Rhythm and tempo in Slovak

Štefan Beňuš[1], Juraj Šimko[2]

[1] Constantine the Philosopher University Nitra, Slovakia & II SAS Bratislava, Slovakia
[2] CITEC, Bielefeld University, Germany

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

Abstract

We present the first quantitative analysis of rhythm in Slovak, using two speech corpora designed to exhibit rich prosodic variability including extreme variation in speaking rate and articulatory precision. Several standard consonant and vocalic interval measures show a close similarity of Slovak rhythmic structure with languages traditionally considered syllable-timed, but suggest counter-intuitive dissimilarities with closely related Czech. Additionally, our data highlight strong dependency of these measures on speech material, speaker and tempo. Overall, our results shed further doubt on the usefulness of CV interval measures in speech rhythm research.
Index Terms: speech tempo, rhythm, Slovak

1. Introduction

Rhythm is an inherent, yet a very complex and elusive aspect of speech prosody. Several studies tried capturing similarities and differences among languages related to the rhythmic properties with two basic types of durational measures extracted from V(ocalic) and C(onsonantal) intervals: variability of these intervals (%V, ∆V, ∆C) and the rate-normalized standard deviations VarcoV and VarcoC [4,5,10,13] and pair-wise variability in consecutive intervals of the same type, raw or rate-normalized (PVI-C, PVI-V, nPVI-V, nPVI-C) [8]. These studies showed that the original division of languages to "stress-timed" and "syllable-timed" [1] is in fact a continuum but also that it is possible to capture relative similarities and differences among the rhythmic characteristics of languages using a 2D space of one consonantal and one vocalic measure.

Measures of variability in C-intervals are supposed to capture differences among languages in terms of permissible phonotactic complexity of clusters: traditionally stressed-timed languages allow more complex clusters than syllable-timed languages. V-interval measures should capture stress-related reduction of vowels: stressed timed languages robustly neutralize the quality and duration of vowels in unstressed syllables while traditionally syllable timed languages display only minimal reduction. It is also assumed that C- and V-intervals reflect the sensitivity of infants to rhythmic differences observed in infant discrimination studies [10].

It is clear that measures of rhythm based on the duration of intervals are greatly dependent on both the linguistic material as well as on speech rate. [11] showed that rhythmic measures differ for different material within a language. Hence, it is important to assess to what extent the material used for rhythm analysis is representative primarily in the phonotactic complexity of syllables. Moreover, inter-language studies are commonly limited to a small number of sentences ([5,6,8,10]).

The correlation between the rhythmic measures and speech rate has been long noticed and some studies tried to control for rate in their material while others tried to use rate normalizations in post-processing of the data. In C-intervals, [6] showed that ∆C is greatly influenced by rate since faster rates result in shorter C-intervals leading to lower variability. Measure %V (and ∆V to a lesser extent) seemed more stable. Similarly, [13] reported significant correlations between ∆C and speech rate and the rate-stability of %V. This relatively greater rate-stability of V-intervals compared to C-intervals is surprising since [7] found that increasing speech rate should affect vocalic rather than consonantal durations, at least in English. Normalizing ∆C for rate (VarcoC) in [13] eliminated the correlation but also caused the disappearance of differences between traditionally stress- and syllable-timed languages. [4] concluded that PVI-C and %V provide the best separation among traditionally stress and syllable timed languages at a variety of speech rates, but also that syllable timed languages are spoken faster and have lower syllable complexity than stress timed ones. Further, he argued that perceived rhythmic differences are strongly correlated with rate and suggested that the traditional rhythmic differences on the syllable-stress timed continuum might just be artifacts of variability in rate.

The goal of this paper is three-fold: 1) provide a first quantitative description of rhythm production in Slovak, 2) investigate the relationship between speech rate and rhythm measures using data with continuous variation of rate and data with more stable rate, and 3) investigate the effect of material on the rhythmic measures by comparing a corpus of 2 sentences repeated in different rates with a corpus with rich variability in lexical material and syllable complexity.

