aloud vs. semi-spontaneous speech, contrasted by recording environment (i.e., individual recording in a studio vs. a semi-public performance). For example, narrating a fairy tale based on a series of pictures proved to be the slowest task by far. There was a clear tendency for a higher speech rate to be associated with the read texts. However, some speakers decreased their speech rate in the read version of a story compared to the original spoken one. This could be explained by an intentional effort to slow down, as the original tempo of the story was criticized by the audience as being too fast.

The present paper was inspired, among others, by two studies by Volín (2019, 2022), published in this AUC Philologica series, both in terms of their topics and methodology. Volín (2022) examined temporal differences in two genres – news reading and poetry reciting – and provided reference values for some tempo metrics. Two target genres differed in both articulation and speech rates (news reading was faster than poetry); however, the results suggested that articulation rate was more stable in both between-genre and within-speaker comparisons, meaning that pausing was more variable. Temporal characteristics between genres showed mostly individual speaker behaviour, while inter-speaker variability within a genre was low, which, according to Volín, might suggest shared communication concepts.

The objective of the present study is to provide data concerning the variability of speech tempo and pauses across two distinct speaking styles. The perspective is motivated by a long-term ambition to examine the behaviour of speakers in various speech situations and to analyze speech performances as a whole, including the impact on listeners. The speech material for the present study was provided by a sample of non-professional speakers, each of whom produced speech in both targeted speaking styles. The first style is news reading, which exemplifies the so-called 'clear speech'. The second is a semi-public speech performance delivered in front of an audience; in the recordings analyzed, this style closely resembles ordinary conversational settings (for details, see the Method section).

It is precisely the inclusion of a semi-public performance that makes the current study unique. To the author's knowledge, such speech material has not been used in recent studies. Recordings of (semi-)spontaneous speech have been analyzed, but these typically consisted of either speech in the presence of an interviewer or recordings of individuals in either pairs or very limited groups. A side benefit of the current analysis was also checking the technical quality of these semi-public recordings and the possibility of their use for further phonetic experiments.

Given the nature of the phenomenon under investigation, the target speaking styles and genres represent an appropriate choice. Non-professional speakers may use speech tempo to emulate professional news readers, given the general assumption – supported by objective measurements – that news readers use a higher speech tempo (Veroňková & Poukarová, 2017). Speech tempo is also one of the factors addressed by rhetoric and is often the subject of evaluation of a speaker's performance, particularly in relation to pausing. Pauses are a noticeable phenomenon for the listener and a very clear signal of a prosodic unit boundary. This is why the inter-pause stretch was taken as the basic unit for measurement in this phase of research. Although articulation rate exhibits regular patterns within a prosodic unit (for Czech, see Dankovičová, 2001), this study focuses on this larger interval. With regard to speech tempo and pausing, the paper aims to examine between- and within-genre variability and inter-speaker variability.

2. Method

2.1 Material

2.1.1 Speaker and recording selection

The recordings were sourced from an archive of student recordings created during mandatory courses in the phonetics programme at the Faculty of Arts, Charles University. None of the recordings were made specifically for the purposes of the research presented in this paper. The author conducted all recordings.

For this study, recordings of 10 female speakers with Czech as their mother tongue were used. Their ages ranged from 19 to 23 years, and they did not report any neurological, speech, or hearing disorders. Two recordings of two speech genres were used from each speaker: a semi-spontaneous self-introduction and read-aloud news.

The speakers and their recordings were selected from the archive using the following procedure. Students from eight phonetics courses who recorded the same news bulletin text formed the baseline group of speakers. From this group, the following speakers were excluded: a) non-native speakers of Czech, b) men – due to their limited representation in the archive, and c) recordings by females that contained a large number of errors and slips of tongue. These recordings were omitted to ensure comparable content across all speakers, as dysfluent passages would have otherwise been removed during the measurement phase. The author then matched the selected news reading recordings with the self-introduction recordings from the same speakers. A few pairs were discarded due to the low technical quality of the introduction recording. Conversely, the degree of dysfluency in the self-introduction did not play a role in the selection process.

2.1.2 Speech material and recording environment

Self-introduction speech

The recording of the self-introduction took place in groups of approximately 12–15 people at the beginning of the Speech elocution course. The speakers' task was to introduce themselves. They were given a list of points to talk about (name, age, high school, university studies, etc.), but were free to decide whether to cover all of them and how to structure their presentation. The speech was intended to last approximately 1 minute. Speakers were given approximately 5 to 10 minutes to prepare before taking turns individually. They were allowed to make notes during preparation; however, they performed without them. The target speaker delivered the speech using a hand-held microphone in front of the others, who served as the audience. The talk was also recorded on a video camera.[2] The entire session, including the preparation, lasted approximately 60 minutes.

With the exception of two respondents, the speakers gave their self-introduction presentation in the first semester of their university studies, often during their very first class, i.e., in an unfamiliar group. They were informed – without further specification – that the recording would provide material for course work and serve as an opportunity for them to get to know each other. The speeches were affected by varying degrees of nervousness, evident from both the recordings and the speakers' accompanying comments. The teacher (the author) attempted to create a friendly atmosphere to reduce tension and encourage the speakers.

The recordings were obtained in a medium-sized classroom (approx. 30 seats) with soundproofed walls. The hand-held microphones (Sennheiser e840 or Rode NT3) were plugged into a computer sound card and recorded using Praat (Boersma & Weenink, 2016) in WAV format with 22.05-kHz sampling frequency using a 16-bit resolution, or using Audacity (48 kHz, 16-bit). Although some recordings were of slightly lower quality, high-quality recording was not essential for analyzing the temporal phenomena under investigation.

