

Article

Occurrence and Duration of Pauses in Relation to Speech Tempo and Structural Organization in Two Speech Genres

Pavel Šturm *  and Jan Volín 

Institute of Phonetics, Charles University, 116 38 Prague, Czech Republic

* Correspondence: pavel.sturm@ff.cuni.cz

Abstract: Pauses act as important acoustic cues to prosodic phrase boundaries. However, the distribution and phonetic characteristics of pauses have not yet been fully described either cross-linguistically or in different genres and speech styles within languages. The current study examines the pausal performance of 24 Czech speakers in two genres of read speech: news reading and poetry reciting. The pause rate and pause duration are related to genre differences, overt and covert text organization, and speech tempo. We found a significant effect of several levels of text organization, including a strong effect of punctuation. This was reflected in both measures of pausal performance. A grammatically informed analysis of a subset of pauses within the smallest units revealed a significant contribution for pause rate only. An effect of tempo was found in poetry reciting at a macro level (speaker averages) but not when pauses were observed individually. Genre differences did not manifest consistently and analogically for the two measures. The findings provide evidence that pausing is used systematically by speakers in read speech to convey not only prosodic phrasing but also text structure, among other things.

Keywords: pause; prosodic phrase; text structure; articulation rate; speech genre



Citation: Šturm, Pavel, and Jan Volín. 2023. Occurrence and Duration of Pauses in Relation to Speech Tempo and Structural Organization in Two Speech Genres. *Languages* 8: 23. <https://doi.org/10.3390/languages8010023>

Academic Editors: Jürgen Trouvain and Bernd Möbius

Received: 1 November 2022

Revised: 16 December 2022

Accepted: 4 January 2023

Published: 11 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is plenty of evidence that the division of the speech continuum into smaller prosodic units matters. Such units are beneficial on several planes of speech processing. Their role has been observed in the neurophysiology of the human brain (e.g., [Peelle and Davis 2012](#); [Ghitza et al. 2013](#); [Martin 2015](#)), language decoding ([Lehiste 1973](#); [Watson and Gibson 2005](#); [Elmers et al. 2021](#)), and speaker acceptance ([Niebuhr and Fischer 2019](#)). One of the most obvious devices of prosodic division is the pause.

However, various types of languages (but also speech styles and speech genres within those) still lack carefully conducted large-scale descriptions of pausal performance with regard to factors affecting the duration and distribution of pauses. For the present study on the Czech language, we investigate a medium-sized sample of more than 3000 pauses in two genres of read speech: news reading and poetry reciting. Such a focus brings the research on pauses to a genre (poetry reciting) that is little studied. Our major research question arises from the difference between the two genres in the rhythmicity and level of structural organization of the texts. Another research question pertains to the relationship between pause characteristics and speech tempo.

1.1. Background to Pausing

Boundaries between prosodic phrases are marked by a variety of acoustic cues, most notably, the presence of pauses, lengthening of the final syllable, or the use of specific melodic contours ([Fougeron and Keating 1997](#); [Pannekamp et al. 2005](#); [Wagner and Watson 2010](#); [Petroni et al. 2017](#); [Paschen et al. 2022](#)). Pauses seem to be especially prominent in marking prosodic boundaries ([Carlson et al. 2005](#); [Männel and Friederici 2016](#)). [Zellner \(1994\)](#) describes pauses as large “beacons” for utterances, structuring them for both speakers

and listeners. In other words, pausal structure is relevant for sentence processing. For instance, both deleting pauses from ‘clear speech’ and inserting pauses into conversations may negatively affect intelligibility (Uchanski et al. 1996). Other studies have shown that pauses, when being the only available cue to the prosodic grouping of digits in synthesized speech, offer a memory advantage to listeners (Elmers et al. 2021). Zellner (1994) also considers pauses in the output of speech synthesis, suggesting that it “will sound more fluent, will be more pleasant to listen to, and will likely be more intelligible when silent and filled pauses are systematically integrated into the verbal stream” (p. 60). Werner et al. (2022) emphasized that what is needed currently in speech synthesis is adding more variability to pause durations, adding breath noises, and considering the optionality of pause locations.

The distinction between ‘silent’ and ‘filled’ pauses is a long-standing and popular one but may be misleading in that ‘silent’ pauses are usually not acoustically silent (Trouvain et al. 2016). Many pauses are in fact produced with breath noises or other sounds (laughter, clicks), or with some type of vocal activity (hesitation sounds, repairs, etc.). Using a specific type of pause may be related to sentence structure or turn-taking devices in conversations. For instance, Maclay and Osgood (1959) showed that the number of filled pauses was higher at phrase boundaries than within, whereas silent pauses occurred more often within phrases. The authors conclude that filled pauses may be produced especially when the speaker wishes to keep the floor and continue speaking, although other uses are also likely and speaker behavior in this respect is highly individual (see van Donzel and Koopmans-van Beinum (1996), who demonstrated that the choice of a pausing strategy is highly speaker-dependent).

Regardless of what type of pause is examined, a large-scale analysis of pause characteristics with regard to various factors is needed. Campione and Véronis (2002) investigated over 6000 silent pauses in several hours of mostly read speech in five languages. Two methodological observations should be noted. First, the authors showed that imposing a duration threshold for measurement (i.e., discarding very short and very long pauses), as is commonly done, leads to incorrect conclusions. Second, pause duration is distributed log-normally, which affects the method of analysis. A multimodal distribution of silences in their data was observed due to a combination of brief (<200 ms), medium (200–1000 ms), and long (>1000 ms) pauses.

The factors that affect pause location and duration are numerous. Syntax and discourse structure has always been at the forefront (Lehiste 1973; Cooper and Paccia-Cooper 1980; Zellner 1994; Kjelgaard and Speer 1999; Frazier et al. 2006; Carlson 2009). Studying Dutch, van Donzel and Koopmans-van Beinum (1996) discovered that, on average, 67% of realized pauses could be related to grammatical structure (16% of pauses were placed after discourse markers or connectives and 51% at syntactic clause boundaries), whereas only a third occurred within clauses. Goldman-Eisler (1972) found that the syntactic structure substantially influenced the duration of pauses as well. There was a scale of “temporal integration in the sentence body from words within clauses, to relative subordinate, to other subordinate clauses, to co-ordinate clauses” (p. 106). In spontaneous speech, 65% of sentence-level utterances were separated by more than 750 ms, while such durations were rare in clause transitions. Words within clauses were associated almost entirely with short pauses.

Zvonik and Cummins (2003) highlighted the relation of pauses to the length of intonation phrases. The probability of pauses below 300 ms increased when either one or especially both of the surrounding phrases were short (<11 syllables). Given this behavior, the authors regard such short pauses as special. More importantly, we can find support in the findings for an argument that the coupling between syntax and pauses is less strong than sometimes thought. If the length of a constituent affects pausing behavior, then the constituent structure may be less relevant, as more and more words can be added without changing the structure. It is thus clear that syntax is only one of the determinants of pause location (Ruder and Jensen 1972; Rochester 1973; Watson and Gibson 2004; Carlson 2009).

Moreover, it should be emphasized that read speech may display a different pattern than unscripted speech. [Werner et al. \(2022\)](#) showed not only that each of their examined languages has several locations in a text that attract pauses, but also that pauses frequently cooccurred with punctuation. As languages may differ in the tightness of correspondence between the use of punctuation and grammatical/prosodic constituency, this factor should be considered in analyses of pausing based on read speech.

Another factor is speaking rate, which affects both pause durations and their number and location, since the prosodic structure may be restructured with changes in tempo. This was explicitly investigated by [Werner et al. \(2022\)](#) who examined 46 speakers from 6 languages reading the same text at five intended speech rates: very slow, slow, normal, fast, and very fast. With the increasing tempo, the speakers used fewer pauses (especially non-breath pauses) and the pauses were generally shorter. The composition of breath pauses (the proportion of silences and inhalations) also changed in specific ways.