Slovak is a west-Slavic language that has been impressionistically sometimes described as stress-time [9], but rigorous experimental studies are missing. Several features of Slovak point to a mixed system. First, observed shortening and centralization in unstressed vowels compared to the stressed ones is minimal, ([3] using only CVCa non-sense words), which points toward a syllable-timed end of the continuum. But, Slovak displays phonemic length distinctions that increase the variability in V-intervals pointing toward the stress-time languages. Moreover, Slovak phonology features the so-called Rhythmic Law in which a long vowel changes to its short counterpart if preceded by another long vowel, which should increase V-interval variability even more. Finally, Slovak allows relatively complex onset clusters, which, however, are underrepresented in natural speech. A closely related Czech with similar features of V-nonreduction, cluster complexity, and phonemic length contrast, was shown to lie in between stress-timed and syllable-timed languages on several measures [4,5] but more closely to the former than the latter.

2. Methodology

Material for this study consists of two data sets. The first one contains 102 Slovak sentences read by one native speaker (MR) at a comfortable most natural speaking rate. The sentences were designed to provide rich prosodic variability comprising of syntactically varied simple and complex declarative sentences, questions, or exclamations Duration varied between 5 to 55 syllables (mean 16.8, total 1707).

The second dataset represents two similar Slovak sentences read by three native Slovak speakers (S1, S2, S3) at wide and continuous ranges of speech rate and articulatory precision [12]. 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'. This procedure yielded 626 sentences in which we could reliably identify individual sounds in the signal.

In both datasets individual sounds were transcribed and aligned with the acoustic signal using standard procedures. Following guidelines in [10,4], V- and C-intervals were identified. Slovak contains unrestricted syllabic liquids, which were treated as V-intervals for this study. For each sentence we extracted the durations of the intervals and calculated the rhythmic measures commonly used in previous studies analyzing rhythm (see [13,4] for formulas): sum of V-interval durations divided by total duration (%V), standard deviation of V- and C-intervals in centiseconds (ΔV, ΔC) and their rate-normalized values $VarcoV$, $VarcoC$, raw pair-wise variability index for consecutive V-and C-intervals (PVI-V, PVI-C) and their rate normalized values ($nPVI$-V, $nPVI$-C), and CV-$rate$ as the number of C- and V- intervals per second. (Results for some of these measures are not reported in this paper.)

3. Results

3.1. Syllable and C/V-interval complexity

We first compare the complexity of Slovak material with those reported for the BonnTempo corpus [4,5], which contains recording of a short text, originally in German (=76 syllables), translated to several languages including Czech (99 syllables). Figure 1 shows the distribution of C/V intervals in the lexically rich MR data. Of 5 languages in [4,5], Slovak is most similar to Czech and French, whose syllable and C-interval complexity are shown in Figure 1 for comparison. Slovak syllables tend to resemble Czech ones whereas C-intervals are more similar to French ones.

Slovak V-interval complexity shows that 76% of the V-intervals contained a short vowel while the remaining 24% were long vowels (12%), diphthongs (6%), or adjacent vowels from different syllables (6%). Hence, long vowels together with diphthongs (both phonemically long) made approximately 18% percent of all vowels, which is similar to Czech data and corresponds to the frequency of long and short vowels of about 15% vs. 85% based on the frequency table of Slovak phonemes in [9]. The composition of syllables and C- and V-intervals in the MR data together with its features mentioned in Section 1 predict that the rhythm measures based on these intervals should put Slovak between stress- and syllable-timed languages but closer in the vicinity of the latter.

