Pitch, speech rate and rhythm as speaking style characterizing parameters in German news presentation

Sven Grawunder[1], Hans Engert[2], Helen Kaiser[3], Stephanie Unger[4], and Ines Bose[2]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig; [2]Martin Luther University, Halle, Germany
[3]University of Trier; [4]University of Cologne, Germany
grawunder@eva.mpg.de, bose@sprechwiss.uni-halle.de

Introduction

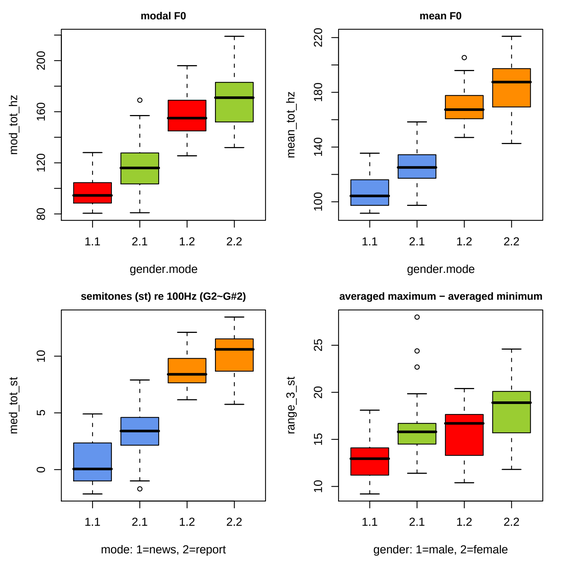
On a new corpus of German radio news (called DieNachrichtenArche) several general prosodic parameters have been investigated towards the distinction of the two presentation styles (or speaking modes) of news and report. The commonly observed *livelyhood* as it had been discussed could not just be interpreted as an expression of pitch variation but also of speaking rhythm. In order to characterize the two speaking styles more precisely, we asked whether recently introduced rhythm-oriented parameters like r/nPVI(V), r/nPVI(C), δ C or RR would be expedient to reflect such subtle differences. Since this is within a language, we reverse the question and ask: To what degree do we have to control for speaking style in order to compare languages? May we apply such measures on samples of varying text sorts. Recently a number of studies have been suggested applicability of such measures for classification [1][6][2]. What do we have to control in cross-linguistic studies?

Data Corpus

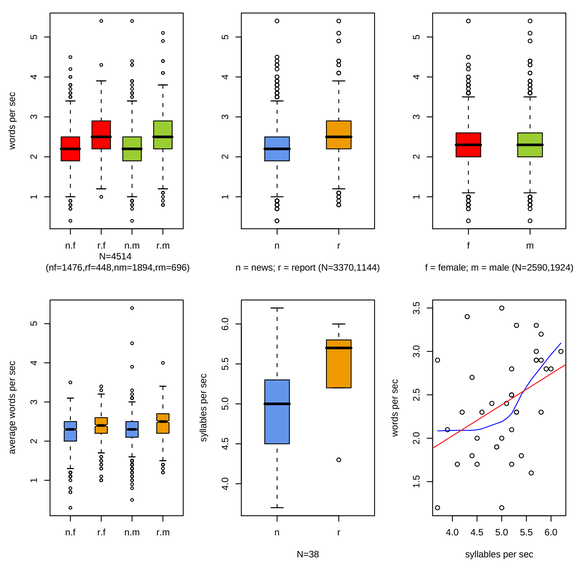
The investigated recordings originate from a corpus of German radio news. The so called *NachrichtenArche* (news archive consists of recordings of news presentations of all kinds, classical news and typical "info-radio reports" from one specific date (11th November 13h00). The aim of this collection is to document the style of news changing over time. It started in 2003 and it is hosted by the 'public' MDR broadcasting station in Madgeburg/Germany and consists currently of 160 files (one per station). It is going to be extended every year.

Pitch

Average speaker fundamental frequency and speech rate already indicate significant differences between these two very closely related but distinct text genres, in the way that reports (vs. news) show clearly higher average F0 for both genders. But also the pitch range shows in both speaker gender higher values for report.