Finally, we may relate pauses to speech styles and speech genres. [Goldman-Eisler \(1972\)](#) examined the modifications introduced when originally spontaneous speech was read out by other speakers. Long pauses (>750 ms) within clauses disappeared in the reading as well as at boundaries preceding relative subordinate clauses; however, they disappeared less abruptly in the other types of subordinate clauses, and a substantial number remained between coordinate clauses and especially between sentences. Additionally, genre differences were investigated by [Hieke et al. \(1983\)](#). The reading of an English translation of a French poem by 24 speakers was compared to several speeches or statements read by the US politicians Reagan and Carter (and the German politician Schmidt). The focus was on very short pauses (130–250 ms) as an argument against Goldman-Eisler's omission of these in her investigations. The results showed that, in this language (English), the proportion of short pauses was much higher in poetry (29%) than in political speeches (8%). The authors discuss the likely differences between the two types of speakers: "The politicians are professional speakers, whereas our poetry readers were ordinary university students; the speeches were given under broadcast conditions, the readings were not; poetic format imposes a specific sequential structure different from the prose of a political text" (p. 209). Furthermore, in the readings of poetry, most of the short pauses occurred "at the end of poetic lines or at punctuated positions" (p. 211).

1.2. The Current Study

Our major research question arises from the difference between two GENRES—poetry reciting and news reading—and the rhythmicity and level of structural organization of the texts. Specifically, we ask how (if at all) such differences influence the pause properties. [Hieke et al. \(1983\)](#) suggested a strong correlation with verse lines and punctuation. However, they compared poetry with speeches delivered by professional politicians. In contrast, we compare two genres produced by the same speakers, canceling any speaker-related confounds that could appear in a between-subjects design. It can be expected that (1) the aesthetic function in poetry reciting would be inductive to a generally higher use of pauses and/or their longer duration than in news reading at corresponding levels of organization and that (2) additional constraints on the organization of poetry (such as the verse line) will be relevant as well.

The second question probes the relationship between the occurrence of pauses in the text and ARTICULATION RATE. [Werner et al. \(2022\)](#) claimed that "at faster rates, non-breath pauses tend to disappear, breath pauses become less frequent, and breath group size increases" (p. 312), and their results supported this assertion. However, in their material, the same speakers read at different tempos, whereas our aim is to discover whether inherently fast-articulating speakers make fewer or shorter pauses and vice versa. Moreover, the authors correctly say that their five intended speech rates might have been interpreted diversely: "while the faster rates may be more naturally limited by how fast a given speaker can produce speech, the slower ones may be less uniform and may thus lead to very long pauses in extreme cases" (p. 315).

Third, the CONTENT of the texts must be considered. Rochester (1973) called for research investigating the relationship between the location and function of pauses, suggesting that their occurrence might be connected to “semantic or propositional factors in the utterance” (p. 78). The levels of unit organization will be predicted following an innovative and grammatically informed approach by Franz et al. (2022). Such prediction will be compared with the occurrence and form of pauses; stronger breaks within sentences are expected to yield a greater number of pauses and/or pauses of longer duration.

2. Materials and Methods

2.1. Material

The material consists of two speech genres: news reading and poetry reciting. A sample of 24 Czech speakers (11 male, 13 female, mean age = 24.3 years, range = 19–33 years) produced speech in both genres. Apart from the age, they were constrained to current or former students of philological programs at Charles University who volunteered, based on the requirement of a positive attitude to poetry. None of them was professionally trained but they had reciting experience from school or afternoon clubs. The speakers were naive to the purpose of the current study, as their recordings were part of a database obtained from students during their coursework in order to generally assess their own speech performance.

News reading was examined in four unrelated texts taken from actual broadcasted news (two reports on domestic events, one foreign political report, and one domestic sports report). The speakers originally produced six texts, but the first and last ones were omitted to balance the compared material in terms of the number of thematic pieces. The key structural unit in the news reading was the SENTENCE, defined as a grammatically complete syntactic structure terminated with a full stop (period). Each speaker produced 19 sentences; the mean length of a sentence was 36.8 syllables (699 syllables in total, 292 words).

Poetry reciting was examined in four Czech rhymed poems of different meter (2× iambic, 1× trochaic, 1× dactylic) written in the early 20th century. The poets were František Gellner (two poems: one tragic, one satirical), Karel Toman (one romantic poem), and Fráňa Šrámek (one ironic poem). Enjambment occurred only once in two of the poems. The key unit of organization was the verse DISTICH, corresponding to the sentence in the news reading. The term distich refers to two verse lines that belong together semantically and grammatically. The two lines usually form two parts of one utterance. For instance, the following stanza contains two distichs [D1, D2]:

<i>Jak lovcí rozstřílený dravec</i>	Like an eagle shot by the hunters	
<i>slét balon v srázy ledných skal.</i>	A balloon crashed into icy rocks.	[end of D1]
<i>Zřítíl se z člunu vzduchoplavec</i>	Fell down with his vessel an aeronaut	
<i>a údy své si roztřískal.</i>	And horribly smashed his limbs.	[end of D2]

Each speaker produced 38 distichs; the mean length was 20.1 syllables, amounting to a total of 763 syllables (410 words) in the four poems. The poetry also has overt organizational structures both above and below the distich (see Section 2.3). In contrast, news items have no such structure. In both materials, the pause after the titles was excluded from analyses.

The recording sessions proceeded as follows. In both genres, the speakers had sufficient time for preparation before reading and they were required to practice the task first (therefore, mispronunciations were rare). They were instructed to strive for fluency and to imagine they are expressing themselves and not impersonating another speaker (e.g., a broadcaster). Each text item (news item or poem) was presented to the reader on a separate paper after finishing the previous item. Crucially, unlike in the genre of news reading, the participants were not simply reading the poetry. They were instructed to recite the poem, to perform it, although not from memory. The aesthetic function of poetry was thus an inseparable part of the production in contrast to the mostly informative function of reading the news.

2.2. Phonetic Segmentation and Pause Measurement

All recordings were processed identically. The first step was automatic forced alignment (words and phones) using the Prague Labeller (Pollák et al. 2007), and the second was a preliminary approximate correction of word and phone boundaries in Praat (Boersma and Weenink 2022) by student research assistants, following the principles of phonetic segmentation in Machač and Skarnitzl (2009). The authors then performed manual segmentation of pauses and the adjacent segments, adhering to the following criteria:

- **Start of the pause aligned with the end of vocal activity;** namely, the end of friction noise after fricatives, affricates, and plosive bursts; the end of formant structure after sonorants when the transition was abrupt and unambiguous; and the end of the devoiced portion of sonorants when their articulation was weakened, leading to voiceless formants (but excluding potential breath noises).
- **End of the pause aligned with the start of vocal activity;** namely, the start of a visible acoustic reflection of articulatory activity (friction, formant structure); however, initial voiceless plosives with silent closures, lacking visible information in the spectrograms, were segmented in such a way that their total duration ranged between 50–100 ms (typically) or between 50–120 ms when a strong emphasis was produced and perceived on the word (mostly in poetry reciting). The remaining portion of the silence was annotated as a pause (see Figure 1).