A similar examination in the limited material of S1-S3 data shows greater frequency of CCV (40%) and smaller of CV (40%) syllables compared to MR's material, resulting in the distribution of C and CC types in C-intervals of 60% vs. 40%, compared to 70%-30% in the MR corpus. The distribution of long/short vowels is almost identical in the two corpora. Hence, we might predict that S1-S3 should show greater variability in C-intervals than MR and relatively similar variability in V-intervals should be observed.
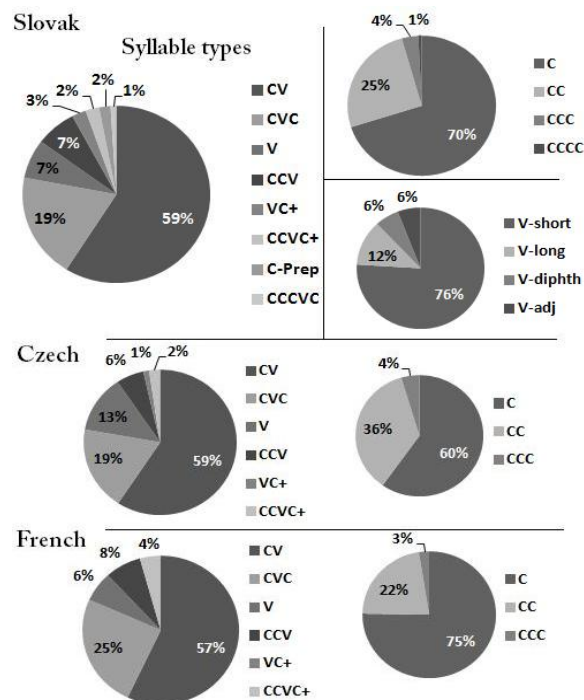

Figure 1: *Syllable and C/V-interval complexity in MR corpus; Czech and French data adapted from [4].*

3.2. Dependency on tempo, material, and speaker

We are interested in how various rhythm measures depend on speaking rate, recorded material (prosodically varied sentences vs. repeated single sentence) and speaker. Figure 2 illustrates the distribution of *CV-rate* for our four speakers. It shows that S1 elicited both the fastest speech and the widest range of speaking rate. *CV-rate* captures speech rate variation for S1-S3, who intentionally varied the rate. For MR, however, no rate variation was intended, and *CV-rate* rather reflects lexical characteristics of particular sentences, like a presence/absence of long vowels and complex C-clusters.
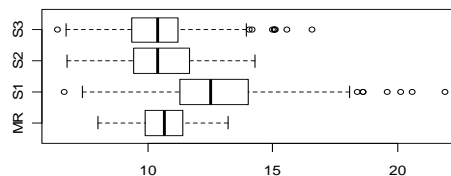

Figure 2: *CV-Rate data distributions for 3 speakers*

One-way Anova with TukeyHSD post-hoc tests showed that mean S1's *CV-rate* (12.7) was significantly higher than the rates of S2, S3, and MR (10.5, 10.4, and 10.6 respectively) that in turn did not differ from each other. Pair-wise Levene tests showed that the standard deviations of *CV-rate* (1.1 (MR), 2.4 (S1), 1.5 (S2), and 1.8 (S3)) were significantly different for each pair of speakers except the S2-S3 pair. Comparing with 5-step (very slow, slow, normal, fast, very fast) rate elicitation for 5 languages reported in [4], mean MR's, S2's, and S3's rates correspond to mean normal rate (slow for French/Italian, normal for Czech/English, and fast for German), S1's rate corresponds to mean fast rate.

Next, we examined the correlation between the rhythm measures and speech rate, shown in Table 2. The significant correlations are mostly negative. However, the fastest of our speakers (S1) is similar to S2 and different from MR despite rate similarity of MR and S2, which suggests the effect of material. %V is the most rate-insensitive measure. Comparing raw *ΔV* with *nPVI-V* that is supposed to filter out the effect of speech rate, we see no effect in rather rate-stable speech of MR, some decrease in S2 and a sharp decrease in S1. Hence, rate normalization is effective for fast rates and high variability (S1) but less so for slower rates and less variability. In short, for our data, none of the rate-normalizations of V-interval measures seems to be robust enough to deal with both rate and material differences.

