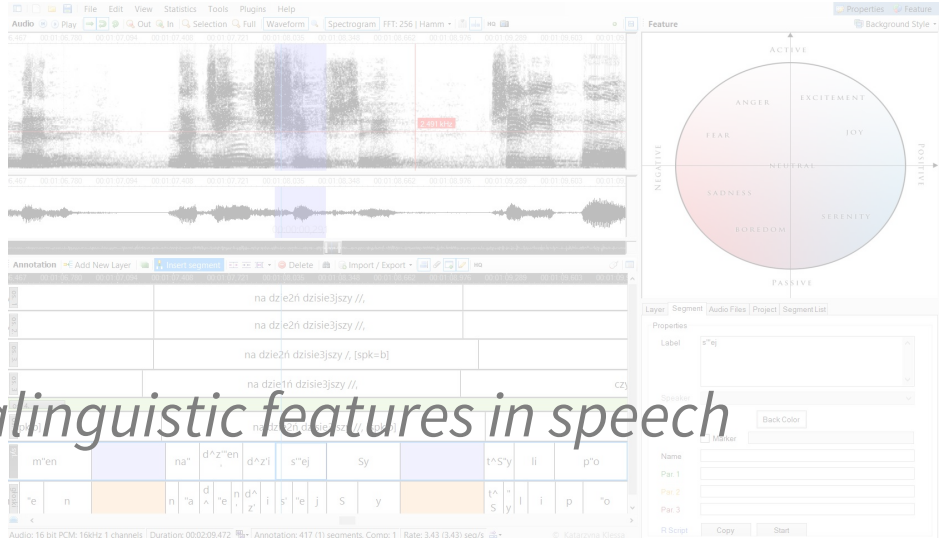
Annotation Pro

annotation of linguistic and paralinguistic features in speech

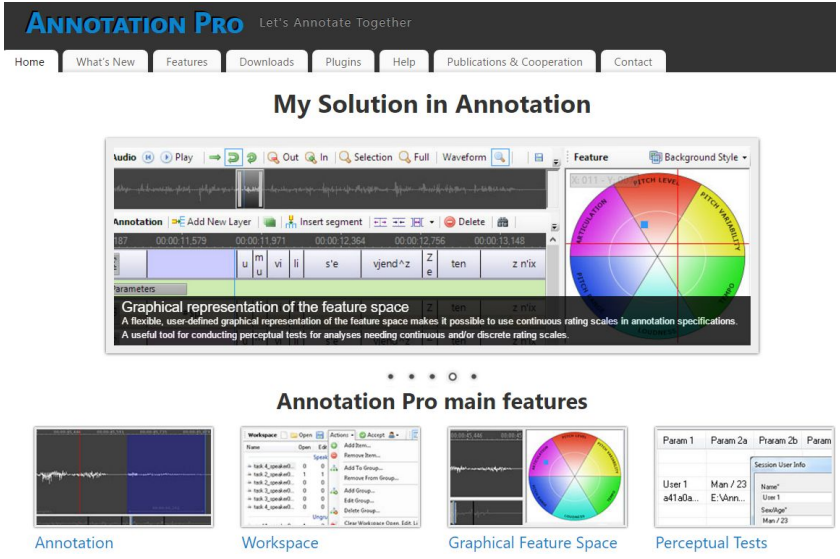


Katarzyna Klessa

Phon&Phon meeting

Faculty of English, AMU Poznań, 25 April 2017

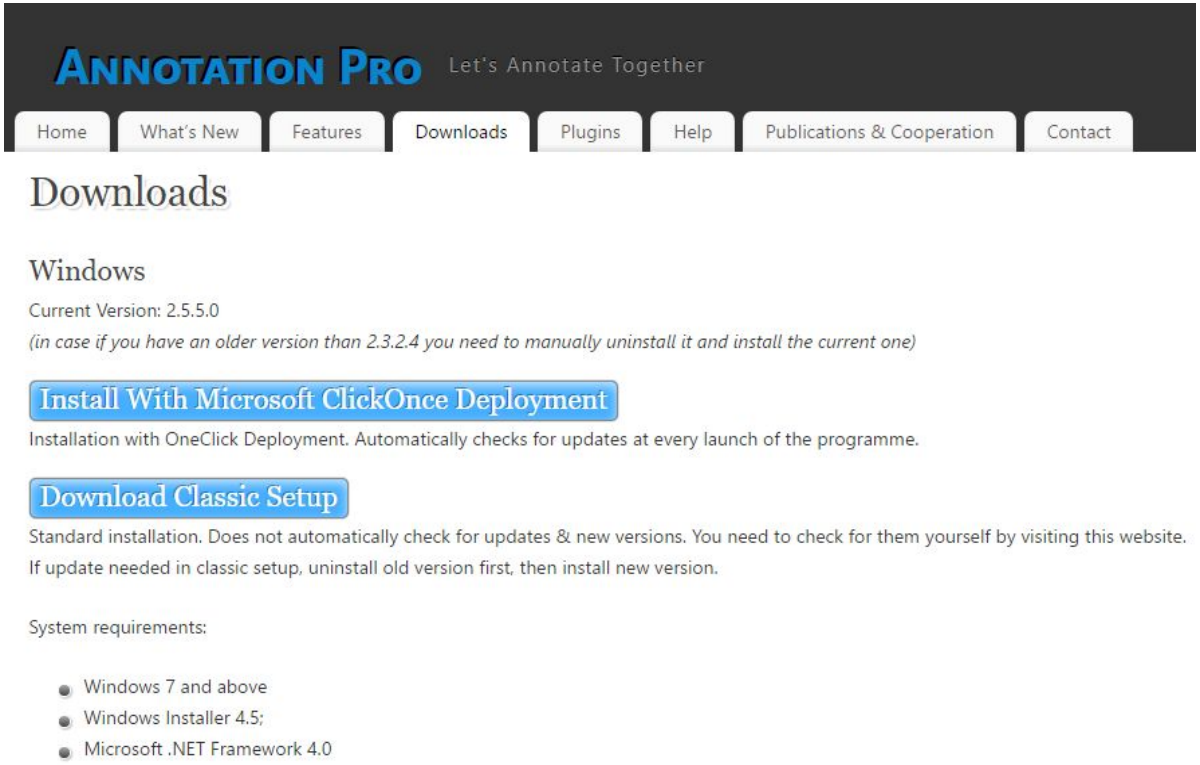
annotationpro.org



More information:

- Quick Start page
- User's Manual @ AMUR
- Other related publications & slides
- This presentation (katarzyna.klessa.pl -> Lectures&Courses)
- Sample files

Download & installation: annotationpro.org/downloads/



Motivation & background

Annotation & annotation mining - some of the earlier experiences:

- conversational speech data (PoInt, Karpiński, 2002; Francuzik et al., 2003) (Praat & dedicated database management system, annotating intonation)
- speech synthesis corpora (Demenko et al., 2010) (Wavesurfer & automatic time-alignment, GTP software, dedicated annotation editor for BOSS format)
- very large speech recognition corpora, including authentic emotions in speech (Klessa & Demenko, 2009) (Wavesurfer, Transcriber -> not supported any more, time-alignment, GTP software, relational database for annotation management, SPEECON specifications)

Rating scales. How to annotate emotions?

- Categories or continua?
- More levels & continuous rating scales reported to be more useful for non-canonical emotions / features while categories found more useful with the canonical emotions.
- Continuous rating scales -> possibility to use the output values as (quasi) continuous variables which can be useful for statistical analyses.

Fig. 1. Screenshot of the modified Feeltrace interface.

from: Metallinou, A., & Narayanan, S. (2013)