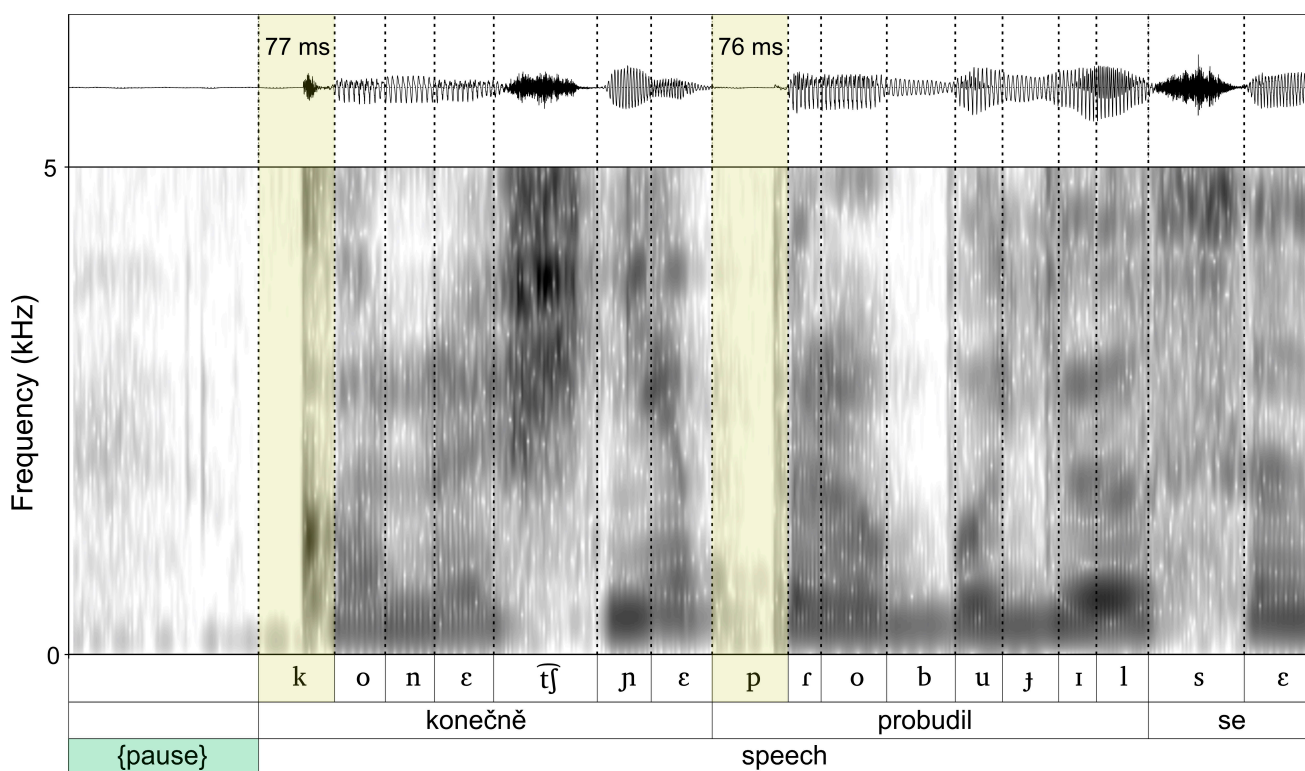


Figure 1. Segmentation of word-initial plosives. Their duration is annotated below a threshold (see text) and any remaining ‘silence’ (which may be filled with acoustic signal) is annotated as a pause.

In the early days of pause research, it was believed that pauses below 250 ms or a similar threshold were not indicative of ‘real’, intended pauses, but rather as articulatory breaks comprising the silent hold phase of plosives, slowing down, or the separation of adjacent sounds (e.g., Goldman-Eisler 1968). However, Hieke et al. (1983) clearly showed that such an interpretation is unfounded, and many others have voiced concerns that the inconsistency in setting up cut-off points for measuring pause duration has profound consequences on the results (see Campione and Véronis 2002 for verification). Hieke

[et al. \(1983\)](#) lowered the threshold of silence to 130 ms, arguing that longer durations are psychologically functional (intentional).

However, the best option seems to be to abandon such a threshold. Instead, pauses before words with initial plosives can be excluded from analyses entirely. Alternatively, we can account for segment-related silences by means of annotating the silences as parts of the segments (e.g., [Werner et al. 2022](#) assigned 50 to 100 ms of silence to word-initial plosives, the rest constituting the pause). In this way, all pauses are annotated, even if short in duration and less easily perceptible. We opted for the latter procedure here (see above). Although the resulting ambiguity in segmentation might play some role in theory, it should be of no practical consequence to the results due to (1) the small extent of the potential bias considering that the duration of pauses is much larger and (2) the randomness in the direction of the potential bias.

2.3. Coding of Pause Contexts (Text Structure)

Our procedure moves from the pause to the description of its context (i.e., its linguistic justification). We suggest two scales of text organization that may constrain the pausing behavior of the speakers. It should be noted, though, that we do not claim that the reversed procedure (proceeding from contexts to pauses) is equally possible, e.g., that each level determines the placement of pauses.

Sentences in news and distichs in poetry (see Section 2.1) were considered salient core units of analyses. In news reading, there was no smaller or larger unit analyzed than the sentence, as the news items did not include, for instance, any coherent paragraphs. However, in poetry, single verse lines and four-line stanzas were also viewed as clear, visually cued smaller and larger units of text organization, respectively. The ends of all these units are often associated with punctuation and loose grammatical coupling. In addition, two other pause contexts are defined within the units, differentiated by the presence or absence of punctuation. Czech comma placement is different from English; it is strictly prescriptive and based on syntactic rules which are taught throughout the educational system. Unlike in English, there are no prosodic considerations involved, as all is governed by syntactic specifics that are usually translated into lists of conjunctions that require a comma. In addition, the Czech language allows for conjunctionless post-positions of coordinated components (e.g., lists of items or structurally equal constituents).

This leads to the following scale of OVERT TEXT STRUCTURE arranged by the predicted probability and salience of a pause (from lowest to highest). Examples can be found in the stanza provided in Section 2.1:

- Within-unit—no punctuation (within a sentence or a verse line);
- Within-unit—with punctuation (within a sentence or a verse line);
- End-of-smaller-unit (end of a verse line);
- End-of-unit (end of a sentence or a distich);
- End-of-larger-unit (end of a stanza).

However, the distinction between punctuated and non-punctuated within-unit elements is probably too crude. Therefore, a further classification of potential pause placement was inspired by the innovative and grammatically informed approach of [Franz et al. \(2022\)](#), who summarized a large body of research and proposed a predictive transcription system for German. Apart from word-class information, the analysis takes into account the size of the syntactic constituents (see tables in [Franz et al. 2022](#), pp. 57–61). We also had to add rhythmical considerations since the above-mentioned authors worked only with read prose, not poetry. This leads to the following scale of what we call COVERT TEXT STRUCTURE, for simplicity's sake. It should be noted that the first two classes are termed “phrasing ex negative” by [Franz et al. \(2022, p. 14\)](#), that is, specifying where no pauses are allowed. In the other three, stronger breaks (looser connections), marked with the # symbol, are predicted to increasingly facilitate pausing:

- -II: Pausing blocked
 - within multi-word personal names, e.g., *Bohuslav Sobotka* (first and family name)
 - after proclitics, e.g., *pod kontrolou* (under control)
 - before enclitics, e.g., *potvrzuje to* (confirms it)
- -I: Pausing not recommended
 - within genitive constructions (without modifiers), e.g., *rozvoje města* (expansion of the city)
 - between an adjective + noun, e.g., *plavovlasou lásku* (fair-haired sweetheart)
 - between a numeral + noun, e.g., *dva stupně* (two degrees)
- I: Weaker break
 - at the subject–verb division when the subject consists of one noun and at most one modifier, e.g., *muži nad sklenkou # půl ironicky sní* (men over glasses # dream half ironically)
 - before a conjunction + one-word constituent, e.g., *lože smrtelné # a hrob* (deathbed # and grave)
- II: Break
 - before longer complements of at least two autosemantic words, e.g., *# kvůli údajným neregulárnostem* (# because of alleged irregularities)
 - before the final adverbial of at least two stress groups, e.g., *matku Bůh povolal # ve svoji slávu* (mother was taken by God # into his glory)
 - rhythmical analogy between the verse lines, e.g., *trochu se vraždilo # trochu se kradlo/pereme pereme # špinavé prádlo* (there were some murders # there were some thefts/ /we wash we wash # dirty laundry); [the object in the second line would not be separated from the verb were it not for the analogy with the first line]
- III: Stronger break
 - after the initial adverbial, e.g., *po dnešním jednání představitelů vlády #* (after today's meeting of government members #); [there is no comma in Czech orthography]
 - before an apposition, e.g., *členka komise # Nikola Nováková* (member of the board # Nicola Newman)
 - before a conjunction + multi-word constituent, e.g., *# a dlouhodobým zatížením rozpočtu* (# and long-term burdening of the budget)

2.4. Data Processing and Statistical Analysis

Data processing was performed using R version 4.2.1 (R Core Team 2022). Duration data were log-transformed for the descriptive and inferential analyses. Visualization was performed with *ggplot2* functions included in the *tidyverse* package (Wickham et al. 2019). Boxplots depict median values and variance (hinges: 2nd and 3rd quartiles, whiskers: $\pm 1.5 * IQR$, dots: outlying values). Effect plots are based on a statistical model and depict the predicted means and their 95% confidence intervals.