News bulletin

The speakers read the same *bulletin* of six *paragraphs* (approx. 550 words in total), and the entire text was included in the analysis. The paragraphs are not completely balanced in terms of syllable count. The first and last paragraphs contain introductory and concluding phrases, respectively (they were also included in the analyses and referred to as *N_ini* and *N_fin*). The news reading genre itself is represented by four paragraphs (*N1–N4*), which are shortened versions of real news from the national broadcaster, Czech Radio[3]. The bulletin text does not contain any lexical items with unclear pronunciation, such as foreign words or names. Apart from the shortening of individual news items, no other modifications were made to the text.

The bulletin recording was part of a session that students completed (with two exceptions) at the beginning of their second year in the phonetics programme, i.e., a year after the self-introduction recording. The news reading was performed individually in the

[2] It does not apply to two speakers from our list.
[3] Český rozhlas, https://portal.rozhlas.cz.

sound-treated studio of the Institute of Phonetics in Prague. The session typically consisted of two rounds, with respondents successively recording the news and then another read text of a different genre (or vice versa). Speakers were given a paper copy of the target text and time to familiarize themselves with it prior the reading. They were allowed to make notes on the sheet and take it into the studio. No special instructions concerning the reading style were provided.

2.2 Procedure

2.2.1 Data segmentation and annotation

The sound processing was performed in Praat (Boersma & Weenink, 2016). Phone and word boundaries (based on orthographic transcripts from archive) were forced-aligned using the Prague Labeller (Volín et al., 2005; Pollák et al., 2007) and Prak (Hanžl & Hanžlová, 2022, 2023). The boundaries were then manually corrected, with special attention to pauses, following the annotation rules presented in Machač & Skarnitzl (2009). For the purposes of this study, prepausal vowels were labelled with an emphasis on perception rather than formant structure.

Pauses were defined as sections without lexical articulation and could be either silent or filled. The boundaries of pauses were labelled manually based on the author's perception; therefore, no minimal pause duration was determined. The vast majority of boundaries were clearly located. Discretion was required in a limited number of cases, for example, when strong glottalization was present or when a canonical word-initial/ /final sound was preceded/followed by an additional schwa. In such cases, the sound was considered part of the articulation if it appeared to be an integral component of the word. If it gave the impression of a hesitation, it was labelled as part of a pause. This process determined the boundaries for the *inter-pause stretches*.

In the bulletin recording, pauses were mainly filled with breaths. In the self-introduction, pauses also contained noticeable hesitation sounds, especially for some speakers. To check whether this could be a source of variability, hesitation passages within the pauses were also separately labelled in these two performances.

One speaker's introduction was twice as long as the others; therefore, only less than the first half of this speech was used for measurement; the cut was made at the point of semantic completion where the speaker finished a subtopic. In one self-introduction recording, a couple of shorter inter-pause streches that the speaker delivered with laughter were excluded.

Due to the speaker selection criteria, the bulletin recordings did not contain significant dysfluencies requiring exclusion. There were a few minor slips of tongue (9 in total, 0–3 per speaker), which were always located in the middle of an inter-pause stretch and did not disrupt the speech flow[4].

The resulting blocks of speech are referred to as *performances*. To obtain units parallel to the bulletin paragraphs, the self-introduction was divided into four parts –

[4] For example: *V České republice se [ro] loni narodilo nejvíce dětí…* Eng: *The Czech Republic had the highest number of children born last year…*

an initial part, two middle parts, and a final part (*P1–P4*). The self-introductions typically contained four subtopics, which made these divisions relatively easy. Not all self-introduction performances contained final phrase (such as *Thank you for your attention*), so this phrase was not included in the final (P4) paragraph but is referred to as *P5* instead, if present.

2.2.2 Measurements and analysis

To determine the number of syllables and calculate the articulation rate, scripts developed by Oceláková and Bořil were used (Oceláková, n.d.; Bořil & Oceláková, n.d.). The data processing and subsequent analysis were partially performed in R within RStudio environment (R Core Team, 2024). Statistical tests were performed using the online calculator at Statistics Kingdom (2017a, 2017b). One-way ANOVA for repeated measures, the post-hoc Tukey HSD test, and Pearson correlation were used to test significance of the results. Outcomes of the statistical tests were considered significant at an alfa level of 0.05.

Adopting the methodology of Volín (2022), this study examined several descriptors of variation: minimum, maximum, their distance (i.e., variation range; all in syll/s), and the coefficient of variation (Cvar, in %), calculated as the ratio of the standard deviation to the arithmetic mean, multiplied by 100. The interpretation of the Cvar values is based on the thresholds presented in Volín (2022): Cvar < 30% represents concentrated values, while Cvar > 50% represents dispersed values.

Pause volume (in %) and pause duration (in seconds) were also monitored. Duration values were used in both non-normalized and normalized forms. Normalization was performed using the following formula: durnorm = dur * ARindiv / ARspeakers (where ARindiv is the AR of a given speaker in a given performance, and ARspeakers is the mean AR of all speakers in that performance). The same descriptors of variation used for tempo were applied to pauses.

A limit value was determined for assessing changes in temporal course; according to Quené (2007), the just-noticeable difference is approximately 5%, which corresponded to 0.3 syll/s for both SR and AR in our speech material.

The analysis employed two analytical units: *performances* and *paragraphs*, with the latter corresponding to the *genre unit* used by Volín (2022). As the long-term focus of this research is on speakers' presentations as a whole and their impact on listeners, the primary analysis encompassed speech stretches of varying lengths, treating them as integral components of the overall performance. In the next step, the data were also recalculated, first excluding 1-syllable stretches and then excluding 1-syllable stretches together with 1-word stretches.

Two sources of data were used: a) *overall values* (the total duration and corresponding number of syllables for a performance/paragraph), and b) the *articulation rate of single inter-pause stretches*. The ARs of individual stretches were used to calculate a weighted mean for a given unit (weighting was done by duration).

The study focuses on two types of performance – news bulletin reading and self-introduction – each produced by ten speakers. Two speaking styles were examined: read-aloud news and semi-public speech presentation.

