TEMPO DIFFERENTIATED ANALYSES OF TIMING IN POLISH

Zofia Malisz

Faculty of Linguistics and Literary Studies, Bielefeld University, Germany
zofia.malisz@uni-bielefeld.de

ABSTRACT

Two methods assessing timing patterns are used with Polish spontaneous data. Syllabic and interstress intervals are grouped into tempo categories. PVI measures are calculated for both classes of intervals and compared with the predictions of a coupled oscillator model of rhythmic variability. Coupling strengths between the syllabic and the stress oscillator are investigated across rates based on a regression analysis. Both analyses strongly support the syllabic character of timing in Polish and the usefulness of tempo differentiated approaches to timing and rhythm.

Keywords: rhythm, duration, timing, speech rate, Polish

1. INTRODUCTION

Much research on speech rhythm variability was devoted to devising and testing the predictions of rhythm metrics. Fundamentally, rhythm metrics are "formulas that seek to quantify consonantal and vocalic variability and use this quantification to classify languages rhythmically" [1]. The classification is mainly restricted to three hypothesised types: stress-, syllable- and mora-timed. The most popular indices include [10, 18] and variations thereof. Recently, the robustness of metrics has been the subject of wide scrutiny. Despite failing to promote any of the measures to the title of a complete rhythm model, subsequent criticisms provided a lot of insight into the workings of crosslinguistic timing variability.

Among others, speech rate turned out to be the factor interacting with the outcomes of classification. In [19] it was found that vocalic variability measures (ΔV and vocalic nPVI) show higher values at slower rates and lower at higher rates. The authors concluded that "rhythm values are a function of articulation rate". In [6], the effect of speech rate on %V and ΔC scores in English, French and German was investigated. Subjects spoke at a speech rate following instructions regarding tempo. The intended speech rate was compared with the actual number of syllables per second. All subjects showed a change in the measured rate as they introduced the intended rate change. The magnitude of differences in scores along the tempo changes in [6] suggested that a simple syllable rate metric alone would potentially be able to distinguish between purported rhythmic groups.

If timing types are a function of speech rate (as shown by [19]) that is being varied consistently by speakers of given languages (as shown by [6]) then it follows that degrees of both timing types can be observed in each of these languages. In fact speech rate can be treated as an order parameter that exposes the dynamical interaction of levels in a prosodic hierarchy. Therefore it should be included in models of rhythm variability.

Additionally, rhythm metrics have been criticised, by e.g. [5] and [8], for the lack of hierarchical structure: metrics effectively imply a flat metrical structure where durations of vocalic and consonantal units (or other units) unfold linearly. Similarly, [2] argued that it is the interplay between a higher prosodic unit, such as the foot, and a lower unit, such as the syllable, that negotiates particular typological patterns. Such structures would be hard to represent using just a single dimension of e.g. vocalic interval variability. Using PVI for both levels [2] demonstrated the different degrees in which the isosyllabic and isoaccentual force is exerted in a given language.

There are models of speech rhythm variability that incorporate both speech rate and hierarchical structure such as [3]. [17] developed a coupled oscillators model able to account for the syllable-timed vs. stress-timed rhythmic dichotomy posited for some languages. The model relies on the analysis of rhythmic gradation (term as defined in [16]) first made mathematically explicit by Eriksson [7]. [7] showed, in several languages, a linear relationship between the duration of an interstress interval and the number of syllables contained in it:

$$I = a + bn \quad (1)$$

where *n* is the number of syllables comprising a stress group, and *a* is constant. Languages differ in the constant term *a* which clusters around 200 for stress-timed languages and 100 for syllable-timed.

[17] proposed a model that could potentially be generalised: the interaction of syllable and stress oscillators exhibiting their own natural frequency (*eigenfrequency*) is expressed by coupling strength that varies continuously. Regression coefficients

obtained empirically by [7] are reinterpreted in continuous terms by [17]: the ratio between the point of interception to the slope is expressed by the coupling strength *r*:

(2) $r = a/b$

Slope *b* is dependent on speech rate and so systematically participates in the model. In this way, a parsimonious account on the basis of just one parameter is possible: for $r > 1$ the stress oscillator dominates and for $r \leq 1$ the syllabic one.

