Disentangling sources of rhythmic variability between dialects

Adrian Leemann[1], Volker Dellwo[1], Marie-José Kolly[1], Stephan Schmid[1]

[1]Phonetics Laboratory, Department of General Linguistics, University of Zurich
{adrian.leemann, marie-jose.kolly, schmidst}@pholab.uzh.ch, volker.dellwo@uzh.ch

10.21437/SpeechProsody.2014-127

Abstract

Speech rhythm is highly variable. Previous studies reported variability between languages, dialects, speakers, and labelers. Research further revealed an effect of *sentence* in the rhythmic characteristics of speakers of the same language. In the present study we tested whether the effect of sentence material is constant across varieties of the same language. We addressed this question by an example of analyzing rhythmic variability between eight dialects of Swiss German in three different sentences. Results showed a significant interaction for *dialect*sentence* for most of the tested rhythm metrics. We take this as evidence that differences between dialects are contingent upon the sentences used in the experiment. We further investigated which sources in the sentence material caused between-dialect differences in rhythm scores to vary. We found exemplary evidence that dialect-specific phonological and morphological phenomena contained in the individual sentences are the prime suspects. Implications for future speech rhythm research are discussed.

Index Terms: Speech rhythm; dialectology; Swiss German, rhythm metrics

1. Introduction

Acoustic measures of speech rhythm that are based on temporal features of speech have been reported to vary significantly between and within languages. Yet, relatively little is known about the actual sources behind this variability. [1] suggested that metrics such as the percentage over which speech is vocalic (%V) reflect the degree of vowel reduction, and metrics such as the standard deviation of consonantal intervals (ΔC) capture syllable complexity. [2], however, provided evidence that differences in rhythm metrics of typologically different languages emerge even when syllable structure complexity is controlled for. They reported that the durational marking of prosodic heads or pre-final heads accounted for more rhythmic variability for the language set investigated. [3, 4] examined the degree to which sentence material affected rhythm scores. Both studies reported effects of *sentence*, implying that rhythm scores strongly differ depending on the sentence material being analyzed.

While these studies have shown an effect of *sentence* in the rhythm scores of speakers of the same language, it is unclear whether such effects are constant across varieties of the same language. In this preliminary study we address this question by an example of analyzing rhythmic variability between dialects of Swiss German (SwG) in different sentences. Our paper addresses the following research questions:

(1) Does sentence material have an influence on rhythmic differences between dialects?
(2) Which parameters exactly in the sentence material cause between-dialect differences in rhythm scores to vary?

To test these hypotheses, we applied rhythm metrics on the following 8 SwG dialects, see Table 1.

	West	**East**
Midland	**BS:** Basel	**TG:** Thurgau
	BE: Bern	**ZH:** Zurich
Alpine	**SB:** Sensebezirk	**SZ:** Schwyz
	VS: Valais	**GR:** Grisons

Table 1: Selected dialects and their abbreviations.

The sentences used in the present study are a subset of those used in [5], where the declarative intonation patterns of the dialects mentioned were examined based on 7 read sentences per speaker. Since the current study is preliminary in nature, we analyzed only 3 of [5]'s 7 sentences per speaker. The other 4 out of 7 sentences used in [5] were the basis of between-dialect rhythmic analyses in [6]. [6] reported significant differences for these 8 SwG dialects particularly in the variability of vocalic intervals, while consonantal variability was less discriminant. The current study is thus a follow-up of [6].

We hypothesize that dialect-specific phonological processes present in some but not all sentences are possible reasons as to why rhythm scores vary according to *sentence*. Consider, for instance, the following example: a typical feature of Basel SwG is extensive vowel lengthening before [r]; often the [r] is elided completely. In *stark* 'strong', for example, Basel speakers articulate [ˈʃtɑːx], while most other SwG dialect spekaers realize [r] before the [x] without vowel lengthening, resulting in [ˈʃtɑrx]. If some sentences in the data set contain words that allow for this phonological process to take place, it is conceivable that this has an effect particularly on %V and vocalic variability measures such as VarcoV, ΔV, and nPVI_V. To test this hypothesis, we selected three sentences that vary in the proportions of vowels and consonants as well as in syllabic make-up. Given these examples, we expect to find significant interaction of *dialect*sentence*.

