The influence of tempo and speaking style on timing patterns in Polish

Agnieszka Wagner, Katarzyna Klessa
Institute of Linguistics, Adam Mickiewicz University in Poznan, Poland
https://doi.org/10.36505/ExLing-2015/06/0022/000259

Abstract

The study investigates the effect of different speech rates and speaking styles on timing patterns in Polish using a recently developed methodology (Annotation Pro+TGA) and a new speech corpus for rhythm research. We report on preliminary results indicating changes in syllable rates according to intended tempi (i.e. higher/lower number of syllables per second at faster/slower rates) and discuss results of regression analysis of duration differences in inter-pausal time groups, e.g. the variability of duration difference regression slopes in the three speaking styles.

Key words: timing patterns, tempo, speaking style

Introduction

The relation between timing patterns, speech rate and speech style has been investigated for many languages, but not for Polish – the exceptions are the recent studies by Malisz (2013) who investigated tempo-differentiated utterances using coupled oscillator model or by Klessa & Gibbon (2014) and Yu et al. (2014) who looked at acceleration and deceleration patterns within interpausal groups of syllables. Malisz observed the general dominance of the syllable oscillator (pointing to syllable isochrony), especially with increasing speech rate, and a tendency towards greater stress-timing at slower rates. The findings concerning acceleration and deceleration patterns showed significant differences in duration regression slopes and intercepts between read and conversational speech styles.

The goal of this study is to investigate the effects of speech rate and style variation on timing properties of Polish and to investigate the relationship between intended and laboratory measured tempo. For this purpose we apply a recently developed methodology of comprehensive, multidimensional speech timing analysis – Time Group Analysis (TGA) proposed by Gibbon (2013) and its implementation as Annotation Pro+TGA (Klessa & Gibbon, 2014) to a new speech corpus created specifically for the needs of speech rhythm research and consisting of phonotactically, prosodically and rhythmically differentiated speech material (isolated sentences, text "The North Wind and the Sun", excerpts of poems, spontaneous speech and mini-dialogues) produced by 20 native and 15 non-native speakers of Polish.

Speech material & methods

The present analyses were based on recordings of 5 native speakers of Polish reading a text (*text* subcorpus), sentences consisting of 17-22 syllables (*sent* subcorpus) and poetic verses representing different metres (*poe* subcorpus) in five different intended tempi (very slow to very fast, referred to as: *vslow, slow, norm, fast, vfast* further in the paper).

In order to investigate the possible differences in timing variability in the three types of texts and at the five different tempi, we applied the TGA approach which is based on computational annotation mining procedures and provides information on the perceptually significant properties of the temporal structure of utterances including isochrony, iteration and alternation. The main TGA parameters are the *slope* (representing speech acceleration and deceleration) and *intercept* regression values.

The speech material was annotated with Annotation Pro (Klessa et al., 2013). First, the recordings were semi-automatically labelled on the orthography level, then a GTP conversion was performed on the level of phones, syllables and words using Salian plugin for Annotation Pro. Afterwards, all segment boundaries were inspected by trained labellers, and manually adjusted in case of a need. Finally, the data were automatically segmented into inter-pausal time groups for the needs of TGA, using the Annotation Pro+TGA plugin (Klessa & Gibbon 2014).

Results

Altogether, 1265 inter-pausal groups were distinguished in the present material. The number of the groups varied among the speakers (min. average of 38 groups, max. average of 64 per speaker) and among the different tempi with more groups produced at slower than at faster rates.

The mean values of duration difference *slopes* obtained for the recordings at the five tempi for the three different speaking styles are shown in Figure 1. Generally, the slopes tended to be slightly higher at normal or slower rates than at the very fast or fast rates, depending on the speaking style. The low values of slope in the *vfast* condition (and also in the *norm* condition for the *sent* subcorpus) indicate that speakers realized this part of the recording task not only by speeding up on average, but also by applying less deceleration within the time groups. More deceleration, i.e. more dynamics in within-group timing can be found at slower rates. When comparing the speaking styles, grater differences in slopes are present in the utterances from the *poe* and *text* subcorpora than in the isolated sentences – this might be attributed to the wider range of possibilities given by longer texts or poetic verses, e.g. in terms of interpretation by individual speakers.

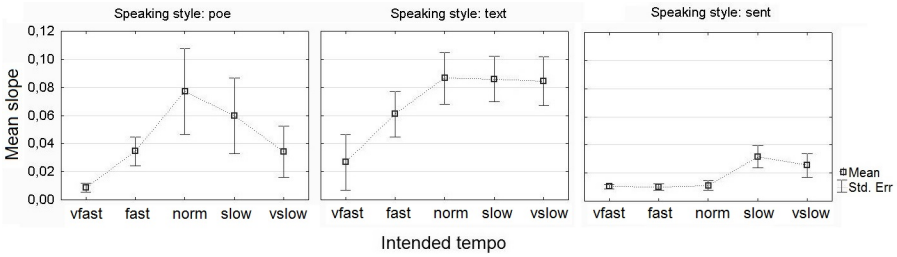

Figure 1. Mean values of duration difference slopes.

