Vocalic and consonantal intervals under continuous rate variation

Juraj Šimko[1], *Štefan Beňuš*[2,3]

[1]Bielefeld University, Germany, [2]Constantine the Philosopher University, Nitra, Slovakia,
[3]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia

juraj.simko@uni-bielefeld.de, sbenus@ukf.sk

There is a long tradition of research of the temporal characteristics of vowels and consonants and their relationship to linguistic (e.g., vowel tenseness, consonant voicing) and non-linguistic (speech tempo, number of syllables) features (e.g., Gay, 1978, Klatt, 1976, Port, 1981). Several generalizations were identified: a) C-durations shorten less than V-duration when tempo is increased, b) slowing down from normal tempo preserves phonologically meaningful contrasts but speeding up might wipe them out (Port 1976). These studies are often limited to analysis of discrete levels of speaking rate with a strict control for syllable type (e.g., identical CVC or VC sequences). Languages studied include primarily stress-timed languages like English. Another line of research attempts to use durations of individual vocalic and consonantal intervals extracted from the acoustic signal for characterizing rhythmic properties of languages (e.g. Dankovičová & Dellwo, 2007, Grabe & Low, 2003, Ramus et al., 1999, White & Mattys, 2007). This research supports the idea that a) languages tend to differ along the traditional stress-timed–syllable-timed continuum, and that percentage of vocalic intervals (%V) tends to be a measure robust with respect to rate variation. However, it has been shown that both speaker and material introduce considerable variability, questioning the use of these durational features for analyzing rhythmic and temporal characteristics (e.g., Beňuš & Šimko, 2012). In this study we present evidence contradicting several findings presented above, namely, that %V is tempo-resistant (Dankovičová & Dellwo, 2007) and that %V shows negative correlation with tempo (Gay, 1978).

Data come from Slovak, whose phonology and phonotactic patterns suggest mixed rhythmic properties (neither stress- nor syllable-timed), and whose productive phonological pattern called Rhythmic Law shortens a phonologically long syllabic nucleus if preceded by another long nucleus. Three native speakers of Slovak read two similar semi-meaningful sentences at wide and continuous ranges of speech rate and articulatory precision. Both sentences contain 10 syllables, share 2 words (i.e., 3 syllables) and have very similar syntactic and prosodic structures. Subjects were instructed to repeat the sentences in two blocks. In one block, an experimenter signaled a desired relative speech rate by moving an indicator along the axis between ‘extremely slow’ and ‘extremely fast’. In the second block, the same was repeated with the axis of ‘extremely precise articulation’ and ‘extremely relaxed articulation’. Individual sounds were transcribed and aligned with the acoustic signal. As we found no differences between the two sentences relevant for the analysis presented here, the results for only one of them are presented. The dataset thus consists of 309 repetitions (106, 107 and 96 for speakers S1, S2 and S3, respectively) in which individual sounds could be reliably identified.

Analysis of common rhythm measures that included the same data (Beňuš & Šimko, 2012) showed that %V is the most robust measure of the rhythmic properties of the dataset with respect to speaking rate changes. The left pane of Fig. 1 plots %V and PVI-C measures calculated from our data in bins of increasing speaking rate (indicated by arrows) and compares them with mean values for five languages in Dankovičová & Dellwo (2007). The plot reveals two observations. First, surprisingly, the local prediction that C-durations shorten less than V-duration when tempo is increased does not hold when the unit of analysis is the sentence. In fact, %V *increases* for very fast. The second generalization in is that this relationship does not seem to be linear, supporting the observations in Klatt (1976) and Port (1981). In slower rates, speeding up decreases %V, in moderate ones %V remains relatively stable, while the relationship is reversed in very fast rates where %V clearly increases with rate.