Statistical analysis involved linear mixed-effects (LME) regression modeling using the *lme4* package (Bates et al. 2015) and *emmeans* (Lenth 2022) for plotting effects and interactions. Count data were modeled with the generalized glmer function ('family = poisson') while duration data were modeled with the standard lmer function. The random effect structure involved random intercepts (SPEAKER, TEXT) and—if the model still converged and did not involve singular fits—varying slopes for the fixed effects. In the analysis of counts, the predicted variable is the pause rate per unit of exposure (i.e., it has an upper bound or offset which corresponds to the number of potential pauses in the given cell). In plotting the effects, pause rates are related to a sample of 100 potential pauses to resemble percentages. The significance of effects was evaluated with likelihood-ratio tests by observing whether removing an effect (or an interaction term) from the full model significantly lowers the model's fit. Tukey post-hoc tests were used to evaluate pairwise comparisons

with the Bonferroni correction applied (i.e., the α level is 0.05 but the reported p -values have been multiplied by the number of performed tests).

3. Results

3.1. Data Presentation

3.1.1. General Overview

The material included 3249 pauses, of which 2398 (74%) occurred in poetry reciting and 851 (26%) in news reading. Individual speakers produced, in the combined material, between 91 and 173 pauses (mean = 135.4). The duration of pauses was higher in poetry reciting (median = 471 ms, mean = 553 ms, SD = 409 ms) than in news reading (median = 398 ms, mean = 487 ms, SD = 351 ms). Figure 2 shows the distribution of pause duration. The upper panels demonstrate that raw durations are highly skewed (cf. [Campiono and Véronis 2002](#)); the lower panels display the data after logarithmic transformation. Both speech genres seem to be bimodal, with one mode around 100–150 ms and another around 500–600 ms. The second peak in news reading might itself be composed of two distributions; however, more data would be needed to verify this suggestion. In any case, there seem to be at least two types of pauses, with a tentative boundary of around 200 ms, although a structural definition of these types is needed rather than positing a temporal threshold.

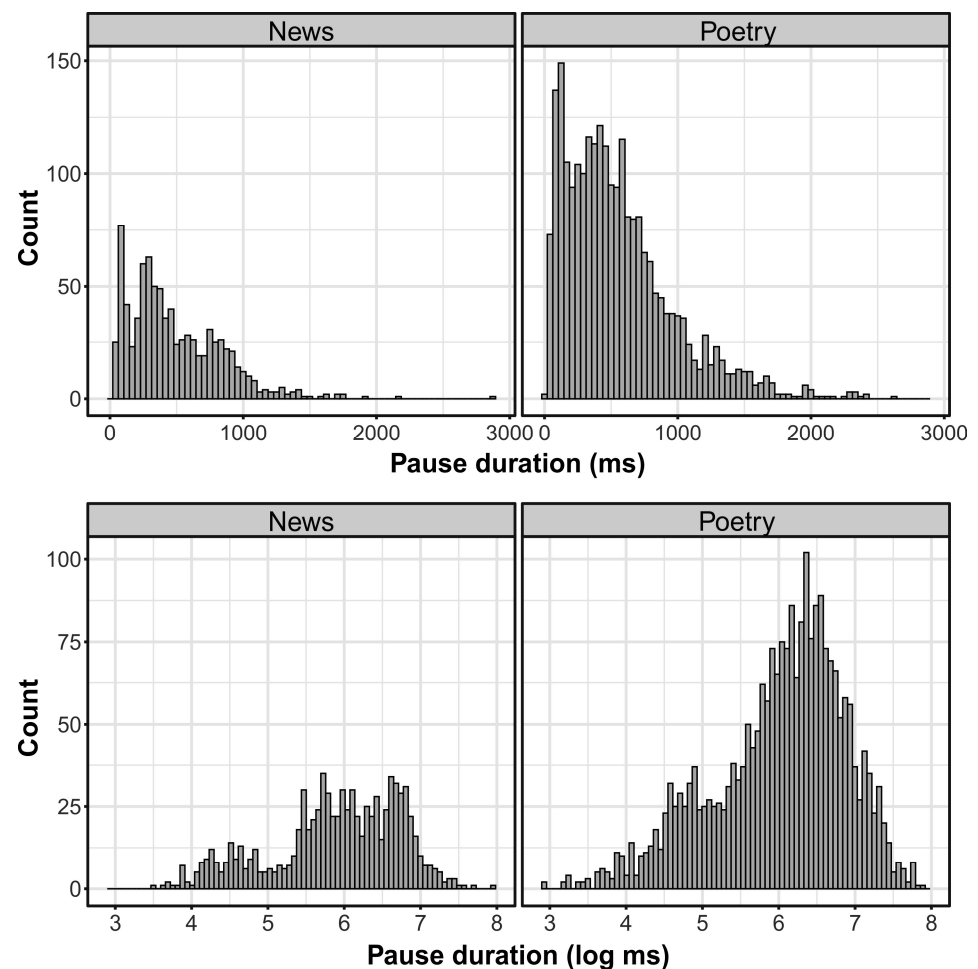


Figure 2. Histograms of pause duration (raw or logarithmic) in two types of materials.

3.1.2. Effect of Overt Text Structure

The occurrence of pauses is summarized in Table 1. In news reading, all sentences were terminated with a pause; within sentences, the key factor was whether there was

punctuation or not. Poetry offered more levels of overt text organization: ends of stanzas were unanimously produced with a pause, and there was a decreasing proportion of pauses at the ends of smaller units (distichs, lines). Similarly to news reading, punctuation mattered greatly within lines: 79% of potential contexts with punctuation involved a pause, whereas the proportion was only 3% for potential contexts without punctuation.

Table 1. Occurrence (absolute and relative) of pauses and their duration as a function of the level of overt text structure in two genres (for an explanation of levels, see Section 2.3). The relative occurrence denotes the percentage of realized to potential pauses in the given context (for instance, the ends of lines offer more opportunities to pause than the ends of stanzas due to their different numbers).

Genre	Level	Pause Context	Absolute Count	Relative Count	Median dur. (ms)	Mean dur. (ms)	SD (ms)
News reading	1	within-unit (no punct.)	240	4%	218	235	187
	2	within-unit (punct.)	275	76%	296	306	159
	3	end-of-unit	336	100%	779	815	293
Poetry reciting	1	within-unit (no punct.)	217	3%	136	185	143
	2	within-unit (punct.)	628	79%	265	321	242
	3	end-of-smaller-unit	757	44%	450	471	243
	4	end-of-unit	532	65%	727	800	388
	5	end-of-larger-unit	264	100%	1082	1144	397

Textual organization also influenced the duration of pauses (see Table 1 for raw durations, Figure 3 for logarithmic durations). In both genres, pauses at the end of the basic unit (sentence/distich) were larger than pauses within those units. Moreover, there seems to be a systematic relationship between the depth of textual cohesion (three levels in news reading and five levels in poetry reciting) and the durations of pauses, i.e., the looser the connection between the flanking words, the longer the pauses. Such an effect suggests a functional use of pause duration in conveying paragraph/stanza structure. See Section 3.2.1 for statistical evaluation.

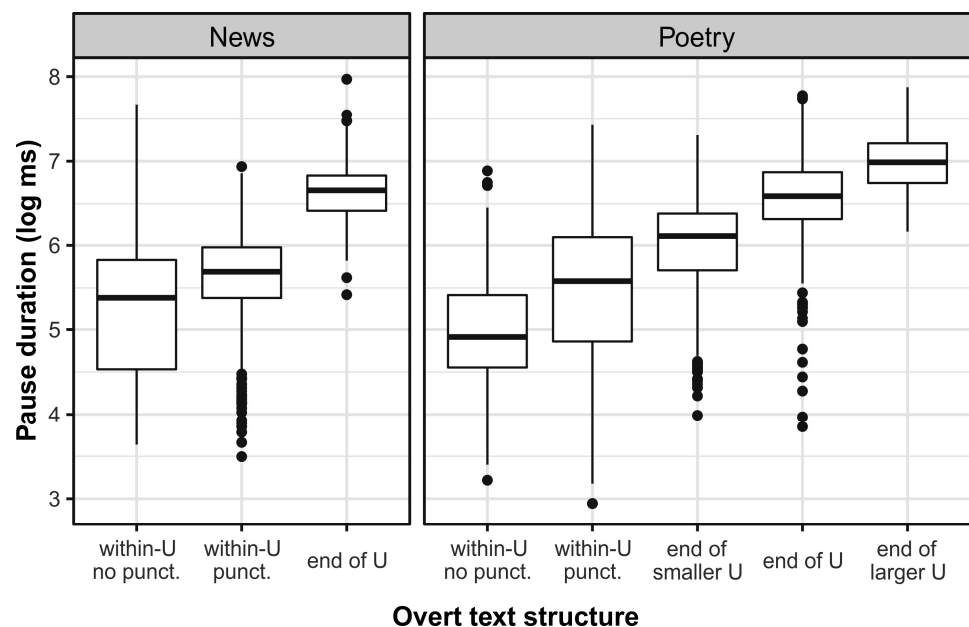


Figure 3. Boxplots of log pause duration in two genres as a function of overt text structure (U = unit; punct. = punctuation). For an explanation of levels, see Section 2.3.