2. Data and methods

2.1. Subjects

6 speakers were recorded for each of the 8 dialects, adding up to a total of 48 speakers (33 females, 15 males). The subjects aged between 17 and 69 confirmed to be using the dialects in question on a daily basis. None of the speakers reported speaking or hearing problems.

2.2. Material

Each speaker was asked to read in their respective dialects three sentences that were written in Standard German:

(1) *Warum verfolgt der Hund die Katze?*
'Why does the dog chase the cat?'

(2) *Der Bildschirm leuchtet stark*
'The monitor glows brightly'

(3) *Die Union von den Nonnen hat einen neuen Namen*
'The union of the nuns has a new name'

Sentence 1 on average consisted of 12 consonants and 6 vowels, sentence 2 of 14 consonants and 6 vowels, and sentence 3 of 12 consonants and 13 vowels. With respect to the syllabic make-up of the sentence material, the three sentences demonstrated the following distributions of syllable structures averaged over all dialects, see Figure 1.

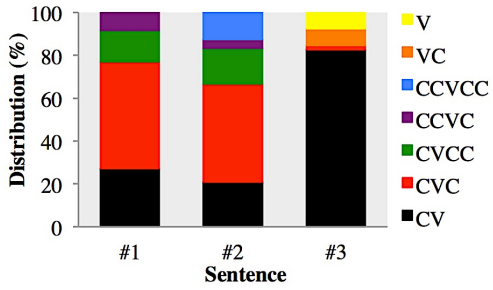

Figure 1: Syllabic make-up of sentences 1, 2, and 3.

Figure 1 reveals that sentences 1 and 2 were very different from sentence 3. Averaged over all dialects, the latter contained 83% CV syllables (black) while sentences 1 and 2 featured only 27% and 21% CV syllables. Sentences 1 and 2 contain much more CVC syllables (red) (sentence 1: 50%, sentence 2: 49%, sentence 3: 2%), however. Syllabic make-up of the sentences was further different between the dialects examined. Figure 2 shows the syllabic make-up of sentence 1 by dialect, for example.

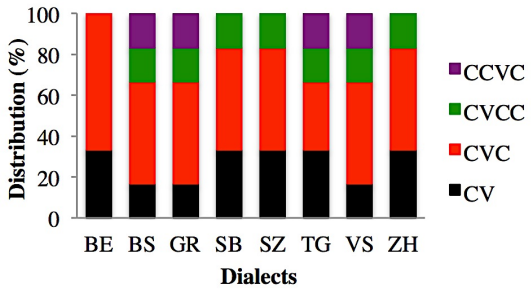

Figure 2: Syllabic make-up of sentence 1 by dialect.

Figure 2 reveals that dialects differed with regard to the relative proportions of syllable types in sentence 1. The BE speakers only exhibited CV (33%) and CVC (67%) syllables, for instance.

2.3. Procedure

Subjects were recorded in their respective locations with a Marantz PMD-671 solid-state recorder (sampling rate of 44.1 kHz and 16 bit quantization) and a Sennheiser clip-on ME2 omni lavalier microphone. Sentences were transcribed in SAMPA and labeled on the segmental level in *Praat* [7]. Consecutive vowels or consonants were merged into vocalic and consonantal intervals respectively, which provided the preferred labeling format for a subsequent application of rhythm metrics and statistical analyses. The following metrics were calculated using the *Praat* plugin 'Duration analyzer' (http://www.pholab.uzh.ch/leute/dellwo/software.html).

C:V ratio measure
The percentage over which speech is vocalic: %V [1]

Vocalic variability measures
a). The rate-normalized standard deviation of vocalic intervals: VarcoV [8]
b). The rate-non-normalized average differences between consecutive vocalic intervals: rPVI_V [9]
c). The rate-normalized average differences between consecutive vocalic intervals: nPVI_V [9]
d). The rate-non-normalized standard deviation of vocalic intervals: ΔV [1]

Consonantal variability measures:
a). The rate-normalized standard deviation of consonantal intervals: VarcoC [10]
b). The rate-non-normalized average differences between consecutive consonantal intervals: rPVI_C [9]
c). The rate-normalized average differences between consecutive consonantal intervals: nPVI_C [9]
d). The rate-non-normalized standard deviation of consonantal intervals: ΔC [1]