The speakers achieved the targets assumed for the five intended tempi by producing utterances significantly differing in terms of rates in these five conditions for each of the three speaking styles. The mean speaking rates (measured in syllables per second) realized by the speakers were in agreement with the assumptions of the recording scenarios, and were the fastest in the *vfast* condition and the slowest in the *vslow* condition. Interestingly, when the speaking rates were normalized by speakers and files, it was observed that on average, the rates were slightly higher in the *sent* subcorpus than in poetic verses and read text (Figure 2).

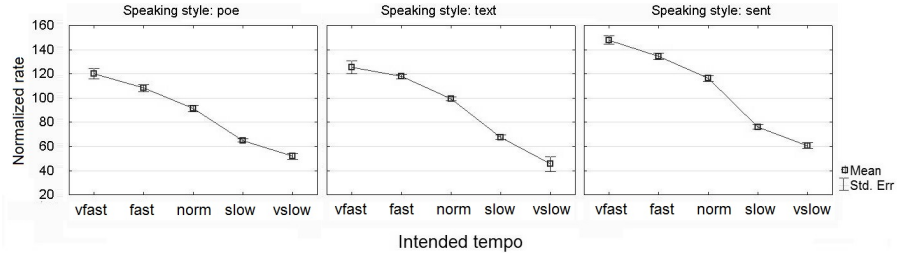

Figure 2. Mean values of speech rates (syll/sec) realized by the speakers.

The mean values of regression *intercepts* are expected to correlate with segmental durations and thus – to negatively correlate with speaking rates. This was also generally confirmed for the present data (Figure 3).

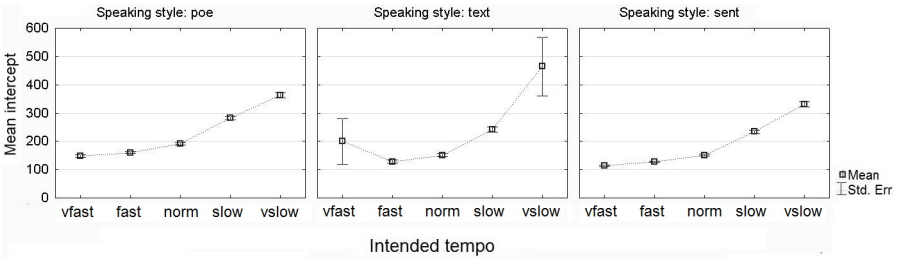

Figure 3. Mean values of duration difference intercepts.

An exception is the *text* subcorpus where the mean intercept in the very fast tempo was relatively higher than in the *fast* and *norm* conditions. However, it should also be noted that the result in the *vfast* condition had the highest standard error (and standard deviation) values, which shows more individual differentiation in this speaking style. The largest dispersion around the mean was found for the *vslow* condition, also in the *text* subcorpus.

Discussion and conclusions

The present paper provides the first insights into a new corpus designed primarily for analyses of durational and rhythmic phenomena in Polish. As it has been discussed, the variability of the actual average values of speech rates was generally in agreement with the assumed five tempo-related conditions. The intercept values were negatively correlated with speaking rates and the patterns were similar across speaking styles (except for *vfast* condition in the *text* subcorpus). The highest variability was found in mean slope values, indicating that the deceleration and acceleration patterns vary more among the speaking styles than intercepts or mean speech rates.

Due to the limited amount of speech material used, the present paper should be treated as a pilot study. Further work will include a replication of the present analyses based on the whole corpus, as well as a series of follow-up timing analyses with the use of other methods and rhythm metrics.

Acknowledgements

This work was supported by National Science Centre (NCN) grant no. 2013/11/D/HS2/04486.

References

Gibbon, D. 2013. TGA: a web tool for Time Group Analysis. Proc. of TRASP. In Hirst, D. & Bigi, B. (eds.), Proc. of TRASP, 66-69, Aix-en-Provence, France.

Klessa, K. and Gibbon, D. 2014. Annotation Pro + TGA: Automation of speech timing analysis. Proc. of LREC, Reykjavik, Iceland.

Klessa, K., Karpiński, M. and Wagner, A. 2013. Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. In Hirst, D. & Bigi, B. (eds.), Proc. of TRASP, 51-54, Aix-en-Provence, France.

Malisz, Z. 2013. Speech rhythm variability in Polish and English: A study of interaction between rhythmic levels. Unpublished PhD thesis. Adam Mickiewicz University in Poznan.

Yu J., Gibbon D. and Klessa K. 2014. Computational annotation-mining of syllable durations in speech varieties. In Campbell N., Gibbon D., Hirst D. (eds.), Proc. of the 7th Speech Prosody Conference, 443-447, Dublin, Ireland.