Annotation and annotation mining tools for analyzing speech prosody in the Polish-German Borderland database

Katarzyna Klessa, Maciej Karpiński, Brigitte Bigi

Annotation and annotation mining tools for analyzing speech prosody in the Polish-German Borderland database

Katarzyna Klessa[*], **Maciej Karpiński**[*], **Brigitte Bigi**[†]

[*]Adam Mickiewicz University in Poznań
The Institute of Applied Lingustics, The Department of Multimodal Communication
Al. Niepodleglości 4, Poznań
{klessa,maciejk}@amu.edu.pl

[†]Aix-Marseille Université
Laboratoire Parole et Langage, CNRS
5 avenue Pasteur, 13100 Aix-en-Provence
brigitte.bigi@univ-amu.fr

Abstract

In this demo presentation, we discuss tools and techniques developed within Borderland: a project dedicated to the analysis of paralinguistic phenomena in the conversations of Polish and German teenagers, entitled: *Language of Boundaries and Boundaries of Language* (see more at: `http://borderland.amu.edu.pl/`).

1. Introduction

The Borderland database recordings were collected in the area of Frankfurt (Oder) and Słubice. The data management system and annotation specifications were discussed in an earlier publication (Karpiński and Klessa, 2018).

In the present contribution, we focus on the work aimed at supporting the investigation of speech prosody. We present and discuss software tools that can be used to automatize certain steps in the process of both describing speech recordings as well as annotation mining and data analysis.

Some of these tools have already been used for gesture analysis with the data coming from the same corpus. An example is the Re-Occurrence plugin for Annotation Pro (Karpiński et al., 2018) that enables the calculation of the number of occurrences of an annotation label found in one annotation layer (e.g., labels for Speaker 1) in another annotation layer (e.g., labels for Speaker 2). The number of re-occurring segments is calculated within the segments appearing after the end boundary of the original segment. Another example is the SRMA (Segment Rate Moving Average) plugin, used earlier with different sets of data. This plugin makes it possible to calculate the number of segments or measure segment durations within a moving frame window (see also e.g. Kousidis et al. 2010). One of the applications of the SRMA plugin is to investigate communicative alignment in the time domain (Karpiński et al., 2014).

In order to carry out similar analyses with pitch data, additional steps must be involved. In case of direct pitch measurements, we use the Probabilistic YIN (PYIN) algorithm (Mauch and Dixon, 2014) implemented in Annotation Pro, the smoothing in Praat (Boersma and Weenink, 2014) or spline function based on pitch contour stylisation in Momel (Hirst, 2011) in order to get rid of micro-changes that are irrelevant to perception and facilitate local minima and maxima detection. In further steps, for some analyses, pitch frequency is normalised in order to safely compare voices of different average pitch (e.g. male and female). With automatically added pitch labels (Hirst, 2011) as in INTSINT, more annotation mining and analyses can be done, including pitch pattern (represented as a sequence of labels) comparison and sequential analysis. Momel and INTSINT data can be obtained using SPPAS (Bigi, 2015) integrated speech analysis system. The numerical data are explored for convergence between the speakers using regression analysis while symbolic data are analysed for co-occurrence and re-occurrence of symbol sequences using the abovementioned SRMA and Re-Occurence plugins.

The starting point for many prosodic analyses is the segmentation and labelling of speech recordings. Several tools have been tested at the preliminary stage for both Polish and German data,e.g., (Kisler et al., 2016; Koržinek et al., 2017; Szymański and Grocholewski, 2005; Wypych et al., 2003). Finally, we decided to use Annotation Pro + CLARIN-PL Align (Klessa and Koržinek, 2019) for Polish and SPPAS (Bigi, 2015) for German. Both of the tools are:

(a) freely available for research,

(b) dedicated to desktop off-line mode uses.

The latter is particularly important with the Borderland data, as the recorded persons are children which restricts the access rights, hinders sending the data to external servers etc. Furthermore, the recording sessions resulted in long (or very long) media files that could be problematic for on-line processing. The reason to use Annotation Pro + CLARIN-PL for Polish (although SPPAS also support the alignment of Polish) was the fact that the whole audio annotation work had been done in Annotation Pro, and thus this choice was the simplest solution.

In order to enhance the processing of the Borderland German recordings and their annotations, a phone and word level time-alignment module of SPPAS has been developed. The pronunciation dictionary used in SPPAS has been manually inspected for the present purpose by a

NARODOWY PROGRAM ROZWOJU HUMANISTYKI

team of German philologists (see acknowledgements). The data annotated with SPPAS are compatible with Annotation Pro as the import / export options were implemented for the .ant (Annotation Pro) and .xra (SPPAS) file formats. Therefore, the SPPAS output can subsequently be used for further analyses conducted with any of the two tools, e.g. with the SRMA or Re-Occurence plugins describe above in this demo abstract.

2. Acknowledgemnents

3. References

Bigi, Brigitte, 2015. SPPAS-multi-lingual approaches to the automatic annotation of speech. *The Phonetician*, 111(ISSN: 0741-6164):54–69.

Boersma, Paul and David Weenink, 2014. Praat: Doing phonetics by computer (ver. 4.3.14) [computer program]. (Last viewed Dec. 2017).

Hirst, Daniel, 2011. The analysis by synthesis of speech melody: from data to models. *Journal of speech Sciences*, 1(1):55–83.

Karpiński, Maciej, Agnieszka Czoska, Ewa Jarmołowicz-Nowikow, Konrad Juszczyk, and Katarzyna Klessa, 2018. Aspects of gestural alignment in task-oriented dialogues. *Cognitive Studies| Études cognitives*, (18).

Karpiński, Maciej and Katarzyna Klessa, 2018. Methods, tools and techniques for multimodal analysis of accommodation in intercultural communication. *Computational Methods in Science and Technology*, 24(1):29–41.

Karpiński, Maciej, Katarzyna Klessa, and Agnieszka Czoska, 2014. Local and global convergence in the temporal domain in Polish task-oriented dialogue. In *Proc. 7th Speech Prosody Conference*. Dublin.

Kisler, Thomas, Uwe D. Reichel, Florian Schiel, Christoph Draxler, and Bernhard Jackl, 2016. BAS Speech Science Web Services-an update of current developments. In *Proceedings of LREC 2016*. Portorož: European Language Resources Association (ELRA).

Klessa, Katarzyna and Danijel Koržinek, 2019. Annotation Pro + CLARIN-PL Align: automatic segmentation and transcription module for desktop uses. In *Proceedings of 2nd Language & Technology Conference*. Poznań.

Koržinek, Danijel, Krzysztof Marasek, Łukasz Brocki, and Krzysztof Wołk, 2017. Polish read speech corpus for speech tools and services. In *Selected papers from the CLARIN Ann. Conf. 2016, Aix-en-Provence, 26–28.10.2016, CLARIN Common Language Resources and Technology Infrastructure*, number 136. Linköping University Electronic Press.

Mauch, Matthias and Simon Dixon, 2014. pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.

Szymański, Marcin and Stefan Grocholewski, 2005. Transcription-based automatic segmentation of speech. In *Proceedings of 2nd Language & Technology Conference*. Poznań.

Wypych, Mikołaj, Emilia Baranowska, and Grażyna Demenko, 2003. A grapheme-to-phoneme transcription algorithm based on the SAMPA alphabet extension for the Polish language. In *Proc 15th ICPhS*. Barcelona.