3. Results

The results will be presented in the following order: (1) the differences between genres regarding speech tempo and pauses, (2) within-genre differences regarding speech tempo, (3) inter-speaker differences. Self-introduction is further referred to as *introduction* or *intro*.

3.1 Between-genre differences

3.1.1 Rates in performances

The measurements depicted in Figure 1 are based on overall performance values of the bulletin and the introduction.

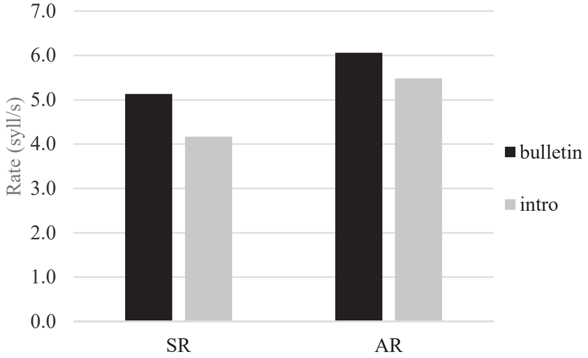

Figure 1 Mean speech and articulation rates in two performances: bulletin (N = 10), and introduction (N = 10). The results are based on overall values.

Although the study focuses on the difference between genres, the relationship between SR and AR within the same genre will be examined first. Given the calculation of SR and AR differs in pauses, which are present in both performances, we assumed that SR and AR would differ. As expected, SR and AR differed within both the bulletin and introduction performances (compare the left and right sides of each pair in Figure 1). A one-way ANOVA for repeated measures returned strong significant results for both performances. Bulletin, $F(1, 9) = 329.7$, $p < 0.001$; introduction, $F(1, 9) = 227.4$, $p < 0.001$.

Regarding between-genre differences, both SR and AR differed between the bulletin and introduction (compare SRs and ARs separately in Figure 1). The bulletin reading was faster, with an SR higher by 1.0 syll/s and an AR higher by 0.6 syll/s. A one-way ANOVA confirmed a significant difference for both rates, although the effect was slightly weaker for the articulation rate. SR: $F(1, 9) = 42.9$, $p < 0.001$; AR: $F(1, 9) = 19.6$, $p \approx 0.002$.

In addition to using overall values, AR was also calculated as a weighted mean of the AR from individual inter-pause stretches. A one-way ANOVA confirmed a significant difference in these recalculated ARs between the bulletin and intro as well. $F(1, 9) = 18.6$, $p \approx 0.002$.

The computation of variation metrics (Table 1) was based on 60 data points for the bulletin (10 speakers × 6 paragraphs) and 40 data points for the introduction (10 speakers

× 4 paragraphs). This means that transition pauses between paragraphs and any potential final phrases (P5) in the introduction were excluded from this calculation.

Table 1 Variation metrics for speech rate (SR) and articulation rate (AR) across the bulletin and introduction performances, expressed in syllables per second (syll/s).

Rate – genre	C_{var} (%)	Min	Max	Range
SR – bulletin	8.9	4.4	6.9	2.5
SR – intro	16.0	3.2	6.2	3.0
AR – bulletin	6.8	5.3	7.1	1.7
AR – intro	11.2	4.3	6.7	2.4

The values for the bulletin are more compact than those for the intro, for both SR and AR. The minimum rates observed in the bulletin were higher than those recorded in the intro (by 1.2 syll/s for SR and 1.0 syll/s for AR). The maximum rates recorded in the bulletin were also higher, although the difference was less pronounced (0.7 syll/s for SR and 0.4 syll/s for AR). Consequently, the range is narrower for the bulletin for both rates.

The coefficient of variation (Cvar) for bulletin was below 10%, indicating highly concentrated values. The Cvar for the intro was somewhat higher, particularly for SR (16.0%), while the AR value of the intro (11.2%) only slightly exceeded 10%.

3.1.2 Pauses in performances

The measurements depicted in Figure 2 are based on overall performance values.

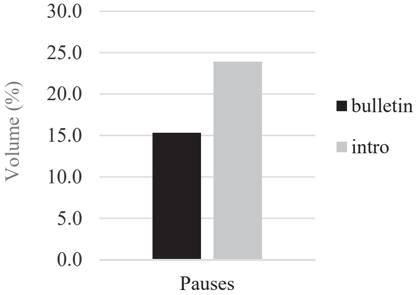

Figure 2 Mean volume of pauses (as a percentage of total duration) in two performances: bulletin (N = 10), and introduction (N = 10). The results are based on overall values.

It is clear that the genres differ in their overall pauses volume; in the bulletin, pauses constitute, on average, approximately 15% of the performance, while in the intro, they account for about a quarter of the duration. An ANOVA for repeated measures returned a significant difference between the bulletin and intro, $F(1, 9) = 29.3$, $p < 0.001$. A higher Cvar value was measured for the intro (25.1%) than for the bulletin (16.0%); however, both values still indicate relatively compact data.

In addition to pause volume, pause duration was also measured. Figure 3 depicts the mean pause duration in two forms: as measured (non-normalized) data and normalized data. The contrast between these two displays is evident: while the measured mean pause duration reveals differences between bulletin and introduction (left side of the figure), the normalized pause durations remain relatively consistent across both performance types (right side). This observation is confirmed by statistical significance testing. A one-way ANOVA returned a non-significant result for the normalized data, $F(1, 9) = 0.1$, $p \approx 0.7$. However, for the non-normalized pause duration, the result is significant, $F(1, 9) = 10.6$, $p \approx 0.010$.

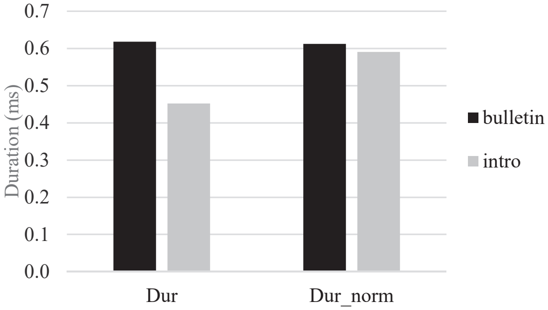

Figure 3 Mean pause duration (non-normalized and normalized) in two performances: bulletin (N = 10), and introduction (N = 10). The results are based on overall values.

