CHANGES IN SEGMENTAL TIMING IN SLOW AND FAST METRONOME-SYNCHRONIZED SPEECH

Eliška Churaňová, Pavel Šturm, Lenka Weingartová

Institute of Phonetics, Charles University in Prague, Czech Republic
eliska.churanova@ff.cuni.cz

ABSTRACT

Temporal characteristics of individual segments are affected by various global and local factors, such as tempo, syllable structure or position of a segment within the word. This has been shown for many languages, but since the general trends may be language specific, the present study investigates such effects for Czech. The material included 28 natural phonotactically complex words spoken by 24 subjects in synchrony with metronome beats at two different tempi. The lexical units were carefully selected with regard to their phonological structure. The results demonstrate interactions of tempo with the effects of final lengthening, segment type and onset complexity, suggesting non-linearity in the temporal changes.

Keywords: segmental duration, speaking rate, syllable structure, final lengthening, Czech

1. INTRODUCTION

Durational features of consonants and vowels in speech are influenced by a variety of interacting factors in a given communicative situation. Individual phones have been demonstrated to possess intrinsic temporal properties based on physiological and/or language-specific principles, e.g., voiced obstruents or high vowels being mostly shorter in duration than voiceless obstruents or low vowels [2], [18], [30]. Extrinsic factors affecting phone duration include for instance pitch accent, word stress, syllable structure, position within a word or prosodic phrase [18], [31]. Evidence also suggests an influence of the lexical status of the word itself: content words tend to be articulated at a slower rate than function words [24]. Some results suggest that this effect may play a role in Czech as well [35]. Moreover, speaking style influences the durational properties of both consonants and vowels as well [12], [25].

1.1. Tempo

Tempo, measured as speech rate (SR) or articulation rate (AR), is one of the most evident agents influencing phone duration: at a higher AR segments manifest shorter duration [18]. Nevertheless, such changes are non-linear, with some segments being affected more than others. For example, in American English the temporal features of long vowels and glides vary greatly with changes in SR, whereas plosives and affricates are almost resistant to these shifts [4]. Similar results were obtained for Czech by Kaiser [15], who discovered that shortening at higher SR was greater in phonologically long vowels than in their short counterparts, or more recently by Machač [21], who verified the temporal stability of Czech plosives at varying ARs.

1.2. Prominence and position within prosodic units

Duration of segments can also be affected by prominence manifestation on the word or phrase level. Generally, stressed/accented syllables tend to be longer than unstressed/unaccented ones. This fact is supported by research on various languages [6], [8], [11], [27], [28], [29]. However, in Czech, where word stress is fixed on the first syllable of the word or stress group, it is not manifested by increased duration of the stressed syllable by default [13], [14]. A recent study demonstrated the relationship between phone duration and prominence on the sentence level, though: words with phrasal/contrastive prominence appeared to be longer than others [33].

Another salient factor that affects local temporal features is the position of the phone within a higher prosodic unit (word, stress-group or phrase). Some of these phenomena are often considered universal, such as pre-boundary or final lengthening [18], [19], [31]. Although observed in many languages, its exact extent can be language- [9] or even speaker-specific [36]. Dankovičová [7] studied rate variation in fluent Czech and concluded that final lengthening manifests itself in the domain of the prosodic phrase, but some inter-speaker variability in the extent of lengthening appeared in her material as well. In some languages phrase- and word-initial lengthening of consonants or vowels occurs [5], [10] [16].

1.3. Temporal compensation

Temporal compensation between a vowel and a neighbouring consonant has been observed for

several languages. Frequently, vowels can be shorter in duration if followed by voiceless rather than voiced consonants [18], [20], [26]. In Czech, such compensatory relationships between segments have not been elaborated in detail. However, according to [7] longer Czech words tend to be pronounced faster than shorter units (the more syllables in a word, the shorter the duration of its segments), which might imply that syllable-timing of Czech is not as strong a principle as is commonly assumed.