In C-intervals, *VarcoC* provides effective rate-normalization for MR but is less successful with S1-S3 data. These observations are in line with [6] in which the greatest rate variation was shown on the C-interval axis (ΔC in their case) and minimal on %V.

Table 2. *Pearson's correlations of rhythmic measures with CV-rate; dark fill p<.01, light fill p < .05.*

	%V	ΔV	nPVI-V	ΔC	VarcoC	PVI-C
MR	0.18	-0.34	-0.34	-0.39	0.09	-0.15
S1	0.05	-0.73	-0.07	-0.79	0.35	-0.59
S2	0.02	-0.82	-0.53	-0.76	0.42	-0.40
S3	0.01	-0.68	-0.15	-0.63	0.47	-0.36

Table 3. *F-values of mixed-models tests; dark fill p<.01*

	%V	ΔV	nPVI-V	ΔC	VarcoC	rPVI-C
Material	47.9	1.8	3.0	5.0	0.1	0.1
Speaker	58.2	52.9	50.0	43.5	57.1	9.5

After investigating rate-dependence, we tested the effects of material (MR vs. S1-S3) and speaker on the rhythm measures. First, we divided *CV-rate* data into 5 bins separated by the following values: 8.17, 10.45, 13.37, 17.1. These values were computed by dividing the log transformed interval of all *CV-rate* values into sub-intervals of equal length, and then exp-transforming the border values back to the original scale. Table 3 shows F-values from mixed-models testing [2] for the effects of Material (with speaker and discrete *CV-rate* as random factors) and Speaker (with Material and discrete *CV-rate* as random factors) on the rhythm measures. Values greater than 8.6 are considered significant at p < 0.01.

Speaker variation affects all rhythm measures significantly, but the difference in material between the two corpora has no significant effect apart from *%V*. Hence, while *%V* is resistant to within-speaker speech rate variation (Table 2), it is sensitive to both material and speaker differences.

3.3. Comparison with Czech and other languages

Table 4 shows means and stdevs for our rhythm measures. Comparing our *%V* data (2nd row) with data in [5], MR's speech falls between Czech (~46) and French (~49), which is in line with the phonotactic patterns discussed above. [10] reported for French mean *%V* of 43.6 and [8] 50.6 with MR's data falling in-between these values. However, S1-S3 have significantly lower values (confirmed with Anova and post-hoc HSDTukey tests) falling closer to 'stress-timed' languages. Hence, *%V* in MR's material put Slovak closer to

syllable-timed languages while S1-S3 data show a more stress-timed character.

Table 4. *Means and standard deviations (in the brackets) for C- and V-interval measures.*

	MR	S1	S2	S3
%V	47.7 (5.6)	43.8 (2.2)	43.2 (1.9)	44.1 (3.5)
ΔV	4.1 (1.5)	3.2 (0.7)	3.0 (0.8)	4.0 (1.2)
varcoV	41.1 (15.9)	45.1 (6.5)	35.62 (5.8)	45.2 (7.8)
nPVI-V	38.8 (11.3)	46.7 (7.2)	44.7 (8.2)	53.8 (9.9)
ΔC	4.7 (1.0)	5.2 (0.9)	5.2 (0.7)	5.8 (1.0)
VarcoC	49.8 (9.2)	57.0 (6.6)	47.4 (4.5)	53.4 (6.8)
PVI-C	44.8 (12.2)	42.5 (9.5)	43.3 (10.5)	49.3 (12.8)

Interestingly, reported *nPVI-V* values for Czech in [5] (~45) are different from MR's, and S3's but similar to S1-S2 data. Comparing with data in [8], MR's mean value is lower than traditionally 'syllable-timed' languages while S3's 53.8 is higher than syllable-timed but lower than stress-timed ones.