3. Results

3.1. Statistical analyses

All data were analyzed using R [11] and the R packages *lme4* [12] and *languageR* [13, 14]. If not indicated otherwise, we analyzed data using linear mixed effect models (LMEs). Normality was checked by visual inspection of quantile plots. *Dialect* was treated as a fixed effect, *speaker* and *sentence* as random effects. Effects were tested by model comparison between a full model, in which the factor in question was entered as either a fixed or a random effect, and a reduced model without this effect. p-Values were obtained by comparing the results from the two models using ANOVAs. For the assessment of the relative goodness of fit, we report *AIC* (Akaike Information Criterion) values that decrease with goodness of fit. Only p-values that are considered significant at the $\alpha = 0.05$ level are reported.

3.2. Interaction of *sentence* and *dialect*

Table 2 summarizes the statistics for the effects of *dialect*, *sentence*, and *dialect*sentence* by rhythm measure.

Rhythm measure	Factor	Result	Rhythm measure	Factor	Result
ΔV	dialect	AIC=-748, p<.0004*	ΔC	dialect	AIC=-502, p<.007*
	sentence	AIC=-748, p<.0001*		sentence	AIC=-502, p<.0001*
	dialect*sentence	AIC=-519, p<.003*		dialect*sentence	AIC=-519, p<.003*
VarcoV	dialect	AIC=-180, p<.0001*	VarcoC	dialect	ns.
	sentence	AIC=-180, p<.0001*		sentence	AIC=-110, p<.0001*
	dialect*sentence	AIC=-194, p<.006*		dialect*sentence	AIC=-124, p<.004*
rPVI_V	dialect	AIC=603, p<.0005*	rPVI_C	dialect	AIC=719, p<.0008*
	sentence	AIC=603, p<.0001*		sentence	AIC=719, p<.0001*
	dialect*sentence	AIC=572, p<.0001*		dialect*sentence	AIC=705, p<.01*
nPVI_V	dialect	AIC=1195, p<.0001*	nPVI_C	dialect	ns.
	sentence	AIC=1195, p<.0001*		sentence	AIC=1147, p<.0001*
	dialect*sentence	AIC=1173, p<.0002*		dialect*sentence	AIC=1132, p<.008*
%V	dialect	AIC=846, p=.024*			
	sentence	AIC=846, p<.0001*			
	dialect*sentence	ns.			

Table 2: Summary of mixed model comparisons for *dialect* and *sentence* by rhythm measure.

The comparison between the full and reduced models showed a significant difference in all rhythm metrics for *dialect* and *sentence* (except for VarcoC (cf. [6]) and nPVI_C for *dialect*), with the full model exhibiting an increased goodness of fit. Moreover, for all metrics except for %V we obtained a significant interaction of *dialect*sentence*. We take this as evidence that between-dialect variability strongly depends on the sentence used. To study simple effects of *dialect*, we conducted individual model comparisons for each of the three sentences for those rhythm metrics that showed significant interaction. 16 of the 24 model comparisons (3 sentences*8 rhythm metrics that showed interaction) revealed significant effects of *dialect* (Bonferroni adjusted for sentence, α=.017).

To visually illustrate that rhythmic differences between the dialects are contingent upon *sentence*, Figures 3 and 4 show boxplots of the dialects' VarcoV and nPVI_V by *sentence*.

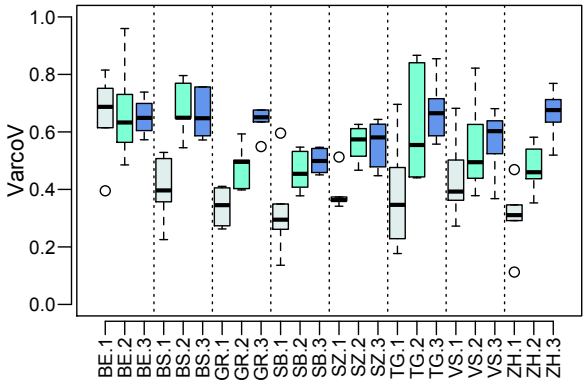

Figure 3: Boxplots of the dialects' VarcoV by sentence.