1.4. Phonetic environment

Durational variability of phones in consonant clusters seems to be caused among others by physiological constraints. The extent of variation can depend for example on the distance between articulatory targets of individual phones or the disposition of a particular sequence for an overlap of articulatory gestures [18]. Klatt [17] examined durational effects induced by phonetic environment and found consonants in clusters to be generally shorter than in CV syllables, but the relationship was not linear for all segment classes.

Given these findings, several hypotheses can be formulated for the experiment. The subjects are expected to produce shorter segmental durations at the faster tempo. The extent of shortening should depend at least on some of the following factors: (1) type of segment [18]; (2) occurrence in a consonant cluster [17]; (3) position within the word [23]. Since Czech temporal structure has not been investigated thoroughly, the aim of the current study is to examine to what extent the general principles observed for other languages apply to Czech as well.

2. METHOD

2.1. Material

The present study used material from [34] containing speech synchronized with metronome pulses in two tempi. 24 native Czech speakers were asked to repeatedly synchronize the first (stressed) syllable of a target word displayed on the screen with metronome beats. They pronounced each word eight times. The interval between beats was 857 ms for the slow tempo (70 beats per minute) and 667 ms for the fast tempo (90 bpm). These rates thus approximated a SR of 4 and 6 syllables per second. Although the material was obtained in laboratory conditions, the speakers were asked to pronounce the words in a natural way and avoid reciting or chanting (see [34] for details).

In total, 28 one-, two- and three-syllable real Czech words with controlled phonological structure were selected for analyses. Each word constitutes a stress-group, with stress on the first syllable. The aim was to obtain pairs or triplets of words with a phonologically comparable structure so that the influence of phonetic and phonological factors could be studied. The complete list of target items was as follows: *mim, my, stál, sál, klus, kus, deka, děkan, těká, těkám, těkáš, štěká, štěkám, štěkáš, stékáš, stékám, létá, slétá, slétám, slétáš, splétá, splétám, splétáš, vůbec, názor, nasadit, podávat, podíváš*.

The material contained 5,254 realisations of 25 Czech phones (9 Vs, 16 Cs), but the analyses were focused primarily but not exclusively on the most frequent coronal consonants /s/, /t/, /l/ ([1]; see [32] for erratic durational behaviour of American English coronals). As columns in Table 1 indicate, the onset of the first syllable consisted of either a single C (sonorant or obstruent), a CC cluster (two obstruents or obstruent + sonorant), or a CCC cluster (two obstruents + sonorant). All intervocalic Cs were singletons. Rows in the table correspond to the final segment in the word, which was either a vowel (open syllables), or a single coda consonant (sonorant or voiceless obstruent).

Table 1: Examples of analysed words according to the structure of the first syllable onset and last syllable coda.

coda	**onset of the 1st syllable**				
	C	**CC**	**CC**	**CC**	**CCC**
	/c/	/ʃc/	/st/	/sl/	/spl/
no coda	cɛkaː	-	-	slɛːtaː	splɛːtaː
nasal	cɛkaːm	ʃcɛkaːm	stɛːkaːm	slɛːtaːm	splɛːtaːm
obstruent	cɛkaːʃ	ʃcɛkaːʃ	stɛːkaːʃ	slɛːtaːʃ	splɛːtaːʃ

2.2. Extraction of data and statistical analyses

Recordings of metronome-synchronized speech were processed in Praat [3] and manually annotated in accordance with [22]. Durations of all segments were measured. For methodological details, see [34].

Repeated measures ANOVAs were used with TEMPO (slow vs. fast) as the repeated measures factor. Tukey HSD post-hoc tests evaluated the differences between individual levels of the factors (SEGMENT CLASS: short V, long V, voiceless plosive, voiced plosive, voiceless fricative, nasal C, lateral C); POSITION WITHIN WORD: onset C1, onset C2, onset C3, intervocalic C, final C, first V, second V); ONSET COMPLEXITY: single C, CC cluster, CCC cluster). Occasionally, independent measures ANOVAs and t-tests were used when items were analysed only in one of the tempi.