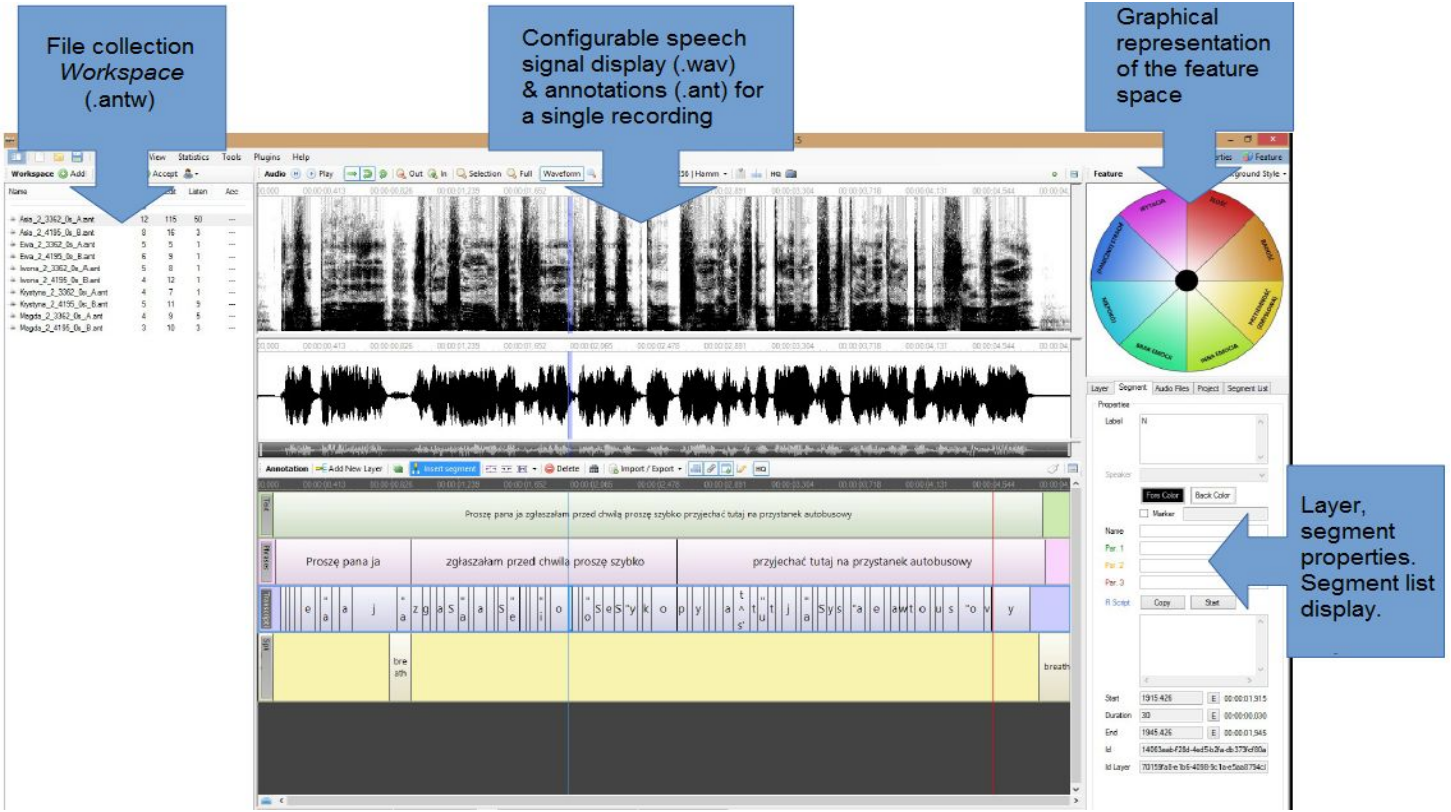
For further discussions & comparisons see eg.:
Cowie & Cornelius, 2003, Clore et al., 1987, Laukka, 2004, Arnold et al., 2012.

Motivation & background

Emerging challenges:

- ★ annotation of paralinguistic features in speech - a need to use **not only categorial but also continuous rating scales**
- ★ a need to integrate **various types of information within one framework** (multilevel annotations of both linguistic & paralinguistic features)
- a need to use many tools within one project: interoperability needed
- everyday work ergonomics, non-standard user interfaces and many other “small” issues
- ….

Annotation Pro: main window



Speech annotation tasks

Annotation seen as more than just time-aligned transcriptions

- Any number of layers available for time-aligned annotations
- Possibility to use various types of specifications using both discrete and continuous rating scales;
- Keyboard shortcuts & extensive navigation support (several zoom types, flexibility in segment editing, relocating);
- Annotation mining support;
- Data export and conversion to many external formats.