It must be noted that although the total volume of pauses in the bulletin is significantly lower than in the intro, the mean non-normalized duration of a single pause is higher (and vice versa for the intro).

3.2 Within-genre differences

This analysis is based on paragraph-level data, which excludes transition pauses. For the bulletin, six paragraphs were examined, and for the introduction, four paragraphs were analyzed. Each paragraph represents one *genre-unit* (cf. Volín, 2022).

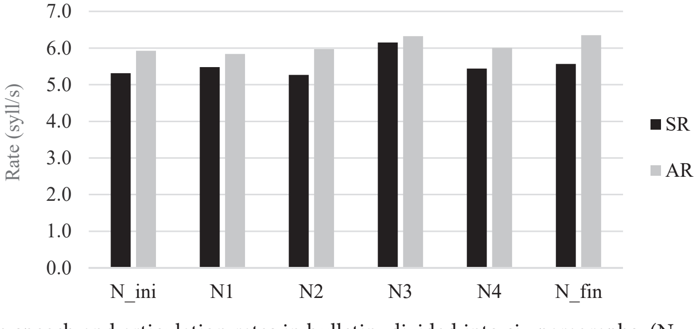

Figure 4 Mean speech and articulation rates in bulletin, divided into six paragraphs. (N = 10 for each paragraph)

3.2.1 Bulletin

Figure 4 shows that the tempi of the paragraphs are very similar, except for N3, where the SR and AR do not differ to such extent from each other. An ANOVA test for repeated measures indicated that there is a significant difference in SR among the news paragraphs $F(5, 45) = 6.53$, $p < 0.001$ ($\alpha = 0.05$). The post-hoc paired t-test using a Bonferroni corrected $\alpha = 0.0033$ indicated that the means of the following three pairs are significantly different: N3 on the one hand, and N1, N4 and N_fin on the other.

The difference between paragraphs is also significant for AR as returned by an ANOVA for repeated measures: $F(5, 45) = 17.28$, $p < 0.001$ ($\alpha = 0.05$). The post-hoc paired t-test using a Bonferroni corrected $\alpha = 0.0033$ indicated that the means of the following eight pairs are significantly different: N3 and N_fin on the one hand, and N_ini, N1, N2, N4 on the other.

The correlation between the SR and AR values for the paragraphs was also tested to examine whether higher SR scores tend to co-occur with higher AR scores. The Pearson's $r(58)$ value is 0.838, $p < 0.001$. This represents a strong positive correlation, statistically significant.

As noted previously, the paragraph-level analysis excludes transition pauses. Within the bulletin, a noticeable difference was attested between the duration of transition pauses (pauses between paragraphs) and inner pauses (within paragraphs). The mean non-normalized duration was 1.41 s for transition pauses (SD = 0.28 s) and 0.43 s for inner pauses (SD = 0.09 s). The lower Cvar for transition pauses suggests greater uniformity in their duration. However, at nearly 30%, this value approaches the upper limit of what is considered compact data. In contrast, inner pauses exhibited the Cvar of almost 60%, indicating highly dispersed data.

The question arises as to what happens when the distinction between transition and inner pauses in the bulletin is considered comparing pause durations between the bulletin and the intro (3 types of pauses x 10 speakers = 30 datapoints). A one-way ANOVA returned a highly significant result: $F(2, 18) = 123.8$, $p < 0.001$. A post-hoc paired t-test using a Bonferroni corrected $\alpha = 0.01667$ indicated significant differences between the transition pauses in the bulletin on the one hand, and the inner pauses in the bulletin and the pauses in the intro on the other; the difference between the inner pauses in the bulletin and the pauses in the intro was not significant.

3.2.2 Introduction

Figure 5 shows the SR and AR values for the introduction divided into four paragraphs.

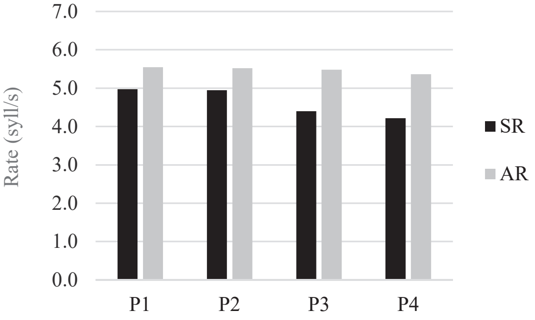

Figure 5 Mean speech and articulation rates for the introduction divided into four paragraphs. (N = 10)

The relationships among the paragraphs of the introduction were examined. A one-way ANOVA returned a significant result for the SR of paragraphs P1–P4 (4 paragraphs x 10 speakers), $F(3, 27) = 4.1$, $p \approx 0.017$. However, the multiple comparisons did not identify a significant difference between any of the pairs. For AR, the lack of significant difference is apparent from the display and was confirmed by an ANOVA test: $F(3, 27) = 0.3$, $p \approx 0.80$.

3.2.3 Course of tempi in the bulletin

The similar tempo patterns observed across speakers suggest the text itself influences the delivery. (These data also relate to inter-speaker variability, see Section 3.3.)

Figure 6 depicting the tempo course between paragraphs with respect to individual speakers shows a conspicuous pattern for SR, in particular in the last four paragraphs. A clear trend may be observed: an increase from N2 to N3, followed by a decrease to N4. The increase to N3 is evident in all speakers (the increase exceeds even 1.0 syll/s in three speakers). The subsequent decrease to N4 is not as extensive and occurs for nine of the ten speakers. For AR, this trend may also be noted, though less prominently (Figure 7).