Polish has proven to be difficult to classify rhythmically. The present study considers hierarchical timing relations between two prosodic domains in Polish. A PVI analysis on syllables and interstress intervals is conducted as well as a regression analysis with the number of syllables as predictors of ISI duration. Both analyses are tempo differentiated.

2. METHODS

2.1. Material

A subset of DiaGest2, a Polish multimodal corpus of task-oriented dialogues [12], was used for data mining. The data in the present study come from 8 speakers of standard Polish (4 female and 4 male undergraduates) who instructed a dialogue partner in a paper folding task.

2.2. Syllabic segmentation

Syllabic boundaries were marked according to sonority principles. The Maximal Onset Principle was not used resulting in the closing of syllables in case of medial clusters such as in: "miasto"→ mias.to, "mokry"→ mok.ry.

2.3. Phrase selection

Intonation, prominence and syllabic annotations were inspected in Praat. Phrases with false starts, hesitation markers and lengthening, unintelligible speech portions, overlapping laughter etc. were omitted.

Phrase selection was based on minor Intonational Phrases (IPs) [20]. All portions of minor IPs with no discernible pause in the flow of speech were selected. Prepausal syllables annotated on the prominence tier were used as phrase boundary indicators. The procedure resulted in very fluent portions of speech, spoken on one breath and with a discernible and coherent pitch contour. Moreover, prepausal syllables were not included in phrase selection. Polish uses lengthening liberally to mark phrasal boundaries; in some varieties multiplying the duration of an average syllable by a factor of 5 [13]. This way durational effects of final lengthening were avoided. Overall, 411 phrases were selected.

2.4. Speech rate

Speech rate information was expressed in syllables per second (mean syllable length in each phrase per second was calculated). Four tempo categories were defined according to quartile ranges of phrase rate: Tempo 1 up to 6 syllables per sec, Tempo 2 between 6 and 7, Tempo 3 between 7 and 8 and Tempo 4 from 8 up to 12 syllables per second. All phrases were grouped according to the four tempo categories.

2.5. Interstress intervals

Interstress intervals (ISI) in the preselected phrases were extracted from the prominence tier. Prominence annotation was prepared by two expert annotators. Two labels were used, denoting perceptually strong and weak prominences. Phrases with at least two prominences and therefore at least one full interstress interval were considered. In utterances where a prominent syllable was noninitial the anacrusis was necessarily omitted. However, the anacrusis was included in the speech rate estimation of a given phrase and PVI analyses (subsection 3.3).

There is no consensus as to acoustic correlates of stress in Polish. It is generally considered to have perceptually weak stress that is mainly based on pitch accents and intensity peaks rather than duration, similarly to Czech, Finnish and Estonian [9, 11, 14]. Therefore it is rather difficult to systematically mark ISIs and often phonological expectations play a great role in the perception of stress. Lexical stress is placed on the penultimate syllable with few exceptions. It was observed that prominence labels largely correlated with the main pitch accented syllable in a phrase in case of strong beats and lexical stress in case of weak beats.

3. RESULTS

3.1. General rate effects

Given that phrase rates between approx. 6 and 8 syll/sec describe 50% of the data (with the overall mean at 6.9 syll/sec), it can be assumed the range defines the "normal" rate in the corpus. The results are supported by individual speaker rate means that deviate from the overall mean by only 0.4.

The mean value for speech tempo in Polish corresponds to what was empirically found by [6] to be the normal speaking rate for French subjects (normal rate range between approx. 6.3 and 7.2, overall mean = 7.3 syll/sec).

[6] explain that crosslinguistic differences in the normal speaking rates observed in their study (with

English at mean 5.9 and German 5.6) might depend on phonotactics. It is plausible that the simpler syllable structure of French allows for a more liberal management of articulation rate than e.g. English. Polish in turn, could be placed in the tempo "inflexible" group, with English or German, due to its complex syllable types and lack of vowel reduction. However, our result is supported by frequencies of complex syllable types in Polish. They are in fact rather low. In a large corpus analysed by [14], simple syllable types (CV, CCV, CVC, CCVC) described 89% of tokens. [15] suspected that the regularity of syllabic intervals in their Polish corpus (nPVI = 38) could be accounted for by a Zipf effect: the longest clusters are rare.