Choosing the most informative C-interval measure is less straightforward. [5] found that *VarcoC* and *PVI-C* for Czech showed more similar results to 'stressed-timed' languages than to 'syllable-timed' ones, but also noted problems with both measures in differentiating English from French and Italian in their data. Our data in the bottom two rows of Table 4 put Slovak into the 'syllable-timed' group with the exception of S1's *VarcoC*. Comparing our *ΔC* with values reported in [10] and [6], the syllable-timed character of Slovak is supported.

Figure 3 shows our data plotted with the most used measures in other studies *(%V, VarcoC, rPVI-C, and nPVI-V)*. For comparison, mean values for five languages from [5] are plotted in the background. Values for different speakers are plotted as circles (MR), squares (S1), triangles (S2), and diamonds (S3). We used the binning procedure as described in 3.2. Note that not all speakers elicited CV-rate values falling into all 5 bins; the arrows indicate increasing speaking rate.

In the left-most panel (*nPVI-V~PVI-C*), Slovak data, at least in medium rate, seem most similar to French and Italian. Both dimensions show considerable rate dependency. This is expected for *PVI-C*, but not for *nPVI-V* that shows substantial variation. A tendency for faster speech to be more 'syllable-timed' can be observed. The best separation between traditional language groups with Slovak data added seems to be provided by *PVI-C*.

The second panel of Figure 3 shows the same data in *%V~VarcoC* space. *VarcoC* shows considerable rate dependency and Slovak data cover considerable area (in particular fast speech for S1 & S3, which weakens the proposals in [4,13] that *VarcoC* provides a good rate-filter). MR's data group well with French/Italian. Interestingly, faster rates do not seem to 'induce' more 'syllable-time' characteristics since data from all speakers seem to move 'up' rather than 'down', which would be expected following a conclusion in [4] about faster rates being perceived as more syllable-timed.

It seems that the dimension doing the best job in separating the languages are %V and, surprisingly, PVI-C. The plot with these two dimensions is in the third panel of Figure 3. It indeed shows both a good separation from English, German and Czech, highlights material dependency on *%V*, small speaker dependency, and also supports positive correlation between faster speaking rate and more 'syllable-timed' characteristics. On the contrary, the dimension that seems to
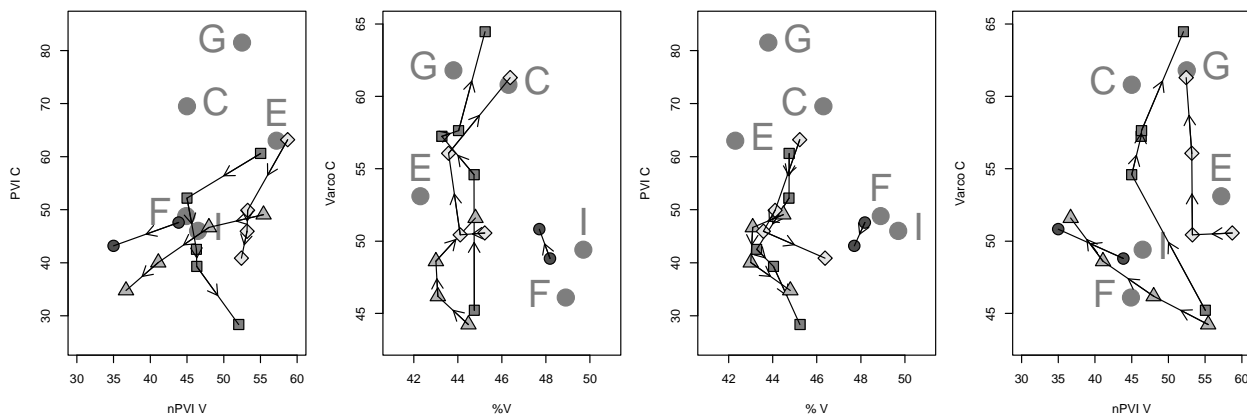
Figure 3: *Comparison of Slovak with C(zech), G(erman), E(nglish), F(rench) and I(talian) in four 2D spaces (from the left): nPVI-V~rPVI-C, %V~VarcoC, %V~rPVI-C, and nPVI-V~VarcoC; see text for label explanations.*

be the least informative is *nPVI-V~VarcoC* that is shown in the rightmost panel of Figure 4. Here, data from a single language (Slovak) span the continuum between syllable-timed and stress-timed languages on both dimensions and speaker/material variability is substantial.