3.1.3. Effect of Covert Text Structure

Given that higher units of overt text organization are predominantly associated with boundaries involving either loose syntactic relations or punctuation, only Level-1 pauses (*within-line/within-sentence without punctuation*) were considered for the grammatically informed analysis based on Franz et al. (2022). There were 458 realized pauses out of 3096 potential ones. The occurrence is summarized in Table 2. The number of potential pauses corresponding to each pause location is 24, that is, the number of speakers (locations where no speaker produced a pause are not counted as potential places). News reading and poetry reciting both seem to reflect the predicted pause adequacy: we found the lowest proportion of pauses (5%) in -II (grammatically forbidden) and -I (not recommended) contexts, while the highest proportion (33–41%) was bound to the linguistic context III (where the pause is possible, frequent, and usually salient). Table 2 also lists the average raw durations of pauses (for logarithmic values, see Figure 4). However, a similar correspondence between durations and the linguistic level is not clear from the data. See Section 3.2.2 for statistical evaluation.

Table 2. The occurrence (absolute and relative) of pauses and their duration as a function of the grammatically informed analysis (Franz et al. 2022) in two genres. Only pauses in within-unit non-punctuated contexts are analyzed. The relative occurrence denotes the percentage of realized to potential pauses in the given pause category context. For an explanation of levels, see Section 2.3.

Genre	Level	Pausing	Absolute Count	Relative Count	Median dur. (ms)	Mean dur. (ms)	SD (ms)
News reading	-II	Blocked	10	5%	118	132	51
	-I	Not recommended	15	5%	185	190	128
	I	Weaker break	30	16%	272	256	155
	II	Break	66	13%	202	245	274
	III	Stronger break	119	33%	231	238	141
Poetry reciting	-II	Blocked	5	4%	211	177	64
	-I	Not recommended	30	6%	194	238	236
	I	Weaker break	33	11%	126	193	177
	II	Break	91	19%	125	173	154
	III	Stronger break	59	41%	147	193	130

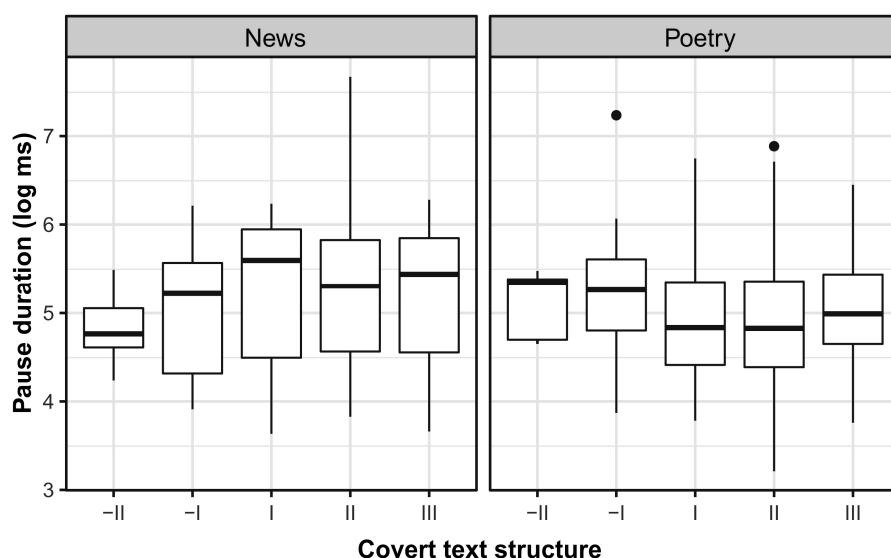


Figure 4. Boxplots of log pause duration in two genres as a function of the grammatically informed analysis according to Franz et al. (2022). For an explanation of levels, see Section 2.3.

3.1.4. Effect of Articulation Rate

Volín (2022) analyzed the same material as used here in terms of articulation rate (AR) and speech rate (SR) and computed mean values for each speaker and genre. Correlating the average AR values with the number of pauses per speaker and median pause duration per speaker yields the results summarized in Table 3. Although the coefficients for the first parameter (pause count) are all negative, i.e., there was a tendency for relatively faster speakers to produce fewer pauses (at least in news reading), the correlation did not reach significance. The second parameter (pause duration) differs between the two genres in sign (direction of the correlation), but only poetry reciting yielded significant values. With increasing speaker tempo, the median duration of pauses decreased. Speakers who were on the faster side of the tempo continuum thus produced shorter pauses while reciting poetry but not while reading the news.

The reader is encouraged to notice that the two tempo metrics (i.e., segments per second and syllables per second) produce almost equal results. The comparison is of interest because Czech is a language with complex syllabic structures, and various consonant clusters in both onsets and codas are legal and common. Our results suggest that despite this, the difference between the two metrics is not crucial for pausing.

Table 3. Pearson’s correlations between the speaker tempo and the frequency or duration of pauses per speaker in two genres. Coefficients (incl. 95% CIs) and their associated *p*-values are given.

Genre	Correlation	<i>r</i>	<i>p</i>
News reading	AR slb/s ~ N of pauses	−0.31 [−0.63; 0.11]	0.142
	AR seg/s ~ N of pauses	−0.23 [−0.58; 0.19]	0.284
Poetry reciting	AR slb/s ~ N of pauses	−0.08 [−0.47; 0.33]	0.701
	AR seg/s ~ N of pauses	−0.07 [−0.46; 0.34]	0.743
News reading	AR slb/s ~ median of pause duration	0.29 [−0.13; 0.62]	0.169
	AR seg/s ~ median of pause duration	0.29 [−0.12; 0.62]	0.164
Poetry reciting	AR slb/s ~ median of pause duration	−0.47 [−0.73; −0.08]	0.020 *
	AR seg/s ~ median of pause duration	−0.45 [−0.72; −0.05]	0.028 *

Note: * refers to a statistically significant correlation.

3.2. Statistical Evaluation

3.2.1. Overt Text Structure and Other Predictors

A Poisson regression model was constructed for the rate of pauses in the full dataset, in which the effects of genre and overt text organization were evaluated while controlling for the speaker’s average articulation rate. However, such a model did not converge so the tempo predictor had to be dropped. The resulting model thus predicts pause rate as a function of GENRE and OVERT TEXT STRUCTURE with SPEAKER and TEXT as random effects (intercept-only). The count variable involved the number of pauses (within a text) that a given speaker produced in the given category, offset to the number of potential pauses of that category (this varied depending on category and text). The interaction between GENRE and OVERT TEXT STRUCTURE was highly significant ($\chi^2(2) = 18.7, p < 0.001$) and is plotted in Figure 5. Post-hoc Tukey comparisons revealed that within news reading, all three levels of organization differed significantly ($p < 0.01$), while within poetry reciting, the differences were significant ($p < 0.001$) for all except for the within-unit punctuated and end-of-unit levels ($p = 0.114$). Across genres, the paired levels generated similar pause rates, except for the end-of-unit level which was associated with different rates ($p < 0.001$). Figure 5 suggests that the end-of-unit pause rate in news reading (i.e., at the end of a sentence) better resembles the end of a larger unit in poetry (i.e., that of the stanza and not that of the distich).