Multilayer annotations of speech corpora

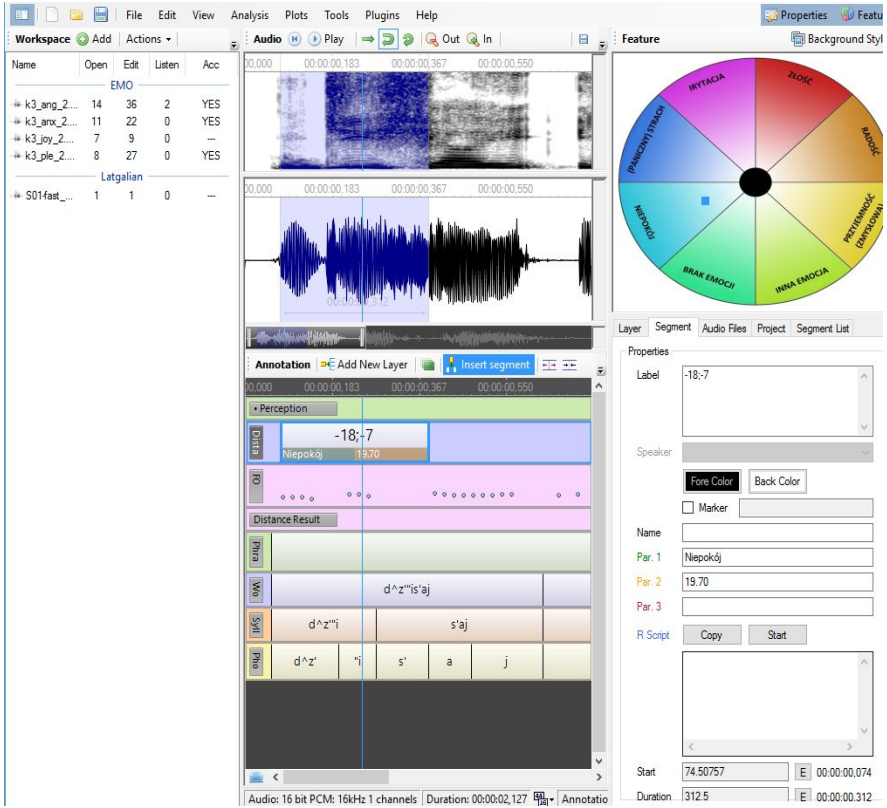
Annotations including time-aligned transcriptions on the levels of phrases, words, syllables & phones (orthographic and phonetic alphabets) as well as additional tags referring to linguistic or paralinguistic features of utterances (emotions, silent/filled pauses, non-speech events).

Example uses of Annotation Pro for corpus annotation include:

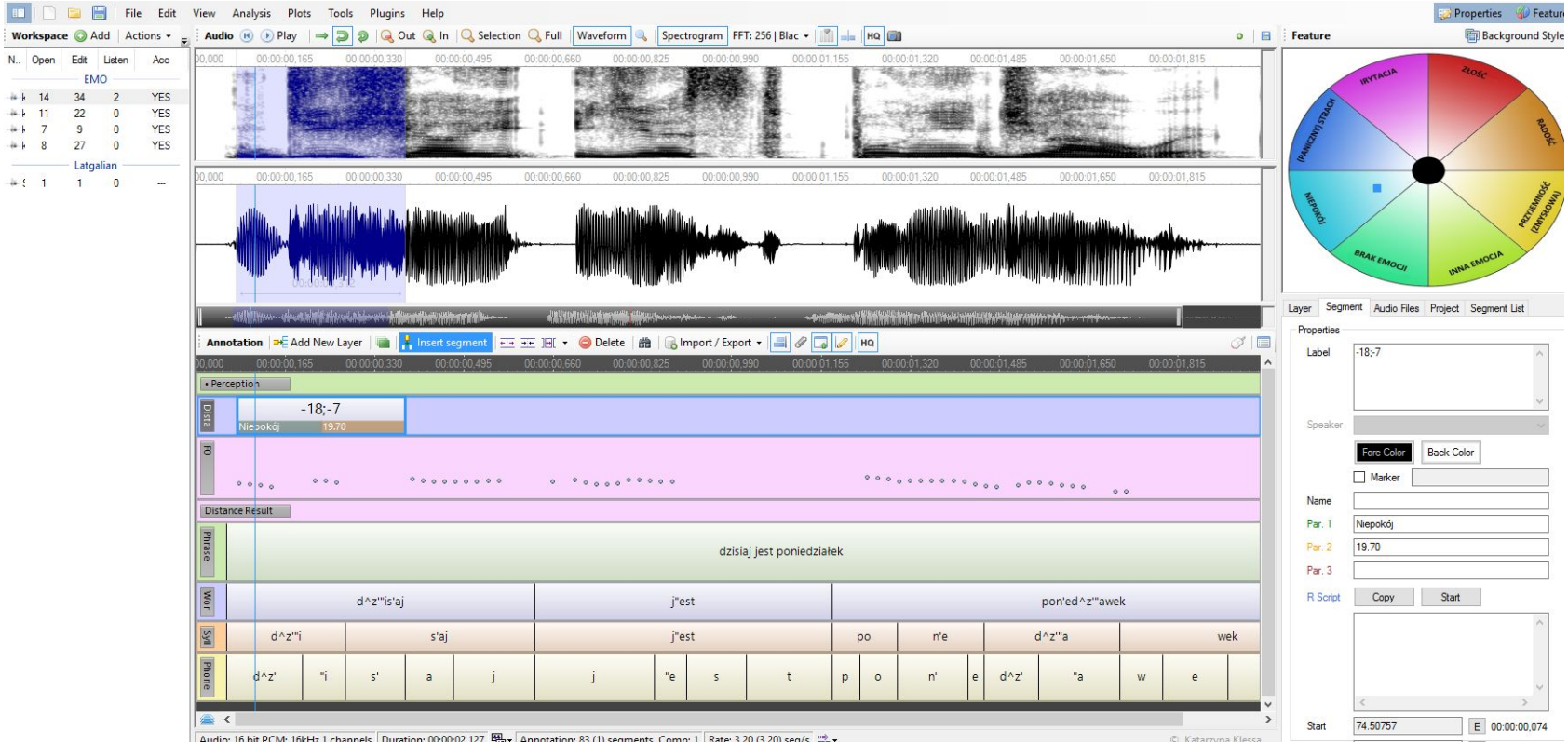
- *Paralingua* corpus (Klessa et al., 2013)
- A small corpus of Latgalian readings in varied speaking rates (Klessa et al., 2017)
- Polish timing & duration database (Wagner et al., 2016)
- *Borderland* corpus of Polish & German conversational speech (Karpiński & Klessa, in press)

Annotation layers as an analysis workspace

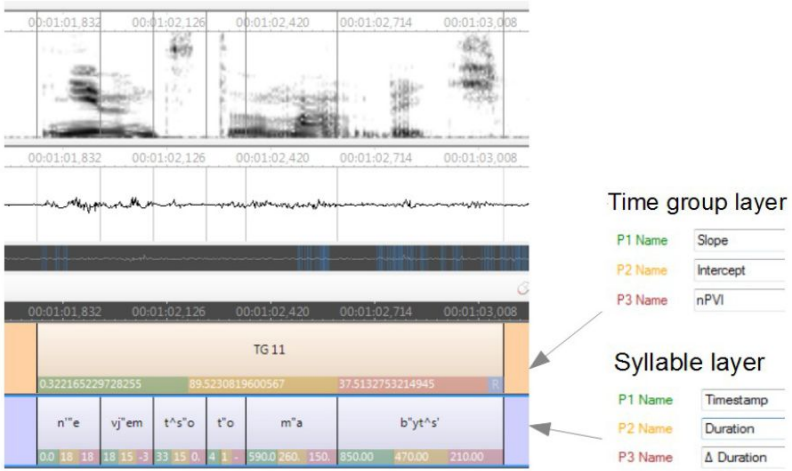
- Annotation layers (tiers) can include not only the time-aligned transcriptions of speech and non-speech events as well as the results of perception tests & annotation data processing and additional tags.
- Thanks to this solution, it becomes easily possible to simultaneously analyse various types of input within one workspace.
- Possibility to create plots based on layer contents or parameters.