Interestingly, the three speakers who did not increase their AR from N2 to N3 were among the fastest speakers overall. This observation leads to the assumption that speakers with a higher baseline articulation rate might exhibit less tempo variability. To test this, a Pearson correlation between each speaker's overall AR and their coefficient of variation (based on the inter-pause stretches) was calculated. The result shows only a weak, non-significant negative relationship ($r = -0.16$, $p \approx 0.65$). Furthermore, the Cvar values themselves are consistent across the group, falling within a narrow range of 9.6% to 13.9%, which indicates a similar level of tempo stability among all speakers.

For both SR and AR, the majority of speakers exhibited an increase in tempo between N4 and the final paragraph N_fin. In the case of SR, however, the extent of this increase was notably smaller than the rise observed between paragraphs N2 and N3.

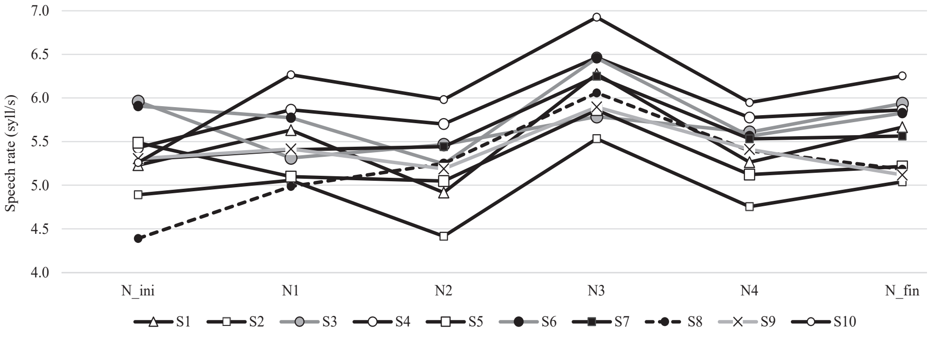

Figure 6 Mean speech rate in bulletin paragraphs, indicating the tempo course for individual speakers (S1–S10).

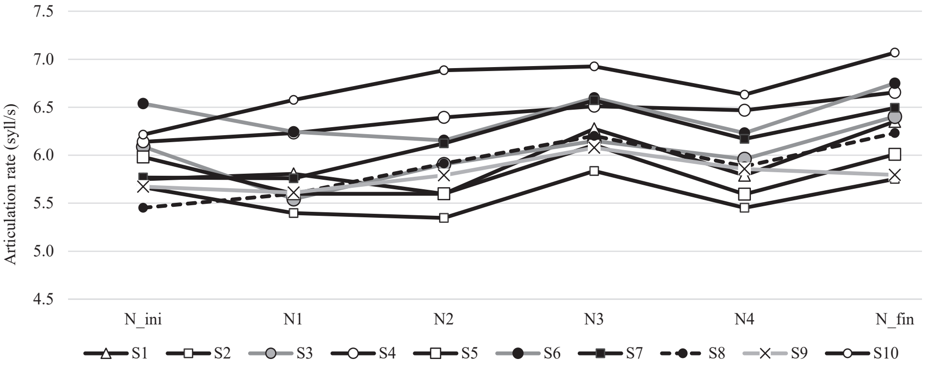

Figure 7 Mean articulation rate in bulletin paragraphs, indicating the tempo course for individual speakers (S1–S10).

3.3 Inter-speaker differences

3.3.1 Speech tempo

The displays in Figure 8 (bulletin) and 9 (introduction) are based on the overall SR and AR for each speaker.

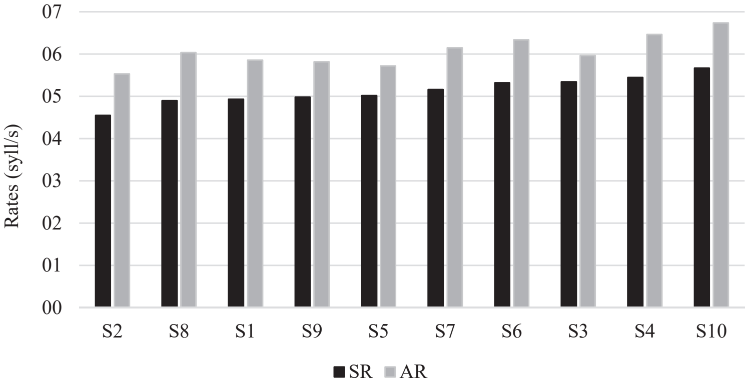

Figure 8 Mean speech and articulation rates produced by individual speakers (S1–S10) in the bulletin, ordered by SR values.

Across both genres, all speakers demonstrated a clear difference between SR and AR. Interestingly, within both the bulletin and the intro, the same speaker produced the minimum SR and AR, while another single speaker produced both the maxima in the bulletin; similarly, both maxima were provided by a single speaker in the introduction as well. According to the Pearson correlation test, SR and AR are strongly and positively correlated within both genres, meaning that high values of AR correlate with high values SR (and vice versa): $r = 0.90$, $p < 0.001$ (bulletin), $r = 0.87$, $p < 0.001$ (introduction).

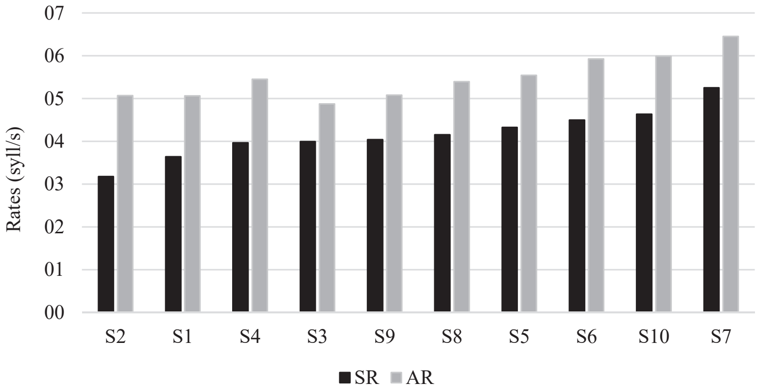

Figure 9 Mean speech and articulation rates produced by individual speakers (S1–S10) in the introduction, ordered by SR values.