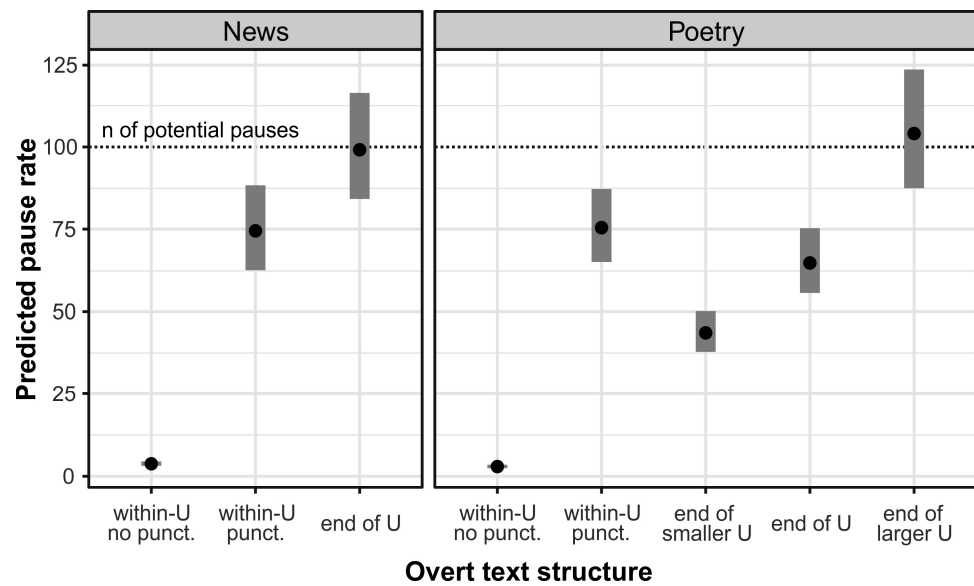


Figure 5. Predicted pause rate as a function of overt text structure (U = unit; punct. = punctuation). For an explanation of levels, see Section 2.3. The predictions of a Poisson regression model are back-transformed from log values to rates and relate to a sample of 100 potential pauses (dotted line).

Further, a linear regression model was fitted to the logarithmic duration of pauses. As observations are individual pauses, the structure of the model included GENRE, OVERT TEXT STRUCTURE, TEMPO (speaker’s mean AR in syllables/s), PRECEDING PROSODIC PHRASE LENGTH (in syllables), and FOLLOWING PROSODIC PHRASE LENGTH (in syllables) as fixed effects and SPEAKER and TEXT as random effects (with a varying by-speaker slope for GENRE). Of these, only OVERT TEXT STRUCTURE and PRECEDING PR. PHRASE LENGTH were significant predictors ($\chi^2(4) = 2303.3, p < 0.001$ and $\chi^2(1) = 15.2, p < 0.001$, respectively; see Figure 6). Tukey post-hoc tests confirmed all comparisons as highly significant ($p < 0.001$). Finally, none of the fixed effects had significant interactions with GENRE ($p > 0.05$).

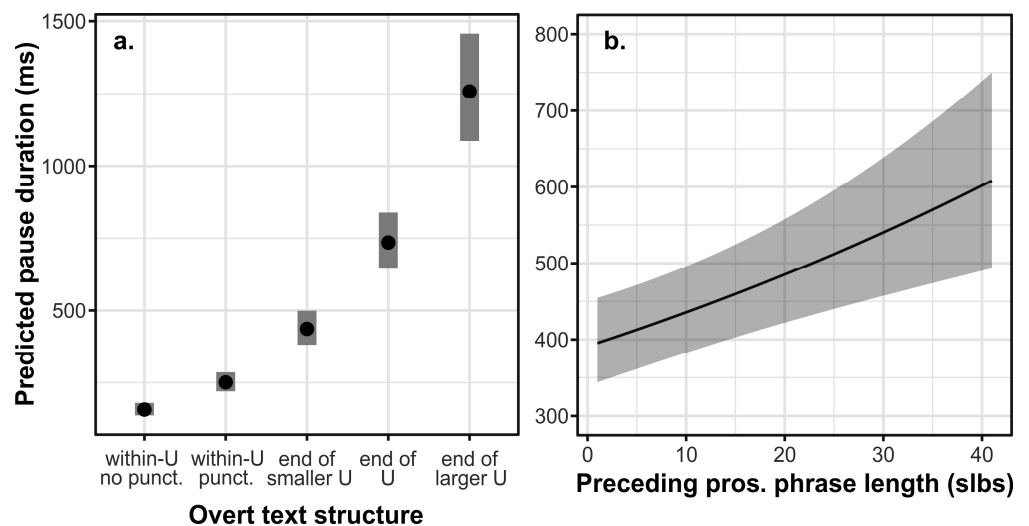


Figure 6. (a) Predicted pause duration as a function of overt text structure (U = unit; punct. = punctuation); (b) predicted pause duration as a function of preceding prosodic phrase length in syllables. For an explanation of levels, see Section 2.3.

3.2.2. Covert Text Structure

The analysis of linguistic structure according to Franz et al. (2022) was based on a subset of the data ($n = 458$) in which only pauses that occurred in within-unit contexts without interpunctuation were considered. A generalized Poisson regression model predicted the rate of pauses as a function of GENRE and COVERT TEXT STRUCTURE with TEXT as a random effect. The count variable involved the number of speakers that produced a pause at a specific location, offset to the number of potential pauses (i.e., 24). GENRE did not reach significance ($\chi^2(1) = 1.2, p = 0.265$) but COVERT TEXT STRUCTURE was a highly significant predictor ($\chi^2(4) = 172.4, p < 0.001$). The interaction between the two effects was not significant ($\chi^2(4) = 5.3, p = 0.261$). Figure 7 shows the non-interactive model's pause rate predictions averaged over genre. Post-hoc Tukey comparisons revealed three groups that differed significantly ($p < 0.001$) from one another: -II/-I formed the first group, differing from I/II, which also differed from III. In other words, there were significantly more pauses in the grammatically favored contexts I/II/III than in the contexts where a pause is grammatically incorrect or disfavored.

A linear regression model was fitted to the logarithmic duration of pauses in the same subset ($n = 458$). As observations are individual pauses, the structure of the model included GENRE and COVERT TEXT STRUCTURE as fixed effects and SPEAKER and TEXT as random effects (with a varying by-speaker slope for GENRE). Neither fixed effect was significant ($\chi^2(1) = 2.1, p = 0.150$ and $\chi^2(4) = 4.8, p = 0.310$, respectively). There was also no significant interaction between them ($\chi^2(4) = 7.4, p = 0.114$). Therefore, unlike the rate of non-punctuated within-unit pauses, their duration was not affected by the linguistic structure.

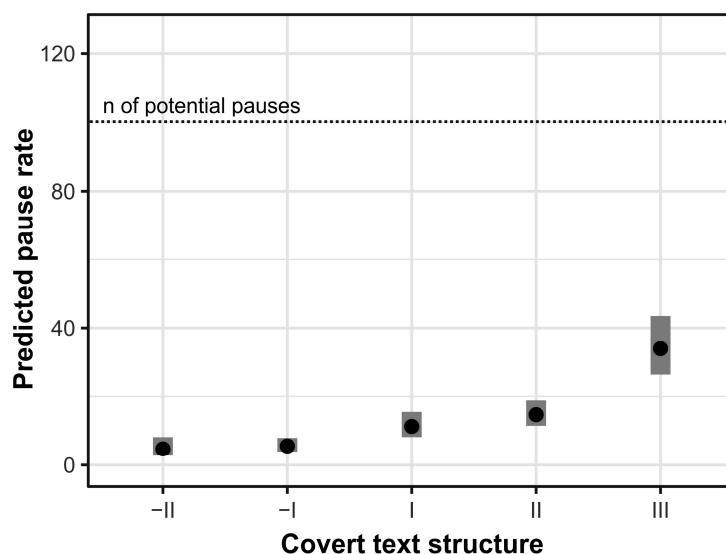


Figure 7. Predicted pause rate as a function of the grammatically informed analysis according to Franz et al. (2022). For an explanation of levels, see Section 2.3. The predictions of a Poisson regression model are back-transformed from log values to rates and relate to a sample of 100 potential pauses (dotted line).

4. Discussion

The current study provided a medium-scale analysis of pausal performance (>3000 pauses) in two genres of read speech in Czech. Although specific hypotheses were also tested, a significant contribution of the paper lies in the detailed description of pause occurrence and duration in relation to various aspects of the utterance structure in terms of overt text organization and covert within-unit linguistic constituency. Examining the variability in pausing at certain linguistically defined places in a spoken text is highly relevant yet often missing or simplified in the published literature.