For VarcoV (see Figure 3), simple effect tests showed significant effects for *dialect* in all three sentences (Bonferroni adjusted for *sentence*, α= .017; sentence 1: *AIC*=-50, *p*<.0001, sentence 2: *AIC*=-51, *p*<.005, sentence 3: *AIC*=-90, *p*<.002). Post-hoc tests (also model comparisons, Bonferroni adjusted for *dialect* and *sentence*, α=.002) revealed, however, that dialectal differences in many cases were contingent upon the sentence being examined. In sentence 1 (grey), for example, BE speakers revealed significantly more vocalic variability (M=.66, SD=.15), than ZH speakers (M=.31, SD=.11). In sentences 2 (turquois) and 3 (blue), however, differences between these two dialects were no longer present.

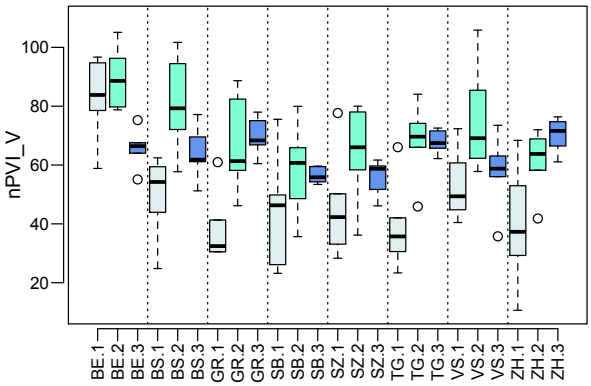

Figure 4: Boxplots of the dialects' nPVI_V by sentence.

For nPVI_V (see Figure 4), simple effect tests revealed significant effects of *dialect* in all three sentences (Bonferroni adjusted for *sentence*, α=.017; sentence 1: *AIC*=413, *p*<.0001, sentence 2: *AIC*=408, *p*<.005, sentence 3: *AIC*=339, *p*<.0005). Post-hoc tests also showed, however, that dialectal differences often depended on the sentence being examined. In sentence 1 (grey), for instance, BE speakers had significantly more vocalic variability (M=.83, SD=14), than BS speakers (M=50, SD=14). In sentences 2 (turquois) and 3 (blue), however, these differences were no longer present.

It is interesting to see that in both Figures, Figure 3 and 4, sentence 1 exhibits the lowest VarcoV and nPVI_V values for nearly all dialects, even though sentence 1 and sentence 2 demonstrate a nearly identical distribution of syllable structures (cf. Figure 1). It is conceivable that this has to do with the rhythm metrics at work in these examples: measures such as VarcoV and nPVI_V capture vowel duration variability and not syllable duration variability.

3.3. Searching for potential triggers in the material

In this section we examine possible sources in the material that cause rhythm scores between dialects to vary. The data provided are exemplary evidence and described impressionistically, largely based on visual inspection. Only VarcoV and nPVI_V are considered because these measures are controlled for speech rate, they revealed significant interaction for *dialect*sentence* (see Table 2), and they have proven to be highly discriminative for the 8 SwG dialects examined [6].

VarcoV: SB and TG

Figure 3 revealed that VarcoV differentiates SB (M=.49, SD=.04) from TG speakers (M=.67, SD=.10) in sentence 3. Post-hoc tests showed that the difference between the dialects was only significant for this sentence. In the other two sentences the dialects do not differ. If we look at a typical SB and TG realization of sentence 3, we detect two phenomena that may contribute to a lower VarcoV in the SB dialect, see Figure 5.

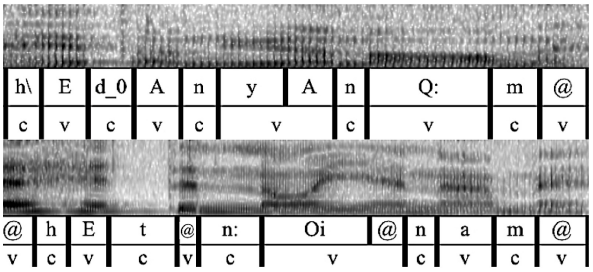

Figure 5: Typical articulation of the constituent *hat einen neuen Namen* (sentence 3) by an SB speaker (top panel) and a TG speaker (bottom panel).