A high degree of similarity among speakers within a single genre is evident, particularly in the bulletin. This observation is confirmed by the data presented in Table 2, which demonstrate low dispersion as evidenced by the Cvar values. Of the measurements, the only Cvar value to exceed 10% was for SR in the introduction. This higher variability corresponds directly to the introduction, having the widest SR range (maximum – minimum). Furthermore, the data show that the minima and maxima for both SR and AR were higher in the bulletin than in the introduction, indicating that the overall tempo was faster in the bulletin genre.

Table 2 Variation metrics for speech rate (SR) and articulation rate (AR), based on the individual speaker values for each performance. The rates are expressed in syllables per second.

Rate – genre	C_{var} (%)	Min	Max	Range
SR – bulletin	6.3	4.5	5.7	1.1
SR – introduction	13.6	3.2	5.3	2.1
AR – bulletin	6.1	5.5	6.7	1.2
AR – introduction	9.2	4.9	6.4	1.6

An examination of the ordering of individual SRs by magnitude reveals that certain speakers retained their relative positions across both genres (e.g., S2, see comments to minima above); some speakers moved their positions slightly (e.g., S1 and S10), while for others the change was greater (e.g., S8 and S4 moved by four and six positions, respectively). The question naturally arises as to how strong the correlation is between the speaker ordering across the two genres. A Pearson correlation test returned only a moderate, non-significant positive correlation. SR: $r = 0.57$, $p = 0.087$; AR: $r = 0.59$, $p = 0.073$.

To test the influence of very short stretches, AR for each speaker was calculated under three conditions. The first, set 1 (used so far), included all speech stretches regardless of their size. For set 2, 1-syllable stretches were excluded, and for set 3, both 1-syllable and 1-word stretches were excluded. An ANOVA for repeated measures indicated a significant overall difference in weighted mean of ARs among the three sets: $F(2, 18) = 23.4$, $p < 0.001$. The post-hoc paired t-test using a Bonferroni corrected $\alpha = 0.01667$ indicated that the means of all three pairs differ significantly. However, the magnitude of the difference between the averages is small. The largest difference observed between set 1 and set 3 was just 0.4 syll/s for one speaker and 0.3 syll/s for three others. For the majority (six speakers), no difference in the calculated AR was noted. These results suggest that while the inclusion of short stretches produces a statistically significant effect overall, its practical impact on AR values is negligible for most speakers.

3.3.2 Volume of pauses

Figure 10 displays the volume of pauses for individual speakers based on overall values.

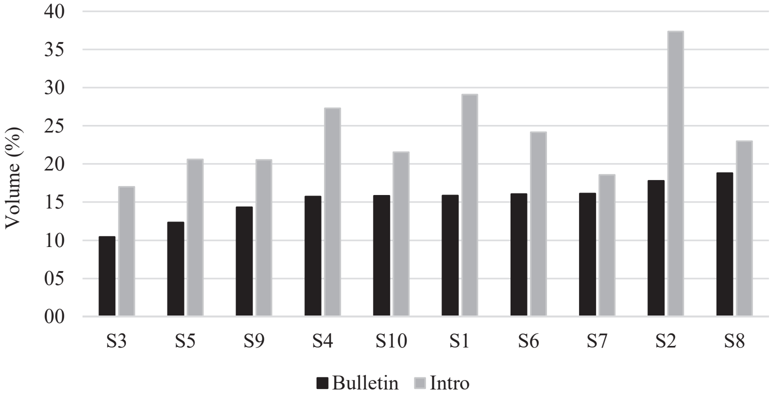

Figure 10 Volume of pauses (in %) produced by individual speakers (S1–S10) in the bulletin and introduction, ordered by percentage in the bulletin.

In contrast to the bulletin, inter-speaker differences in the volume of pauses in the introduction are evident. For example, speaker S7 exhibits a notably low value in the introduction, comparable with the value in the bulletin. In contrast, three other speakers display values up to twice as high in the introduction as those recorded in the bulletin. It is clear that the values of pause volume between both genres are not correlated. A Pearson correlation test returned only a moderate, non-significant positive correlation: $r = 0.57$, $p = 0.083$.

The greater compactness of pause volume in the bulletin compared to the introduction is also evident from the coefficients of variation (15.9% vs. 24.9%) and the range (8.4% vs. 20.2%), as shown in Table 3.

Table 3 Variation metrics for pause volume (in %), based on individual speaker values.

Genre	C_{var}	Min	Max	Range
Bulletin	15.9	10.5	18.8	8.4
Intro	24.9	17.0	37.2	20.2

The speech of some speakers in the intro contained perceptually noticeable passages with hesitation sounds. Figure 11 shows the volume of pauses for individual speakers in the introduction, indicating the ratio between parts without hesitation and those containing hesitation. The volume of hesitation sounds is a source of inter-speaker variability; several speakers' pauses were not filled with hesitation sound at all (S4) or very little (S7, S6). On the contrary, in some speakers the volume of these pauses was relatively high (S2, S5).

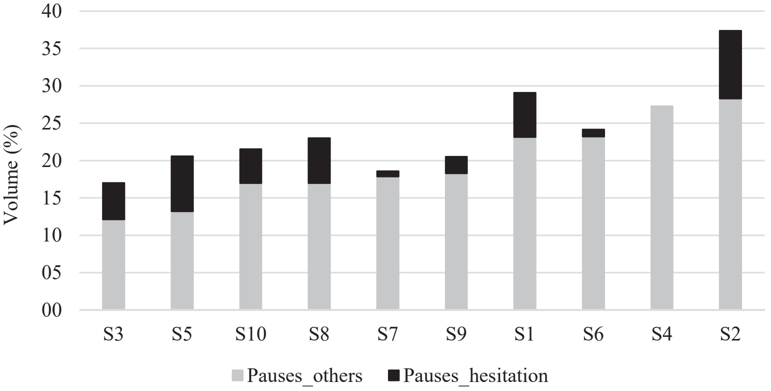

Figure 11 Volume of pauses (in %) produced by individual speakers (S1–S10) in the introduction, indicating the ratio between parts with and without hesitation (ordered by the magnitude of volume without hesitation).