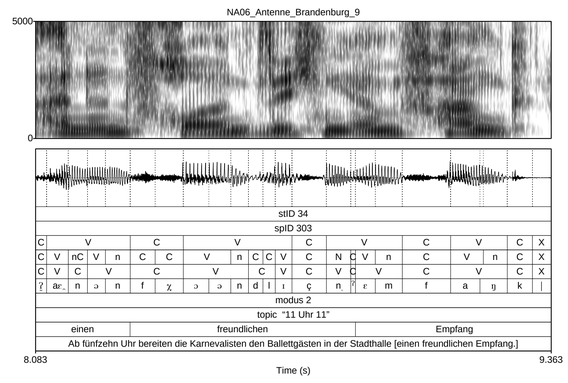


Speech rate (Word rate)

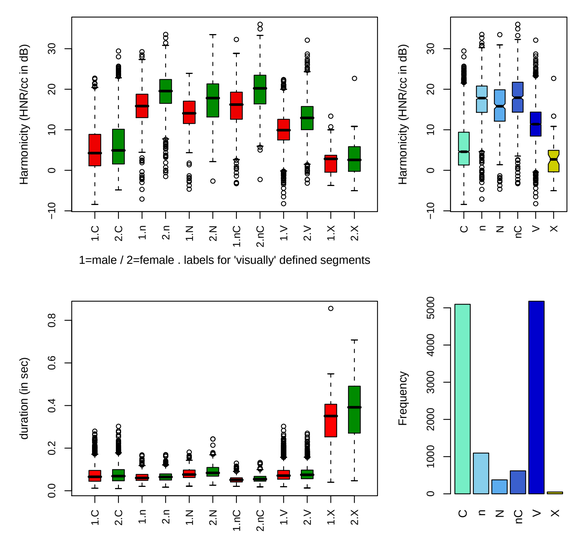


Word rate was chosen as another deliberately simple measure to assess speech rate. Besides applying this measure on the entire corpus, a small-scale sample was investigated, where we correlate words-per-seconds (wps) and syllables-per-second. Wps and its normalized measure Average-words-per-second (avwps) seems indeed to be an appropriate predictor for speech rate at this level.

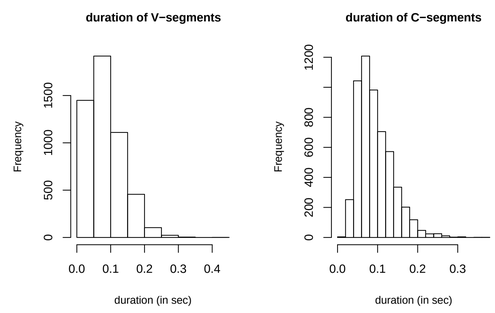
Segmentation



The current automatic segmentation setting achieves a manual-to-automatic p-alignment of p=.6, which can be further developed. The applied rhythm-oriented metric parameters (rPVI(C/V) and nPVI(C/V)) show differences between the two modes which are non-significant at a common level of α =0.05 in the counter-balanced sample.

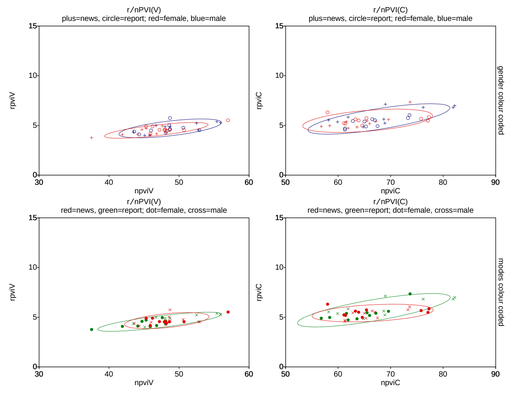


For longer voiced ('vocoid') sequences, additionally a HNR-based segmentation algorithm has been applied. As a result, we receive a gross classification which can be best described as reflecting maximum sonority differences. A parameter approximation including a simple alignment procedure was used to match the closest candidate to manually segmented test-sequences.

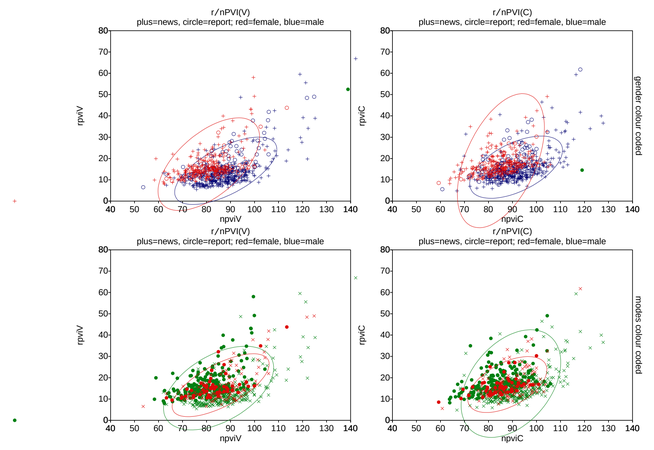


The manually segmented files provide easily accessible thresholds for the automatic algorithm.
By means of simple parameter approximation the confrontation of automatically and manually segmented files show finally a p-alignment of p=0.6 (which can be further developed).