3. RESULTS

The effect of TEMPO induced by the metronome was highly significant ($F(1, 2031) = 1315.6$, $p < 0.001$), with segments (both Vs and Cs) in the slow tempo being on average 1.2 times longer (20 ms) than in the fast tempo. However, there was a strong interaction with SEGMENT CLASS ($F(6, 2025) = 69.4$, $p < 0.001$) and POSITION ($F(6, 2025) = 89.7$, $p < 0.001$). Specifically, there was no significant slow-fast difference for laterals, while the remaining speech sounds showed significant differences (Tab. 2). These were largest for long Vs, voiceless fricatives and nasals, which can be related to their position within words, since the three segment classes also appeared in the word-final position associated with greatest differences between the two tempi.

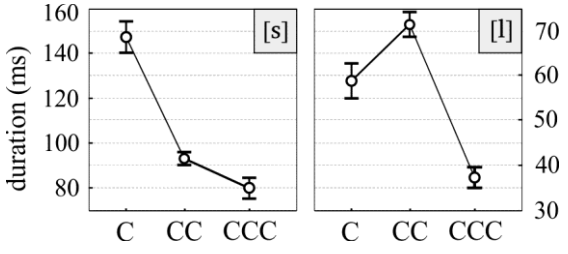
Table 2: Phone duration (ms) in slow and fast tempi; the difference in ms (*** = $p < 0.001$, * = $p < 0.1$) and ratio.

	V:	V	[l]	[n m]	[s ʃ]	[t c k]	[d ɟ]
slow	199.8	87.5	58.3	105.9	119.3	113.1	81.9
fast	167.7	78.2	53.1	85.1	93	102.2	70.7
diff.	32***	9***	5	21***	26***	11***	11*
ratio	1.19	1.12	1.10	1.24	1.28	1.11	1.16

3.1. Onset properties

The duration of [s] in syllable onsets of increasing complexity (/s/; /sl/ and /st/; /spl/) was compared, revealing highly significant differences ($F(2, 198) = 163.6$, $p < 0.001$) but no interaction with TEMPO ($p = 0.76$). Figure 1 (left) captures the main effect of ONSET COMPLEXITY, and is especially illustrative as regards the characteristics of the alveolar fricative in simple vs. complex onsets. This was confirmed by another word pair, /saːl/ vs. /staːl/, which yielded a 40-ms difference between [s] in the C and CC onsets ($t(46) = 6.77$, $p < 0.001$).

Figure 1: Duration of [s] and [l] in C, CC and CCC initial onsets.



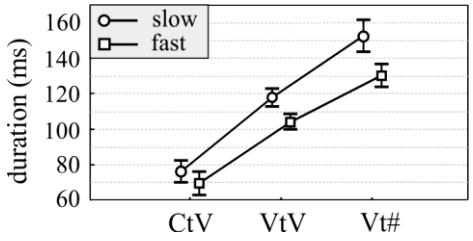
Second, we focused on [l] in C, CC and CCC initial clusters, i.e. in words starting with /l/, /sl/ and /spl/. There was a clear effect of ONSET COMPLEXITY ($F(2, 153) = 279.8$, $p < 0.001$) but again no significant interaction with TEMPO. Surprisingly, [l] was longer in the CC cluster than in the C condition, but [l] in the CCC proved to be shortest, as expected (Fig. 1b).

Finally, a third comparison concerns the onsets /st/ and /ʃc/ differing in the segmental content of a single cluster type. There was no temporal difference between [s] and [ʃ] in these clusters but the interaction of SEGMENT with TEMPO proved significant ($F(1, 159) = 4.0$, $p < 0.05$). Specifically, [s] tended to be longer than [ʃ] in the slow but not in the fast tempo. In contrast, there was no significant interaction of the main effects for the plosive pair, with [c] being consistently longer than [t] in both tempi ($F(1, 90) = 10.08$, $p < 0.01$ for SEGMENT).