4.1. Factors Affecting Pausing

There were several levels of **overt text structure** in the material. Our initial assumption was to equate the *sentence* with the *distich* as the main unit of organization. The rationale was that they are both salient units with clear syntactic boundaries usually characterized by conclusiveness and semantic unity. Moreover, they are in most cases separated with major punctuation marks (full stops, colons, semicolons, etc.). Pausing behavior at these boundaries was predicted to differ greatly from behavior within such units, which was confirmed by the results, provided that we account for the effect of punctuation (see below). However, the rate of pauses at sentence ends (100%) differed from the ends of distichs (65%) and was equal to the ends of stanzas (100%). As a result, there was a significant interaction between text organization and genre, which is likely to disappear if sentences and stanzas are considered as the main unit instead. In contrast, pause duration reflected the initial assumption closely. There was no significant interaction with the genre, and the suggested ordered scale of text organization correlated strongly with increasing pause duration. The two parameters—rate and duration—thus seem to be independent variables in pausing behavior. The main conclusion from the results here is that the pause rate and duration seem to be used systematically in conveying text structure.

Within sentences and verse lines, the key factor was whether words were divided with **punctuation** in the written form. How sentences are written on the page is a strong factor in constraining the oral production of speakers. In both genres, stretches of speech within lines or sentences were interrupted with a pause considerably more often when punctuation was present than when it was absent. This is fully in line with previous findings. For instance, [Hieke et al. \(1983\)](#) also confirmed a strong relation between pause use and verse lines and punctuation. More recently, [Werner et al. \(2022\)](#) showed that for read speech of a non-poetic genre, pauses frequently cooccurred with punctuation, which was a good determiner of pause location.

However, the visual aspects of the text and its division into units are not the sole factors affecting the occurrence and duration of pauses. We postulated five levels of **covert text structure** as well by following the innovative and grammatically informed approach of [Franz et al. \(2022\)](#). Crucially, these predictions about the organization of words reflect word-class information, syntax, the size of the syntactic constituents, and the speech rhythm. In some locations, pauses are not allowed or recommended, while in others, varying degrees of the propensity for pausing are expected. We applied such analysis to within-unit pauses not corresponding to punctuation. Again, the results showed the independent behavior of pause rate and pause duration. Whereas the duration of pauses was not affected by the assumed scale of linguistic cohesiveness, some of the categories yielded different patterns of pause rate. The two deprecated levels (-II and -I) were most resistant to pausing, while the three levels in which pausing is predicted with increasing probability (I, II, III) showed significantly higher rates of pausing (and III the highest of all). This was true of both genres.

In summary, overt text organization influenced both the pause rate and pause duration, while covert linguistic structure influenced only the pause rate. Apparently, it is not only syntax in the narrow sense of the word but also the size of the constituents or rhythm constraints that are implicated. Modeling the delicate interplay of all the factors will require much more research. It is becoming quite clear that syntax is only one of the determinants of pause location (cf. [Ruder and Jensen 1972](#); [Rochester 1973](#); [Watson and Gibson 2004](#); [Carlson 2009](#)). For example, the length of the preceding prosodic phrase in syllables was a significant predictor in our model: longer phrases were followed by longer pauses. [Zvonik and Cummins \(2003\)](#) also investigated the effect of the surrounding phrases on pausing; however, their results support influence from both directions, that is, of the preceding and following phrases. They highlight that short phrases favor short pauses, a pattern discovered here only for the preceding phrase.

Another determiner of pause location and duration, which is (unfortunately) often ignored and considered beyond the purview of linguistic attention, is **speech tempo**. This effect was investigated by [Werner et al. \(2022\)](#), who focused on speakers reading the

same text at five intended speech rates. Among others, they found fewer and shorter pauses in speech produced at faster tempos. However, speaking slower or faster inherently (habitually) may differ from manipulated tempo conditions, as the authors admit (for instance, because of differences in the interpretation of the intended rates). In that respect, it should be interesting to examine our material which provided a range of articulation rates across the speakers, although it is potentially less conclusive. Correlating the speaker means of AR with the number of produced pauses on the one hand and the median duration of pauses on the other yielded disparate results. We found mostly weak and insignificant correlations, with the exception of pause duration in poetry. The pause rate was thus not correlated with habitual tempo in either genre, and inherently faster speakers produced shorter pauses while reciting poetry but not in the news-reading condition. The effect seems to be limited to speaker averages, though. When the speaker tempo was correlated with the duration of individual pauses in the statistical model controlling for other factors, the effect disappeared.

4.2. Genre Differences

The discussion of genre aspects has been present throughout the whole section. The main difference is that speakers produced the majority of pauses in poetry reciting rather than in news reading (there were on average 100 pauses per speaker in the former and 36 pauses per speaker in the latter genre). This is likely connected to the fact that there is an additional level of overt text organization in poetry, namely the line and the stanza. Using the words of [Hieke et al. \(1983\)](#), “poetic format imposes a specific sequential structure different from the prose of a political text” (p. 209), or a news bulletin. Alternating the ends of lines, distichs, and stanzas, which were systematically associated with distinct pause durations, provides another level of rhythmicity—and speech rhythm itself is another source of pause production. If there is a verse line consisting of two independent constituents, it increases the probability of separation in the same position of the following verse line, even if it means breaking a single constituent that, on its own, would not be divided. For instance, the line *Trochu se vraždilo, trochu se kradlo* (there were some murders, there were some thefts) clearly falls apart in the middle. The next line *Pereme, pereme špinavé prádlo* (we wash, we wash dirty laundry) should not have a split between the verb and the direct object. Yet, the speakers often used a pause analogical to the preceding verse line, even though this division is ‘grammatically’ highly marked.

In contrast to the pause rate, there was only a marginal difference in the pause duration between the two genres. We expected that pauses would be longer in poetry reciting, but this prediction was not borne out. However, it should be noted that speakers talked at different tempos in the two genres as well, with poetry reciting being associated with slower speech (4.6 syllables/s compared to 6.1 syllables/s in news reading). Therefore, similar pause durations across the genres are not equivalent perceptually, as the same amount of pause time could be filled with more speech segments in news reading.

4.3. Limitations and Directions for Future Research

Not surprisingly, our study has several limitations which can be addressed in future research. First and foremost, by focusing on read materials and poetic speech, we do not have much to say about pausing in spontaneous speech. It is quite likely that there will be many important differences, but they can hardly be speculated about without a thorough empirical investigation. The imperative is to continue with the research in the spontaneous domain and test whether the methodology used can be applied as well. It might be the case that further considerations such as contextual dependencies or degree of givenness will have to be applied. These factors, however, are explicitly rejected by [Franz et al. \(2022\)](#) and would require a new extensive study. Moreover, pausing in spontaneous speech can relate to the parameters examined here but also to the types of pauses (e.g., breath noises). The phonetic characteristics of pauses should be described in much more detail. Fortunately, this is being done (e.g., [Trouvain et al. 2016](#); [Werner et al. 2022](#)). Another

issue concerns speaker-specific strategies for pausing. First, some of our speakers produced twice as many pauses as others in the same text, yet the properties and distribution of their pauses require detailed contextual analyses. It is also possible that speakers behave differently in the two genres, which could bias the results to some extent. However, the interaction is either captured in the statistical model by specifying varying slopes of the genre effect, or the genre differences are not significant anyway. Second, the current paper deals with pauses, but it would be immensely useful to examine pre-pausal prosodic events such as final lengthening (see [van Donzel and Koopmans-van Beinum 1996](#)), changes in voice quality (increased breathiness), or amplitude lowering. This is a clear objective for follow-up research. What has been left out entirely is the question of pause perception. This is clearly beyond the scope of the study but merits careful attention.

5. Conclusions

The special issue “Pauses in Speech” brings pausing to the spotlight of linguistic inquiry. The current article contributes to the research on within-language variation in terms of different genres and speech styles, comparing the characteristics of pauses in the frequently studied genre of news reading with that of poetry reciting, which is examined rarely if at all. The two genres showed differences in the rate of pauses, but the effects of overt and covert text organization were related to genre differences in a complex way. The sentence level in news reading corresponded to different structures (the distich or the stanza) in poetry reciting depending on whether the occurrence or the duration of pauses was considered. Moreover, the grammatical organization of smaller units influenced the occurrence of pauses but not their duration. The results highlight the fact that speakers use pauses systematically and thus intentionally in read speech—a sufficient argument for incorporating the study of pausing into the description of language in general and prosody in particular.