There appears to be no correlation between the total volume of pauses and the volume of pauses containing hesitation sounds (compare, e.g., the display for speakers S5, S4, and S2). This is confirmed by a Pearson correlation test, which returned a non-significant result: $r = -0.02$, $p = 0.95$.

4. Discussion

The focus of the current study was on tempo variation and pausing in two distinct speech genres: a read-aloud bulletin and a semi-spontaneous self-introduction. The main outcomes are discussed in this section and compared with relevant findings of Volín (2022), who analyzed news reading and poetry reciting.

The two genres examined in the current study differed significantly in tempo. The read-aloud news bulletin was significantly faster than the self-introduction, in terms of both speech rate (SR) and articulation rate (AR). These outcomes are consistent with those of Volín (2022), where news reading was faster than poetry reciting. However, the differences between SR and AR observed in Volín's study were substantially greater than those identified in our dataset. This suggests that tempo of a semi-public introduction is closer to news reading than poetry reciting is.

The tempo in the bulletin was also much more compact than in the introduction, as indicated by variation metrics. The coefficient of variation (Cvar) values for both rates in the bulletin did not exceed 9%, which aligns perfectly with the corresponding value for news in Volín's material. However, a difference emerged: while Volín found that only the speech rate in poetry showed high variation (with articulation rate remaining stable), the current study found that both rates (SR and AR) displayed higher variation in the introduction. Moreover, the Cvar for speech rate in the introduction was somewhat higher than that for poetry reading in Volín's sample (16.0% vs. 12.3%).

Regarding pauses, two aspects were examined: volume and duration. The introduction contained a significantly higher volume of pauses than the bulletin. Although the total pause time was lower, the average duration of individual pauses was significantly greater in the bulletin. When pause duration was normalized to account for each speaker's articulation rate, this difference between the genres became non-significant.

However, these overall data on pause duration masked an important finding. Transition pauses (between paragraphs) in the bulletin were significantly longer than inner pauses (within paragraphs). While the Cvar for transition pauses was close to the limit of compactness, the inner pauses were highly dispersed. Crucially, a between-genre comparison revealed that the inner pauses in the bulletin were not significantly different in duration from the pauses found in the introduction. These outcomes indicate that pauses in the bulletin likely serve the function of structural markers, with longer pauses at paragraph boundaries signalling a topic shift. In the future, not only overall pause volume but also the specific placement, frequency, and acoustic features of pauses should be examined in more detail, including their role in expressing information structure and their perceptual impact on listeners.

Regarding the analyzed speech material, it is important to consider an additional factor that may have influenced the observed genre differences, namely the time interval between the recordings, which was one year for most speakers. The examination of intra-genre differences revealed that certain news paragraphs differed significantly in tempo from others. Moreover, most speakers showed a similar pattern of tempo changes across the paragraphs when reading the bulletin. A preliminary examination suggests that factors such as the number of syllables or sentences may affect the tempo within individual news paragraphs. Volín (2022: 71) noted a mean tempo deceleration in news reading across paragraphs; in this context, he indicated (without further details) the potential influence of the content of the paragraph on tempo (e.g. domestic news vs. foreign news vs. sport). The extent to which linguistic content modulates speech tempo represents a promising area for future research.

In contrast to the bulletin, the introduction did not exhibit a consistent tempo pattern across the paragraphs. While some variation is attested, no single paragraph was consistently faster or slower than another. The articulation rate seemed stable across the paragraphs of the introduction.

Concerning inter-speaker differences, a speaker's SR and AR were strongly and positively correlated within each genre. However, a cross-genre comparison indicated that a speaker's tempo in the bulletin is not significantly correlated with their tempo in the introduction. This suggests that speakers may apply different tempo strategies in different genres, a finding consistent with Volín (2022), who also reported only a moderate correlation between speakers' performances in poetry and news reading.

Regarding within-genre variation, speakers displayed a high degree of similarity, particularly when reading the bulletin. Greater individual differences were noted in the introduction, especially in speech rate. This aligns with Volín's observation of low inter-speaker differences within a given genre.

Large differences between speakers in the volume of pauses were evident only in the introduction; this measure was not correlated between genres, however. Inter-speaker variability in the volume of hesitation sounds was also high in the introduction; nevertheless, there was no correlation between a speaker's total pause volume and their use of hesitation. For example, the speaker with the second-highest overall pause volume produced no hesitations. This suggests that speakers have different strategies for planning speech or masking nervousness; some accumulate pause time through silent pauses (which may be either long or frequent), while others use filled pauses. This could be a fruitful area for future research, including examination of the impact of speech fluency on listeners.

The use of semi-public performances extends the range of speech materials often used in phonetic research. The presence of an audience likely heightened cognitive load and nervousness, thereby influencing speakers' pause behaviour. Compared to news reading, the self-introduction is a less constrained task where individual planning abilities become more prominent.

Although a relatively small sample was analyzed, the results of the current study suggest that speaking style and/or speech genre is not just a minor variable but a force that has the potential to reshape a speaker's entire temporal strategy, influencing not only speed but also the very function and nature of their pauses. The detailed procedure presented in the study enables both the replication and the extension of speech sample.

Acknowledgements

This publication was supported by the Cooperatio Program provided by Charles University, research area Linguistics, implemented at the Faculty of Arts of Charles University.

I would like to thank both anonymous reviewers for their valuable comments and recommendations.