3.2. Position of consonants within the word

The comparison of [t] in different POSITIONS within the word – in an initial cluster, medially between Vs and finally after a V – points to an effect of cluster compression ([t] was 35 ms shorter in onset /st/ than intervocalically; $p < 0.001$) and final lengthening ([t] was 30 ms longer word-finally than intervocalically; $p < 0.001$). Moreover, there was a highly significant interaction with TEMPO ($F(2, 222) = 5.1$, $p < 0.01$). As Figure 2 shows, onset [t] did not behave differently in the two tempi, but there was a highly significant difference ($p < 0.001$) of 12 and 22 ms for the intervocalic and final positions, respectively.

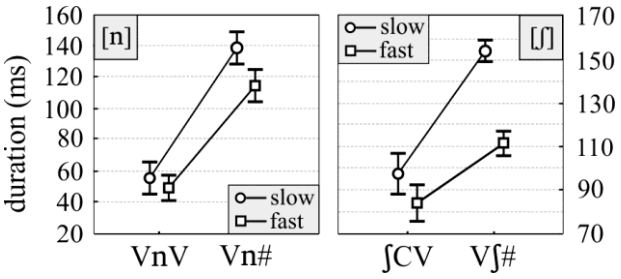
Figure 2: Duration of [t] in the onset (CtV) and in intervocalic (VtV) and word-final position (Vt#) in two tempi.



The effect of POSITION was confirmed by [n] on the one hand (Fig. 3a), which was much longer word-finally than intervocalically, and by [ʃ] on the other (Fig. 3b), which was longer word-finally than initially in a CC cluster. In both cases, there was a significant interaction with TEMPO ($F(1, 43) = 6.6$, $p < 0.05$ for [n], $F(1, 156) = 41.6$, $p < 0.001$ for [ʃ]), leading to greater differences in the slow as opposed to the fast tempo.

As regards the intervocalic position, the voiceless plosives [t] and [k] were quite similar, but the nasal [n] was in comparison more than 2.3 times shorter (by 50 ms). The effect of SEGMENT was highly significant ($F(2, 383) = 137.0$, $p < 0.001$), but not in interaction with TEMPO.

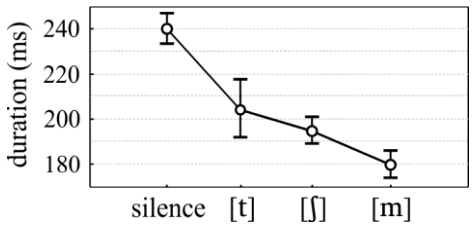
Figure 3: Duration of **a.** [n] in intervocalic (VnV) and word-final position (Vn#) and **b.** [ʃ] in initial (ʃCV) and word-final position (Vʃ#) in two tempi.



3.3. Effect of tempo on vowels

Both the stressed (V1) and unstressed (V2) vowel position showed a highly significant interaction of SEGMENT CLASS (short vs. long vowel) and TEMPO ($F(1, 384) = 49.0$, $p < 0.001$ for V1; $F(1, 384) = 15.1$, $p < 0.001$ for V2), with short vowels being more resistant to change under shifts in speaking rate than long vowels. V1 was represented by /ɛ ɛː/, while V2 by /a aː/. The duration of [aː] was influenced (but without an interaction with TEMPO) by the presence or absence of a final consonant ($F(3, 337) = 78.8$, $p < 0.001$), with substantial (word-final) lengthening in the absence of a coda C (Fig. 4).

Figure 4: Duration of unstressed [aː] before silence and before a coda consonant.



3.4. Word length

The length of the word measured in syllables seems to contribute to the extent of boundary lengthening. Final [m] was significantly longer in monosyllabic words compared to disyllables ($t(141) = -12.49$, $p < 0.001$), while final [ʃ] had an insignificant tendency to be longer in 2-slb than in 3-slb words ($t(135) = -1.21$, $p > 0.05$). An analogous effect was found for initial segments: a single initial [s] in a monosyllable was longer than the same consonant in a disyllabic word ($t(46) = 4.26$, $p < 0.001$), and the results were replicated for [s] in /st/ clusters in mono- vs. disyllabic words ($t(69) = -10.24$, $p < 0.001$). Similarly, initial [n] was longer in a disyllable than in a trisyllable ($t(46) = 3.61$, $p < 0.001$).