Results for Rhythm Related Parameters



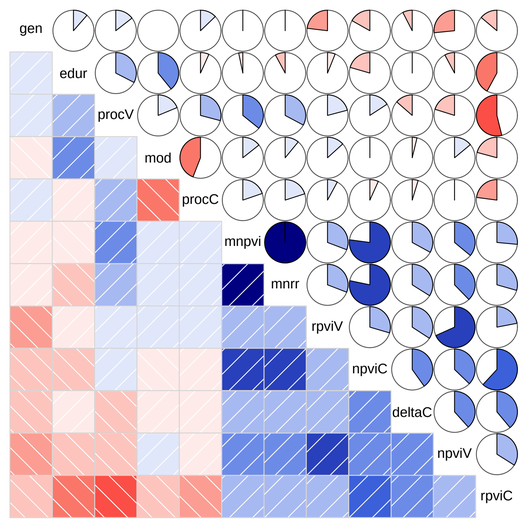
The pairwise variability index would be applied to either C or V 'segments' in its "raw" version: $rPVI = \frac{100}{(N-1)} \times \sum_{n=1}^{N-1} |d_n - d_{n+1}|$ or in the normalized form: $nPVI = \frac{100}{(N-1)} \times \sum_{n=1}^{N-1} \left| \frac{d_n - d_{n+1}}{\frac{1}{2}(d_n + d_{n+1})} \right|$, where d_n is the duration of the indexed sequence (C or V). [4]. Another algorithm regarding a general rhythm ratio [3] (recently adapted by [5]) gives a more global characteristic: $RR = 100 \times \sum_{k=1}^{m-1} \frac{d_i}{d_j}/(m-1)$ "where $d_i = d_k$ and $d_j = d_k + 1$ if d_i is smaller than d_j and $d_j = d_k$ and $d_i = d_k + 1$ if d_i is not smaller than d_j". The measure of the pairwise variability index had been applied to the entire automatically segmented corpus (N=126) and the manually segmented counter balanced sample (N=40) for gender and mode. Additionally the sample had been controlled for speech rate and duration, in the way that only those files similar in speech rate were chosen and their length was set on around 25sec depending on utterance.



Nonetheless we see tendencies at a close to significant level for rPVI(V) and rPVI(C) in the group of male speakers and for nPVI(C) in the group of female speakers. In order to control for the segmentation errors and resulting artifacts, we also confronted the two speech modes based on manually segmented samples. Here we find similar nPVI(C) in both speaking modes but lower nPVI(V) values in news presentations. Whereas male speakers show a higher nPVI(V/C) level in general (significant in news), for female speakers the nPVI(V) and nPVI(C) are slightly higher in reports. Although both modes show the same extent of outliers, the text genre of reports appears more compact than that of news.

Parameter Discussion

We were asking whether the proposed parameters could be applied to various presentation style (text sort), all labeled news presentation. Given the correlation of genre with the two other dimensions pitch and speech rate, it simply remains to state the PVI and its articulatory correlates as covariate (see correlogram). Other parameters seem to be as appropriate to reflect such small variations in segment duration.



References

[1] M. Benton, L. Dockendorf, Wenhua J., Y. Liu, and J. Edmondson. The continuum of speech rhythm: computational testing of speech rhythm of large corpora from natural Chinese and English speech. *Proceedings of the XVIth ICPhS, Saarbrücken, 6-10 August 2007*, pages 1269–1272, 2007.

[2] J. Dankovičová and V. Dellwo. Czech Speech Rhythm and the Rhythm Class Hypothesis. *Proceedings of the XVIth ICPhS, Saarbrücken, 6-10 August 2007*, pages 1241–1244, 2007.

[3] D. Gibbon and U. Gut. Measuring speech rhythm. *Seventh European Conference on Speech Communication and Technology*, 2001.

[4] E. Grabe. Variation Adds to Prosodic Typology. *Speech Prosody 2002, International Conference*, 2002.

[5] H.L. Jian. An Improved Pair-Wise Variability Index for Comparing the Timing Characteristics of Speech. *Eighth International Conference on Spoken Language Processing*, pages 1261–1264, 2004.

[6] I. Steiner. Zur Rhythmusanalyse mittels akustischer Parameter. Ma, University of Bonn, Germany, 2004.

Printed in Universitätsrechenzentrum Leipzig