Firstly, the indefinite article in 'a new name' in Figure 5 (top panel) is realized as a full vowel [ɑ] by the SB speaker, with a duration of 59 ms. The TG speaker, on the other hand, articulates a short schwa with a duration of 42 ms. Alpine dialects such as SB SwG have a tendency of realizing unstressed light syllables as full vowels. Secondly, Figure 5 reveals that the SB speaker lengthens the [ɑ] in *Name*, 'name'. SB speakers have a tendency of lengthening Middle High German short vowels [15]. This [ɑ] in the SB SwG is 155 ms long, while the TG speaker's [a] is only 112 ms long. If a sentence contains tokens that allow for full vocalic articulation and lengthening of short vowels, it is plausible that this leads to less vocalic variability and thus smaller VarcoV values for SB SwG.

nPVI_V: BE and TG

Figure 4 revealed that BE (M=83, SD=14) and TG (M=39, SD=15) differed significantly in nPVI-V in sentence 1. Post-hoc tests showed that the two dialects differ in this sentence, not, however, in sentences 2 and 3. If we look at a typical BE and TG realization of sentence 1, we find one phenomenon which may contribute to the higher nPVI_V values in the BE dialect, see Figure 6.

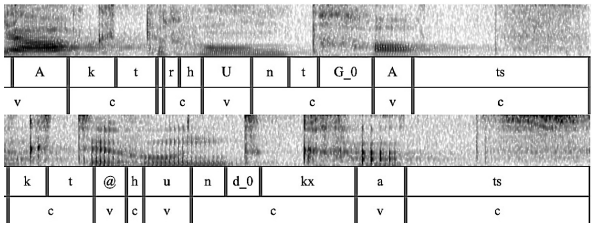

Figure 6: Typical realization of the constituent *jagt der Hund die Katze?* (sentence 1) by a BE speaker (top panel) and a TG speaker (bottom panel).

One of the triggers for a higher nPVI_V value in the BE dialect in this sentence may be the realization of an [r] in the masculine NOM. definite article *der*, as in *der Hund*, for which most other SwG dialects use the form [d̥ə]. Because of this [r], the [ə] is strongly reduced (15 ms). The TG speaker realizes *der* as [d̥ə], where the schwa is articulated more fully (76 ms). This morphological property is particularly typical of BE SwG and may have contributed to the higher proportions of CVC syllables for the BE speaker group in sentence 1 (cf. Figure 2). If a sentence contains such dialect-specific morphological properties, it is possible that this leads to different distributions in syllable structures between the dialects, and possibly more vowel reduction, which in turn may cause an increase in nPVI_V.

4. Discussion and conclusions

Results of the current study support an answer to question (1) *Does sentence material have an influence on rhythmic differences between dialects?* as a *yes*. Effects of *sentence* are not only found within one and the same language, as shown by [3, 4], but also across varieties of the same language: in the majority of the studied rhythm metrics we found significant effect of *sentence* as well as a significant interaction of *dialect*sentence*. That is, cross-dialectal rhythmical differences heavily depend on the sentence material being examined. To complicate matters, the *dialect*sentence* interaction was manifested differently depending on the applied rhythm metric (see Figures 3 and 4). For VarcoV, for example, we found that in VS and ZH SwG sentence 3 has a higher VarcoV than sentence 2. In nPVI_V, however, VS SwG sentence 3 has a lower value than sentence 2, while in ZH SwG, sentence 3 still has a higher value than sentence 2. Rhythmic variability between dialects is more complex than hitherto predicted.

These results provide further support that the selection of sentences in speech rhythm research must be given extra attention, cf. [3, 4, 16]. One can choose to work with a small data set and a special consideration for issues of dialect- or language-specific phonotactic and prosodic representativeness [3, 4], or one can work with a data set large and manifold enough to level out sentence effects. While the applied rhythm metrics have shown an interaction of *dialect*sentence*, it remains unclear to what degree these metrics reflect audible rhythmic aspects of speech [2, 3, 4, 8, 17]. It would be interesting to conduct further experiments to test whether between-sentence variability in different SwG dialects is perceptually salient.