The analysis of the acoustic durations of individual segments in two words subject to the Rhythmic Law (both featuring long stressed syllable followed by short unstressed one) under continuous tempo variation showed that for longer vocalic durations, the inverse relationship between rate and %V predicted by Gay holds, while for short unstressed vowels the trend reverses for faster tempo. Middle pane of Fig. 2 illustrates this pattern. While the quadratic regression curve (R^2=0.12) of data points capturing the ratio of vowel and onset consonant durations in long stressed syllables decreases with increasing tempo, for unstressed syllables the ratio increases for high-tempo sentences (R^2=0.11). Phonemically long vowels are more compressible than consonants but in phonemically short vowels this holds only for slower speech. This points at a possible limitation of generalization of the results of Gay (1978) who analyzed stressed prominent monosyllables. This trend is, however, consistent with Klatt's (1976) and Port's (1981) models that assume minimal intrinsic duration of segments and compressibility of only the portion beyond this threshold. One might assume that this threshold is reached in sentences of approximately 2 s long in our data and for faster rates the pattern is reversed: consonant are now more compressible than vowels that have to maintain the minimal duration. However, as shown in the right pane of Fig. 1, this is not the case since even short vowels shorten linearly with utterance duration and the phonemic contrast in duration is thus maintained even in very fast rates.

In general, our results highlight potential problems with generalizing results obtained in traditional rate-controlled manner for prosodically limited data, and point at strong dependency of widely used measures of rhythmic properties of languages on the analyzed material.

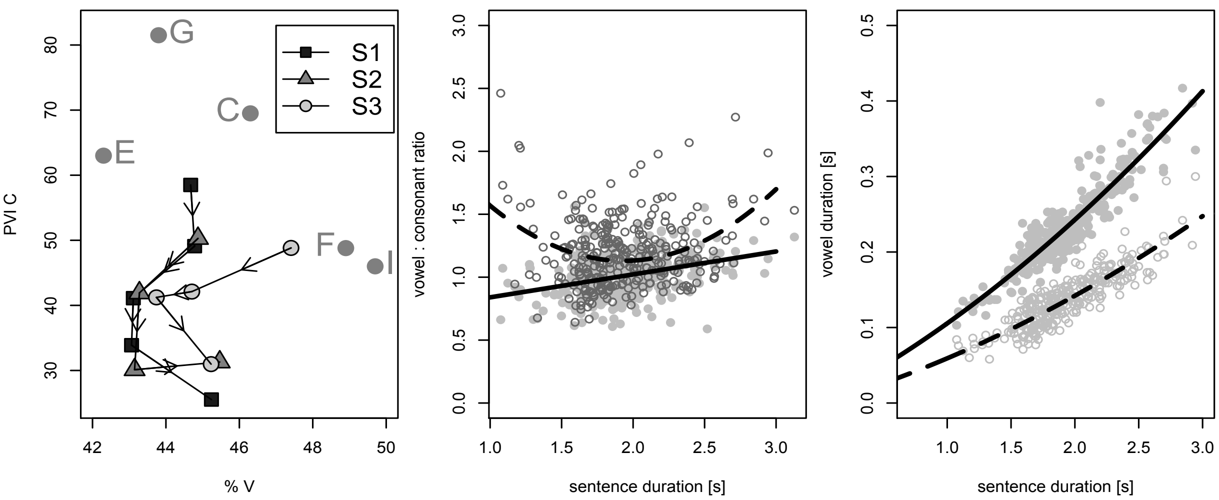

Figure 1. (Left) Mean %V and PVI-C in bins of increasing rate compared with selected languages. (Middle) V:C ratio as a function of sentence duration for long (filled) and short (empty) vowels. (Right) Vowel duration as a function of sentence duration.

References

Beňuš, Š., Šimko, J. 2012. Rhythm and tempo in Slovak. *Proc. 6th International Conference on Speech Prosody*, Shanghai, China.

Dankovičová, J., Dellwo, V. 2007. Czech Speech Rhythm and the Rhythm Class Hypothesis, *Proc. 16th ICPhS*, Saarbrucken, 1241-1244.

Gay, T. 1978. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* 63, 223-230.

Grabe, E., Low, L. 2003. Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology* (7), 515-546.

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

Port, R. 1981. Linguistic timing factors in combination. *J. Acoust. Soc. Am.* 69, 262-274.

Port, R. 1976. *Influence of Speaking Tempo on Duration of Stressed Vowel and Medial Stop in English Trochee Words*. Bloomington: Indiana University Linguistics Club.

Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.

White, L., Mattys, S. 2007. Calibrating rhythm: First language and second language studies, *J. of Phonetics* 35, 501-522.