Supplementary Materials: The following supporting information can be downloaded at: <https://osf.io/9y7et/>. Dataset D1: spreadsheet with the data used in this study.

Author Contributions: Conceptualization, P.Š. and J.V.; methodology, P.Š. and J.V.; software, P.Š.; validation, P.Š. and J.V.; formal analysis, P.Š.; investigation, J.V.; resources, J.V.; data curation, J.V. and P.Š.; writing—original draft preparation, P.Š.; writing—review and editing, J.V.; visualization, P.Š.; supervision, J.V.; project administration, J.V.; funding acquisition, J.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by GAČR (Czech Science Foundation) grant number 21-14758S.

Institutional Review Board Statement: The study did not require ethical approval. Ethical review and approval were waived for this study as it involved no experimenting with humans. The presented analyses were carried out on an already existing corpus of recordings that were made in the past along the lines of seminar recording practice at the Faculty of Arts, Charles University. The practice included informed consent from the recorded participants and reassurance by the institution that the recordings were anonymous and strictly confidential, meant to be used only for research purposes by the institution.

Informed Consent Statement: Informed consent was obtained from all subjects who were recorded during the corpus collection.

Data Availability Statement: The data presented in this study are available in Supplementary Materials. Recordings cannot be shared due to privacy reasons.

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

References

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Boersma, Paul, and David Weenink. 2022. *Praat—Doing Phonetics by Computer (Version 6.2)*. [Computer Program]. Available online: www.praat.org (accessed on 8 April 2022).
- Campione, Estelle, and Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 199–202. Available online: https://www.isca-speech.org/archive/speechprosody_2002/campione02_speechprosody.html (accessed on 3 January 2023).
- Carlson, Katy. 2009. How prosody influences sentence comprehension. *Language and Linguistics Compass* 3: 1188–200. [CrossRef]
- Carlson, Rolf, Julia Hirschberg, and Marc Swerts. 2005. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication* 46: 326–33. [CrossRef]
- Cooper, William E., and Jeanne Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge: Harvard University Press.
- Elmers, Mikey, Raphael Werner, Beeke Muhlack, Bernd Möbius, and Jürgen Trouvain. 2021. Evaluating the effect of pauses on number recollection in synthesized speech. In *Elektronische Sprachsignalverarbeitung 2021/32*. Dresden: TUD Press, pp. 289–95.
- Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101: 3728–40. [CrossRef] [PubMed]
- Franz, Isabelle, Christine A. Knoop, Gerrit Kentner, Sascha Rothbart, Vanessa Kegel, Julia Vasileva, Sanja Methner, Mathias Scharinger, and Winfried Menninghaus. 2022. Prosodic Phrasing and Syllable Prominence in Spoken Prose. A Validated Coding Manual. *OSF Preprints*. Available online: <https://doi.org/10.31219/osf.io/h4sd5> (accessed on 20 June 2022).
- Frazier, Lyn, Katy Carlson, and Charles Clifton Jr. 2006. Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10: 244–49. [CrossRef]
- Ghitza, Oded, Anne-Lise Giraud, and David Poeppel. 2013. Neuronal oscillations and speech perception: Critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience* 6: 340. [CrossRef] [PubMed]
- Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.
- Goldman-Eisler, Frieda. 1972. Pauses, clauses, sentences. *Language and Speech* 15: 103–13. [CrossRef]
- Hieke, Adolf E., Sabine Kowal, and Daniel C. O’Connell. 1983. The trouble with “articulatory” pauses. *Language and Speech* 26: 203–14. [CrossRef]
- Kjelgaard, Margaret M., and Shari R. Speer. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40: 153–94. [CrossRef]
- Lehiste, Ilse. 1973. Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America* 54: 1228–34. [CrossRef]
- Lenth, Russell. 2022. Emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.8.2. Available online: <https://CRAN.R-project.org/package=emmeans> (accessed on 10 October 2022).
- Machač, Pavel, and Radek Skarnitzl. 2009. *Principles of Phonetic Segmentation*. Praha: EPOCH.
- Maclay, Howard, and Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15: 19–44. [CrossRef]
- Männel, Claudia, and Angela D. Friederici. 2016. Neural correlates of prosodic boundary perception in German preschoolers: If pause is present, pitch can go. *Brain Research* 1632: 27–33. [CrossRef] [PubMed]
- Martin, Philippe. 2015. *Structure of Spoken Language*. Cambridge: Cambridge University Press.
- Niebuhr, Oliver, and Kerstin Fischer. 2019. Do not hesitate!—Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance. Paper presented at Interspeech 2019, Graz, Austria, September 15–19; pp. 544–48. [CrossRef]
- Pannekamp, Ann, Ulrike Toepel, Kai Alter, Anja Hahne, and Angela D. Friederici. 2005. Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience* 17: 407–21. [CrossRef]
- Paschen, Ludger, Susanne Fuchs, and Frank Seifart. 2022. Final lengthening and vowel length in 25 languages. *Journal of Phonetics* 94: 101179. [CrossRef]
- Peelle, Jonathan E., and Matthew H. Davis. 2012. Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3: 320. [CrossRef] [PubMed]
- Petrone, Caterina, Hubert Truckenbrodt, Caroline Wellmann, Julia Holzgreffe-Lang, Isabell Wartenburger, and Barbara Höhle. 2017. Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61: 71–92. [CrossRef]
- Pollák, P., J. Volín, and R. Skarnitzl. 2007. HMM-based phonetic segmentation in Praat environment. Paper presented at XIIth Conference on Speech and Computer—SPECOM 2007, Moscow, Russia, October 15–18; pp. 537–41.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 10 October 2022).
- Rochester, Sherry R. 1973. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research* 2: 51–81. [CrossRef]
- Ruder, Kenneth F., and Paul J. Jensen. 1972. Fluent and hesitation pauses as a function of syntactic complexity. *Journal of Speech and Hearing Research* 15: 49–58. [CrossRef]
- Trouvain, Jürgen, Camille Fauth, and Bernd Möbius. 2016. Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech. Paper presented at Speech Prosody 2016, Boston, MA, USA, May 31–June 3; pp. 31–35. [CrossRef]

- Uchanski, Rosalie M., Sunkyung S. Choi, Louis D. Braid, Charlotte M. Reed, and Nathaniel I. Durlach. 1996. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research* 39: 494–509. [\[CrossRef\]](#)
- van Donzel, Monique E., and Florian J. Koopmans-van Beinum. 1996. Pausing strategies in discourse in Dutch. Paper presented at Fourth International Conference on Spoken Language Processing, Philadelphia, PA, USA, October 3–6; vol. 2, pp. 1029–32. [\[CrossRef\]](#)
- Volín, Jan. 2022. Variation in speech tempo and its relationship to prosodic boundary occurrence in two speech genres. *Acta Universitatis Carolinae—Philologica* 1: 65–81.
- Wagner, Michael, and Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25: 905–45. [\[CrossRef\]](#) [\[PubMed\]](#)
- Watson, Duane, and Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19: 713–55. [\[CrossRef\]](#)
- Watson, Duane, and Edward Gibson. 2005. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 59: 279–300. [\[CrossRef\]](#)
- Werner, Raphael, Jürgen Trouvain, and Bernd Möbius. 2022. Optionality and variability of speech pauses in read speech across languages and rates. Paper presented at Speech Prosody 2022, Lisbon, Portugal, May 23–26; pp. 312–16. [\[CrossRef\]](#)
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, and et al. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4: 1686. [\[CrossRef\]](#)
- Zellner, Brigitte. 1994. Pauses and the temporal structure of speech. In *Fundamentals of Speech Synthesis and Speech Recognition*. Edited by Eric Keller. Chichester: John Wiley & Sons, pp. 41–62.
- Zvonik, Elena, and Fred Cummins. 2003. The effect of surrounding phrase lengths on pause duration. Paper presented at Eurospeech 2003, Geneva, Switzerland, September 1–4; pp. 777–80. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.