4. Discussion & Conclusions

The analysis of a wide range of C and V interval measures confirmed their considerable material and speaker dependency. In fact, not a single measure considered in this paper was robust with respect to these variables.

Furthermore, wide speaking rate variation in part of the analyzed material allowed us to evaluate how these measures depend on this prosodic dimension. The absolute, non-rate-normalized variables (*ΔV*, *ΔC*, *PVI-C*) show the expected, mathematically justified, correlations with tempo. More seriously, the measures designed to mitigate the speaking rate influence (*nPVI-V* and *VarcoC*) are correlated with speaking rate, the correlation coefficients being of similar magnitude to those of the raw measures. The only analyzed measure robust with respect to tempo proved to be %V, but this measure is material/speaker sensitive. The rate independence of *%V* is somewhat surprising in the light of Gay's results [7] suggesting different quantitative effects of rate increase on V and C intervals.

A degree of rate dependency of the used measures is to be expected, in particular for the extremely varied material of S1-S3. Our analysis, however, brought up interesting repercussions of this dependency for distinguishing rhythmic properties of various languages: the consequences of rate increase/decrease in terms of speech material classification considerably vary with the measure used.

One of the goals of the paper was to provide the first quantitative analysis of rhythm in Slovak and compare it with data from closely related Czech and other languages. Despite the shortcomings of the used measures outlined above, the various quantitative analyses *consistently* placed Slovak data in the vicinity of languages traditionally perceived as syllable-timed, i.e., French and Italian. The comparison of Slovak CV phonotactic complexities with data from six languages is compatible with this classification.

At the same time, the placement of Slovak very far from closely related Czech, in fact in different, 'syllable-timed' language group, is highly counter-intuitive. Slovak Rhythmic Law should push it towards even more 'stress-timed' end of the spectrum compared to Czech, but the measures we used

failed to capture this postulated effect. This result adds to the doubts about the usefulness of CV interval measures in classifying the rhythmic properties of languages and requires further scrutiny.

5. Acknowledgements

6. References

[1] Abercrombie, D., *Elements of general phonetics*, Aldine: Chicago, 1967.

[2] Baayen, R. H. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R* Cambridge: CUP, 2008.

[3] Beňuš, Š, Mády, K., Effects of lexical stress and speech rate on the quantity and quality of Slovak vowels, *Proc. of 5th Speech Prosody*, 2010.

[4] Dellwo, V., *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence*, PhD thesis, RFW Univesity, Bonn, 2010.

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

[6] Dellwo, V., Wagner, P., *Relationships between speech rate and rhythm*, Proc. of the 15th ICPhS, 471-474, 2003.

[7] Gay, T. Effect of Speaking Rate on Vowel Formant Movements. *J. Acoust. Soc. Am.*, 63(1):223 – 230, 1978.

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

[9] Kráľ, Á., Sabol, J. *Fonetika a fonológia*, Bratislava: SPN, 1989.

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

[11] Ross, T., Ferjan, N., Arvaniti, A., On quantifying speech rhythm. Workshop on Empirical Approaches to Speech Rhythm, 2008, UCL, London.

[12] Šimko, J., Cummins, F., Beňuš, Š., An analysis of the relative timing of coarticulated gestures within VCV sequences, *Proc. of the 17th ICPhS*, 1850-1853, 2011.

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