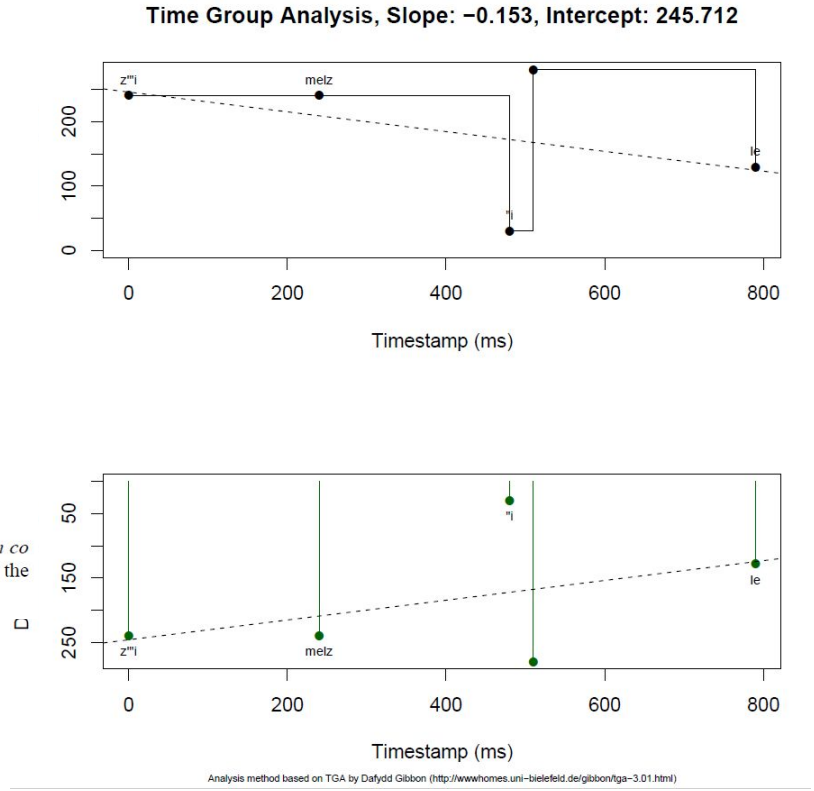
Annotation layers as an analysis workspace: EMO