REFERENCES

Balkó, I. (2005). K výzkumu tempa řeči a tempa artikulace v různých řečových úlohách. *Bohemistyka,* nr 3, 185–198.

Barik, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech*, *20*(2), 116–126.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* (Version 6.0.23). retrieved 28. 12. 2019 from http://www.praat.org.

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America*, *136*, Express letters, 116–121.

Bořil, T., & Oceláková, Z. (n.d.). *Počet slabik a AR (rPraat)*. [Number of syllables and AR (rPraat)], [script] retrieved 23. 5. 2024 from https://fonetika.ff.cuni.cz/vyzkum/skripty-a-nastroje.

Dankovičová, J. (2001). *The linguistic basis of articulation rate variation in Czech*. (Forum Phoneticum 71). Hector.

Ferguson, S. H., Morgan, S. D., & Hunter, E. J. (2024). Within-talker and within-session stability of acoustic characteristics of conversational and clear speaking styles. *Journal of the Acoustical Society of America*, *155*(1), 44–55.

Hanžl, V., & Hanžlová, A. (2022). *Prak*. [software] retrieved 2. 5. 2025 from https://github.com/vaclavhanzl/prak.

Hanžl, V., & Hanžlová, A. (2023). Prak: an automatic phonetic alignment tool for Czech. In *Proceedings of XXth International Congress of Phonetics Sciences*. Prague. ID 525, pp. 3121–3125

Huszár, A., & Krepsz, V. (2022). The development of variability in pausing and articulation rate in Hungarian speakers ten years apart. *Govor*, *38*(2), 121–146.

Jacewicz, E., & Fox, R. A. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, *128*(2), 839–850.

Kohler, K. J. (1986). Parameters of speech rate perception in German words and sentences: duration, F0 movement, and F0 level. *Language and Speech*, *29*, 115–140.

Koopmans-van Beinum, F. J., & van Donzel, M. (1996). Relationship between discourse structure and dynamic speech rate. In *Proceeding of Fourth International Conference on Spoken Language Processing*. https://doi.org/10.1109/ICSLP.1996.607960.

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.

Machač, P., & Skarnitzl, R. (2009). *Fonetická segmentace hlásek*. Nakladatelství Epocha.

Mixdorff, H., Pfitzinger, H. R., & Grauwinkel, K. (2005). Towards objective measures for comparing speaking styles. In *Proceedings of X*th *Speech and Computer – SPECOM 2005*, Patras, Greece, pp. 131–134.

Oceláková, Z. (n.d.). *Počítadlo slabik*. [Syllable calculator] [script] retrieved 23. 5. 2024 from https://keys.shinyapps.io/slabikovac.

Plug, L., Lennon, R., & Smith, R. (2022). Measured and percepived speech tempo: Comparing canonical and surface articulation rates. *Journal of Phonetics*, *95*, 1–15.

Pollák, P., Volín, J., & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In *Proceedings of XIIth Speech and Computer – SPECOM 2007*, pp. 537–541.

Quené, H. (2005). Modelling of between-speaker and within-speaker variation in spontaneous speech tempo. In *Proceedings of Interspeech*, pp. 2457–2460. Lisbon.

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, *35*, 353–362.

R Core Team (2024). *R: A language and environment for statistical computing* (Version 4.4.1). R Foundation for Statistical Computing. Retrieved 19. 5. 2024 from https://www.rproject.org.

Statistics Kingdom. (2017a). *Correlation Coefficient Calculator* [web application]. retrieved 2. 5. 2025 from https://www.statskingdom.com/correlation-calculator.html.

Statistics Kingdom. (2017b). *Repeated Measures ANOVA Calculator* [web application]. retrieved 2. 5. 2025 from https://www.statskingdom.com/repeated-anova-calculator.html.

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*, 297–308.

Veroňková-Janíková, J. (2005). Dependence of individual speaking rate on speech task. *Acta Universitatis Carolinae – Philologica*, *2005*(1), *Phonetica Pragensia X,* 107–123.

Veroňková, J., & Poukarová, P. (2017). The relation between subjective and objective assessment of speaking rate in Czech radio newsreaders. *Acta Universitatis Carolinae – Philologica, 2017*(3), *Phonetica Pragensia*, 95–107.

Volín, J. (2019). The size of prosodic phrases in native and foreign-accented read-out monologues. *Acta Universitatis Carolinae – Philologica, 2019*(2), *Phonetica Pragensia*, 145–158.

Volín, J. (2022). Variation in speech tempo and its relationship to prosodic boundary occurrence in two speech genres. *Acta Universitatis Carolinae – Philologica, 2022*(1), *Phonetica Pragensia*, 65–81.

Volín, J., Skarnitzl, R., & Pollák, P. (2005). Confronting HMM-based phone labelling with human evaluation of speech production. In *Proceedings of Interspeech 2005*, pp. 1541–1544.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of Interspeech 2006 – ICSLP*, Pittsburgh, pp. 541–544.

RESUMÉ

Příspěvek analyzuje mluvní tempo, artikulační tempo a pauzy ve dvou různých typech žánrů, kterými jsou čtené zprávy a polospontánní projev, v němž se mluvčí představuje přítomnému publiku. Řečový materiál tvoří nahrávky od deseti neprofesionálních mluvčích – žen. Mluvní tempo i artikulační tempo zpráv je vyšší než u představení se, zprávy dále obsahují menší objem pauz, ovšem průměrné trvání pauz je naopak vyšší; všechny uvedené rozdíly jsou statisticky významné. Při rozlišení pauz mezi dílčími zprávami a uvnitř těchto zpráv však rozdíl mezi vnitřními pauzami a pauzami v představení se přestane být významný. Příspěvek dále přináší údaje týkající se variability uvnitř žánrů a mezi mluvčími.

Jitka Veroňková
Institute of Phonetics
Faculty of Arts, Charles University
Prague, Czech Republic
jitka.veronkova@ff.cuni.cz