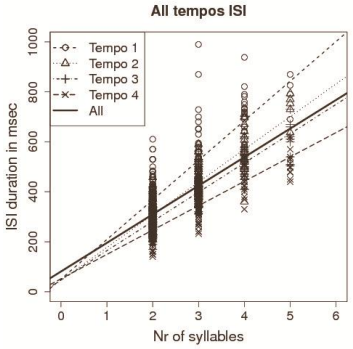
3.2. Rhythmic gradation and rate

The study of rhythmic gradation suggests that Polish has a strongly syllable dominated timing. Values are summarised in Table 1. All regressions on tempo differentiated data stay in the syllabic-timed range, lower than the overall values for Spanish (r = 0.64) given by [17]. The value of r however rises with tempo, indicating that the stress oscillator gains strength over the syllabic one with increasing rate. Correlation coefficients for the tempo-differentiated analyses are high, given the fully spontaneous data. The linear relationship between syllable number and ISI duration is expressed by $I = 90 + 111n$ for all tempos, placing the language between Spanish and Finnish (r = 0.92, after [17]) in terms of type. Fig. 1 presents the results graphically. The slope coefficient decreases with increasing speech rate.

Table 1: Results of regression analysis.

Tempo	Linear regression	Coupling strength	Corr. coeff.
T1	$I = 50 + 158n$	$r = 0.32$	$R = 0.65$
T2	$I = 45 + 130n$	$r = 0.35$	$R = 0.75$
T3	$I = 46 + 117n$	$r = 0.40$	$R = 0.77$
T4	$I = 51 + 97n$	$r = 0.52$	$R = 0.75$
All	$I = 90 + 111n$	$r = 0.81$	$R = 0.53$

Figure 1: Regression results for particular tempo groups (see legend). The black solid line denotes the linear regression model for all tempos.



3.3. PVI and rate

The Pairwise Variability Index [10], is a useful metric of durational "evenness" between two subsequent intervals. The raw PVI (Eq. 3 where m is the number of analysed units and d is the duration of a given unit k) is an averaged distance measure that can be used with different adjacent units, usually vocalic and consonantal but also syllables or feet (as in e.g. [2], normalised). The value of the index is lower when consecutive intervals are more equal in duration.

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}|/(m-1)\right] \quad (3)$$

The raw PVI formula is mostly used a) with consonantal durations known not vary considerably with speech rate and b) in order to preserve syllable structure information. In the present data, raw PVIs were used to preserve the effects of speech rate on the syllable and ISI dynamics.

Tempo differentiated raw PVIs for each type of interval yielded values presented in Table 2. Syllabic rPVI values keep falling until T3, i.e. syllabic intervals get more and more even as tempo increases from slow to normal. The differences are highly significant (paired t-test, conf.level = 0.95). The ISI PVI stays relatively level from T2 to T3 showing that the influence of syllable equalisation is stronger (the syllabic oscillator dominates). From T3 ("faster normal", 7-8 syll/sec), the ISI durations start to regularise but the difference does not reach significance. The syllabic material also reaches an adaptation limit at T3.

Table 2: Raw PVI values for syllables and ISI across tempo groups.

Tempo	ISI rPVI	Syllabic rPVI
T1	160	90
T2	132	67
T3	130	54
T4	121	54

4. DISCUSSION

[19] found that speakers not only approximate to syllable-timing in fast speech but also vary the length of their vowels more in slow speech. It seemed possible therefore, that a grouping effect of stress in slow speech and a regularising isosyllabic effect in fast speech are universal features present in all languages. Our results show that slow rate might indeed correlate with more elaborate phrasing and therefore results in higher PVI values (for equally syllables and feet) that drop around the usual tempos. What makes Polish more "syllabic" in the PVI context is that syllabic dominance is evidenced at

normal tempos. A continuing isosyllabic influence in very fast tempo was not observed.

An explanation based on the character of Polish stress can be offered. If syllable dominance was further enhanced in fast tempos, the weak impact of Polish stress (especially in the duration domain) could be in danger of disappearing, leaving the speakers without a grouping tool that structures utterances. An interesting question remains about the relationship between phonological structure and the threshold points along the speech rate parameter at which a language starts moving its timing towards one type or another.