4. DISCUSSION

The main hypothesis of the study was that temporal changes induced by different speech rates would be non-linear. This was confirmed on several levels. In accordance with hypothesis (1) (see Introduction), segment class was one of the main predictors of the size of the shift between the slow and fast tempi. The clearest difference was found in long vowels, nasals [n, m] and sibilants [s, ʃ], while the lateral [l] seemed to be resistant to tempo-induced change. As regards hypothesis (2), there was no interaction of cluster compression with tempo. The segments [s], [t] and to some extent [l] were significantly shorter in clusters (see [17]), but this effect had similar magnitude in both tempi. Such a result might be expected because of the articulatory or other constraints induced by the cluster. Finally, position of the segment within the word (hypothesis (3)) significantly contributed to the extent of temporal shifts between the slow and fast conditions mostly in the form of word-final lengthening: the difference in phone duration between the two tempi was greater in the final position than in other positions.

It should be noted that average segmental durations were much longer than expected from fluent speech, even in the fast tempo (on average 1.4 times longer compared with [37]). This may be a consequence of the experimental task (metronome synchronization vs. reading or natural conversation). Interestingly, [c] was the only speech sound where there was a negligible difference in duration between this study and [37]. Also, the effect of tempo on [c] was not significant – this implies a certain resistance to both change in tempo and in speaking style.

The surprising behaviour of [l] in CC clusters (see Fig. 1) could be explained by labelling decisions, as the initial boundary of [l] in /sl/ sequences was placed already at the onset of the lateral noise associated with its articulation. However, this seems to be contradicted by the fact that [s] was still longer in /sl/ than in /st/ in both tempi.

In sum, the temporal characteristics observed in other languages on the segmental level appear to be applicable to Czech as well. The study nevertheless demonstrated that timing factors should be studied in a detailed way, as even individual segments differ in their capacity to compress or extend. Ultimately, the results ought to be verified on natural speech.

5. ACKNOWLEDGEMENTS

This output was created within the project FF_VG_2015_015 solved at Charles University in Prague from the Specific university research in 2015.

6. REFERENCES

[1] Bartoň, T., Cvrček, V., Čermák, F., Jelínek, T., Petkevič, V. 2009. *Statistiky češtiny.* Prague: Nakladatelství Lidové noviny.

[2] Beckman, M. E. 1986. Intensity, duration and loudness. In: *Stress and Non-Stress Accent.* Dordrecht: Foris, 133–144.

[3] Boersma, P., Weenink, D. 2014. *Praat: doing phonetics by computer* (Version 5.4.04.) http://www.praat.org.

[4] Brønsted, T., Madsen, J. P. 1997. Analysis of speaking rate variation in stress-timed languages. *Proc. Eurospeech '97* Rhodes, Greece, 481–484.

[5] Byrd, D., Saltzman, E. 1998. Intragestural dynamics of multiple phrasal boundaries. *J. Phon.* 26, 173–199.

[6] Crystal, T. H., House, A. S. 1988. Segmental durations in connected speech signals: Syllabic stress. *J. Acoust. Soc. Am*. 83, 1574–1585.

[7] Dankovičová, J. 1997. The domain of articulation rate in Czech. *J. Phon.* 25, 287–312.

[8] Fant, G., Kruckenberg, A., Nord, L. 1991. Durational correlates of stress in Swedish, French and English. *J. Phon.* 19, 351–365.

[9] Fletcher J. 2010. The prosody of speech: timing and rhythm. In: W. Hardcastle, J. Laver, F. Gibbon (eds.), *The Handbook of Phonetic Sciences*. United Kingdom: Wiley-Blackwell Publishing, 523–602.

[10] Fougéron, C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *J. Phon.* 29, 109–136.

[11] Heldner, M., Strangert, E. 2001. Temporal effects of focus in Swedish. *J. Phon.* 29, 329–361.

[12] Chen, F. R. 1980. *Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level.* Master's thesis. Cambridge, MA: MIT.