In answer of research question (2), this study reported a number of potential triggers in the sentences material that may cause variation in the rhythm scores. By merely looking at two exemplary instances, the following dialect-specific phonological and morphological phenomena were suspected to contribute to variation in rhythm scores: SB SwG: Full vocalic articulation of unstressed light syllables, cf. [15], lengthening of short vowels [15]. BE SwG: NOM. definite article *der* is realized as [d̥ər], which increases the number of CVC syllables in this dialect (cf. Figure 2) and quite possibly results in reduced realizations of [ə].

If, hypothetically, a study uses only a small set of sentences where some sentences contain words that allow for these processes to apply and other sentences do not feature such words and consequently no such processes, dialects are likely to differ in rhythm scores in one sentence yet behave similarly in another sentence. While our preliminary findings clearly showed sentence-dependencies for between-dialect differences, these dependencies are in fact a result of dialect-specific phonological processes present in the sentence material. It is conceivable that the mentioned dialect-specific phonological processes in fact strongly contribute to between-dialect differences in rhythm scores. Such insights call for further microanalyses of the kind presented in Figures 5 & 6. In a follow-up study, one could control for features such full vocalic articulation of unstressed syllables. This would reveal the significance of such a dialect-specific phonological process for cross-dialectal comparisons in speech rhythm.

5. References

[1] Ramus, F., Nespor, M. and Mehler, J., "Correlates of linguistic rhythm in the speech signal", Cognition, 73:265-292, 1999.

[2] Prieto, P., Vanrell, M., Astruc, L., Payne, E. and Post, B., "Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish", Speech Communication, 54:681-702, 2012.

[3] Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O. and Mattys, S.L., "How stable are acoustic metrics of contrastive speech rhythm?", Journal of the Acoustical Society of America, 127:1559-1569, 2010.

[4] Arvaniti, A., "The usefulness of metrics in the quantification of speech rhythm", Journal of Phonetics, 40:351-373, 2012.

[5] Leemann, A. and Zuberbühler, L., "Declarative sentence intonation patterns in 8 Swiss German Dialects", Proceedings of Interspeech, 26.-30.10.2010, Makuhari, Japan:1768-1771.

[6] Leemann, A, Dellwo, V, Kolly, M.-J. and Schmid, S., "Rhythmic variability in Swiss German dialects", Proceedings of Speech Prosody, 21.-25.5.2012, Shanghai, PRC.

[7] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer". www.praat.org, 05.15.2012.

[8] White, L. and Mattys, S.L., "Calibrating rhythm: First language and second language studies", Journal of Phonetics, 35:501-522, 2007.

[9] Grabe, E. and Low, E. L., "Durational variability in speech and the Rhythm Class Hypothesis", in C. Gussenhoven and N. Warner Eds], Laboratory Phonology 7, 515-545, Berlin/New York: Mouton de Gruyter, 2002.

[10] Dellwo, V., "Rhythm and speech rate: a variation coefficient for DeltaC", in P. Karnowski and I. Szigeti [Eds], Language and language processing: proceedings of the 38th Linguistics Colloquium, 231-241, Frankfurt: Lang, 2006.

[11] R Core Team, "R: A Language and Environment for Statistical Computing". Version 3.0.0. http://www.R-project.org, 2013.

[12] Bates, D. M. and Maechler, M., "lme4: Linear mixed-effects models using S4 classes", R package version 0.999375-32, 2009.

[13] Baayen, R. H., "Analyzing Linguistic Data: A Practical introduction to statistics using R", CUP: Cambridge, 2008.

[14] Baayen, R. H., "languageR: Data sets and functions with 'Analyzing Linguistic Data: A practical introduction to statistics using R'", R package version 0.955, 2009.

[15] Christen, H., Glaser, E. and Friedli, M., "Kleiner Sprachatlas der deutschen Schweiz", 5th ed, Stuttgart: Huber, 2013.

[16] Knight, R. A., "Assessing the temporal reliability of rhythm metrics", Journal of the International Phonetic Association, 41(3):271-281, 2011.

[17] 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, Universität Bonn, 2010.