The rPVI results are supplemented by coupling strength values obtained from regression. They increase slowly from "syllable-timed" to more "stress-timed" with tempo. However the trend is not dramatic and on the whole the measure stays within a strongly syllable dominated type of timing with values of *r* remaining well below 1. In [4], intervals between syllabic peaks in Italian were related to ISIs. The opposite trend in *r* was found with speech rate increase: from $r = 1.05$ (stress group dominated) in the slowest tempo, down to $r = 0.57$ (syllable dominated) in the fastest one. In this light, Polish remains quite inflexible in the adjustments of its structure to speech rate, with a tight grip on the coupling strength exerted by syllables on the ISIs.

5. CONCLUSIONS

Speech rate interacts with temporal structure inherently in a constrained manner. Observing variability in the speech rate dimension reveals important differences, e.g. how timing is changed within and across languages. The present approach does not preclude arriving at a predominating mode and adding new dimensions to the typological landscape. A language such as Polish with its under-researched timing patterns and disputed typological status can benefit from the use of models combining speech rate and a hierarchical prosodic structure.

6. ACKNOWLEDGEMENTS

This research was supported by Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 673 and by the Polish Ministry of Science and Higher Education, project N N104 010337.

7. REFERENCES

[1] Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1-2), 46-63.

[2] Asu, E.L., Nolan, F. 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. *Proc. of Speech Prosody 2006* Dresden.

[3] Barbosa, P.A. 2002. Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. *Proc. of Speech Prosody* Aix-en-Provence, 163-166.

[4] Bertinetto, P.M., Bertini, C. 2010. Towards a unified predictive model of Natural Language Rhythm. In Russo, M. (ed.), *Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology.* Naples: Aracne, 43-78.

[5] Cummins, F. 2002. Speech rhythm and rhythmic taxonomy. *Proc. of Speech Prosody* Aix-en-Provence, 121-126.

[6] Dellwo, V., Wagner, P. 2003. Relations between language rhythm and speech rate. *Proc. 15th ICPhS* Barcelona, 471-474.

[7] Eriksson, A. 1991. *Aspects of Swedish Speech Rhythm*. Ph.D. dissertation. University of Göteborg.

[8] Gibbon, D. 2006. Time types and time trees: prosodic mining and alignment of temporally annotated data. In Sudhoff, S., et al. (eds.), *Methods in Empirical Prosody Research*, Berlin and New York: Walter de Gruyter, 281-209.

[9] Gibbon, D., Bachan, J., Demenko, G. 2007. Syllable timing patterns in Polish: results from annotation mining. *Proc. of Interspeech/Eurospeech 2007* Antwerp, 994-997.

[10] Grabe, E., Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In Gussenhoven, C., Warner, N. (eds.), *Papers in Laboratory Phonology. vol. VII.* Berlin and New York: Mouton de Gruyter, 515-546.

[11] Jassem, W. 1963. *Akcent Języka Polskiego.* Wrocław: Ossolineum.

[12] Karpiński, M., Jarmołowicz-Nowikow, E., Malisz, Z. 2008. Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogue. *Speech and Language Technology* 11. Poznań: Polish Phonetic Association, 113-122.

[13] Karpiński, M., Jarmołowicz-Nowikow, E., Malisz, Z., Juszczyk, K., Szczyszek, M. 2008. Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych. *Investigationes Linguisticae XVI* 83-98.

[14] Klessa, K. 2006. *Analysis of Segmental Duration for Polish Speech Synthesis Purposes.* Ph.D. dissertation, Adam Mickiewicz University, Poznań.

[15] Lehiste, I. 1970. *Suprasegmentals.* Cambridge, Mass.: MIT Press.

[16] O'Dell, M.L. 2003. *Intrinsic Timing and Quantity in Finnish*. Ph.D. dissertation, University of Tampere.

[17] O'Dell, M.L., Nieminen, T. 1999. Coupled oscillator model of speech rhythm. *Proc. 14th IcPhS* San Francisco, 1075-1078.

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

[19] Russo, M., Barry, W.J. 2008. Isochrony reconsidered. Objectifying relations between rhythm measures and speech tempo. *Proc. of Speech Prosody* Campinas, 419-422.

[20] Wagner, A. 2008. *A Comprehensive Model of Intonation for Application in Speech Synthesis.* Ph.D. dissertation, Adam Mickiewicz University, Poznań.