[13] Chlumský, J. 1928. *Česká kvantita, melodie a přízvuk.* Prague: Czech Academy of Sciences and Arts.

[14] Janota, P., Palková, Z. 1974. Auditory evaluation of stress under the influence of context. *AUC Philologica 2/1974, Phonetica Pragensia* 4, 29–59.

[15] Kaiser, L. 1964. Phonetic similarity apart from linguistic affinity. *Zeitschrift für Phonetik* 17, 243–249.

[16] Keating, P. A., Cho, T., Fougéron, C., Hsu, C. 2003. Domain-initial strengthening in four languages. In: J. Local, R. Ogden, R. Temple (eds.), *Papers in laboratory phonology 6*. Cambridge, UK: Cambridge University Press, 145–163.

[17] Klatt, D. H. 1973. Durational characteristics of prestressed word-initial consonant clusters in English. *Research Laboratory of Electronics: MIT Quarterly progress report* 108, 253–260.

[18] Klatt, D. H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am*. 59, 1208–1221.

[19] Lehiste, I. 1973. Rhythmic units and syntactic units in production and perception. *J. Acoust. Soc. Am*. 54, 1228–1234.

[20] Lisker, L. 1974. On "explaining" vowel duration variation. *Glossa* 8, 233–246.

[21] Machač, P. 2006. *Temporální a spektrální struktura českých explozív.* PhD thesis. Prague: Institute of Phonetics.

[22] Machač P., Skarnitzl R. 2009. *Principles of Phonetic Segmentation*. Prague: Nakladatelství Epocha.

[23] Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am*. 54, 1235–1247.

[24] O'Shaughnessy, D. 1995. Timing patterns in fluent and disfluent spontaneous speech. *Proc. ICASSP '95* Detroit, MI, 600–603.

[25] Picheny, M. A., Durlach, N. I., Braida, L. D. 1986. Speaking clearly for the hard of hearing, II: Acoustic characteristics of clear and conversational speech. *J. of Speech and Hearing Research* 29, 434–446.

[26] Port, R. F., Al-Ani, S., Maeda, S. 1980. Temporal compensation and universal phonetics. *Phonetica* 37, 235–252.

[27] Rietveld, T., Kerkhoff, J., Gussenhoven, G. 2004. Word prosodic structure and vowel duration in Dutch. *J. Phon.* 32, 349–371.

[28] Sluijter, A., van Heuven, V. 1996a. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proc. ICSLP '96* Philadelphia, 630–633.

[29] Sluijter, A., van Heuven, V. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.* 100, 2471–2485.

[30] Solé, M. J., Ohala, J. J. 2010. What is and what is not under the control of the speaker. Intrinsic vowel duration. In: C. Fougéron, B. Kühnert, M. D'Imperio, N. Vallée (eds.), *Papers in laboratory phonology 10*. Berlin: de Gruyter, 607–655.

[31] van Santen, J. P. 1992. Contextual effects on vowel duration. *Speech Communication* 11, 513–546.

[32] van Son, R. J., van Santen, J. P. 1997. Strong interaction between factors influencing consonant duration. *Proc. Eurospeech '97* Rhodes, 319–322.

[33] Volín, J. 2009. Metric warping in Czech newsreading. In: R. Vích (ed.), *Speech Processing - 19th Czech-German Workshop.* Prague: Institute of Photonics and Electronics AS CR, 52–55.

[34] Volín, J., Churaňová, E., Šturm, P. 2014. P-centre position in natural two-syllable Czech words. *Proc. Speech Prosody* Dublin, 920–924.

[35] Volín, J., Weingartová, L. 2012. Idiosyncrasies in local articulation rate trajectories in Czech. *Proc. Perspectives on Rhythm and Timing* Glasgow, 67.

[36] Weingartová, L. 2013. Rhythm metrics for speaker identification in Czech. *AUC Philologica 1/2014, Phonetica Pragensia XIII*, 33–42.

[37] Weingartová, L. in preparation. *Speaker identification in the temporal domain of speech.* PhD thesis. Prague: Institute of Phonetics.