Annotation layers as an analysis workspace: TGA

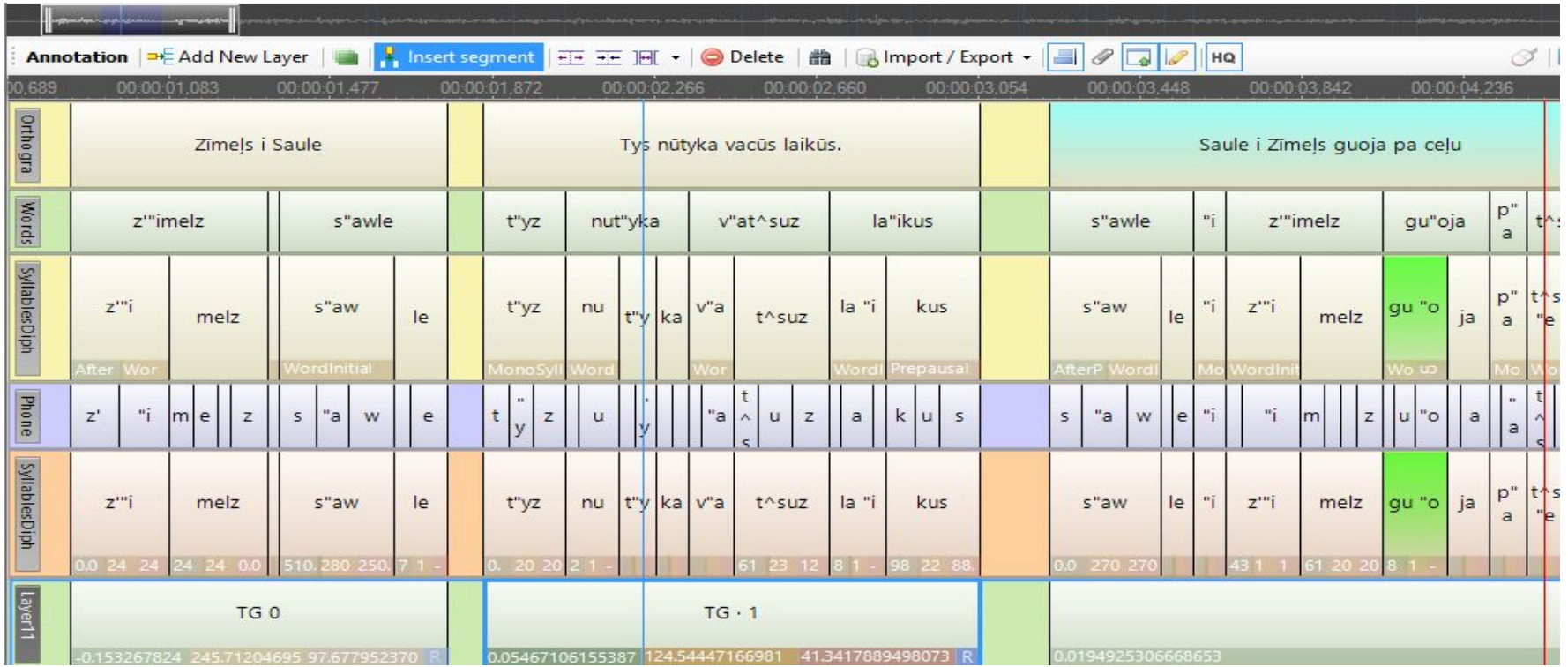


Example *TGA* results as displayed in the *Annotation Pro* interface for a Polish utterance: *Nie wiem co to ma być*. The obtained values are visible as segment parameters (P1, P2, P3) with coloured background at the bottom of each segment (a time group segment or a syllable segment respectively).



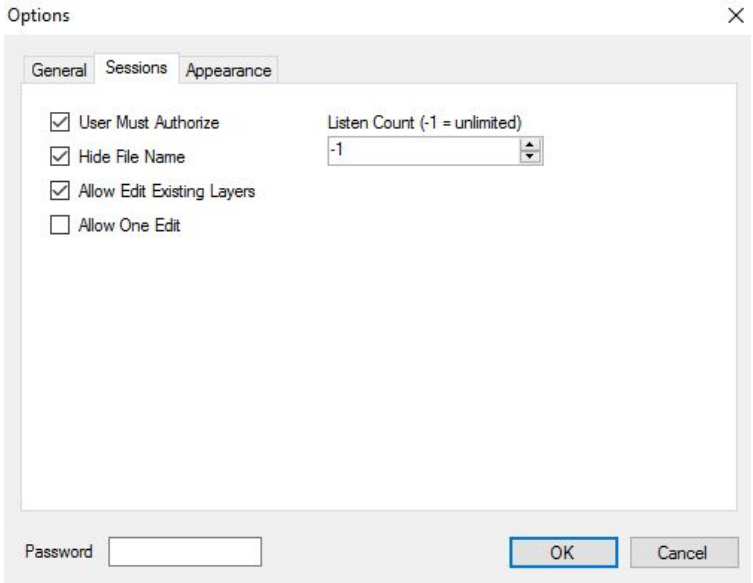
Figures from Klessa & Gibbon (2014)

Annotation layers as an analysis workspace: LTG



Annotations and analysis workspace, a sample from Latgalian read speech corpus, Klessa et al., 2016

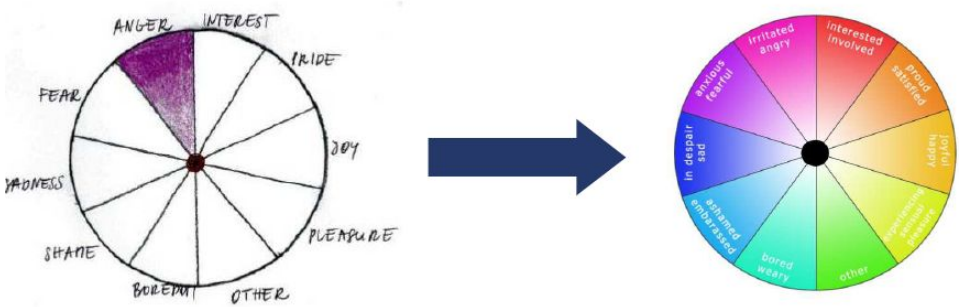
Listening tests

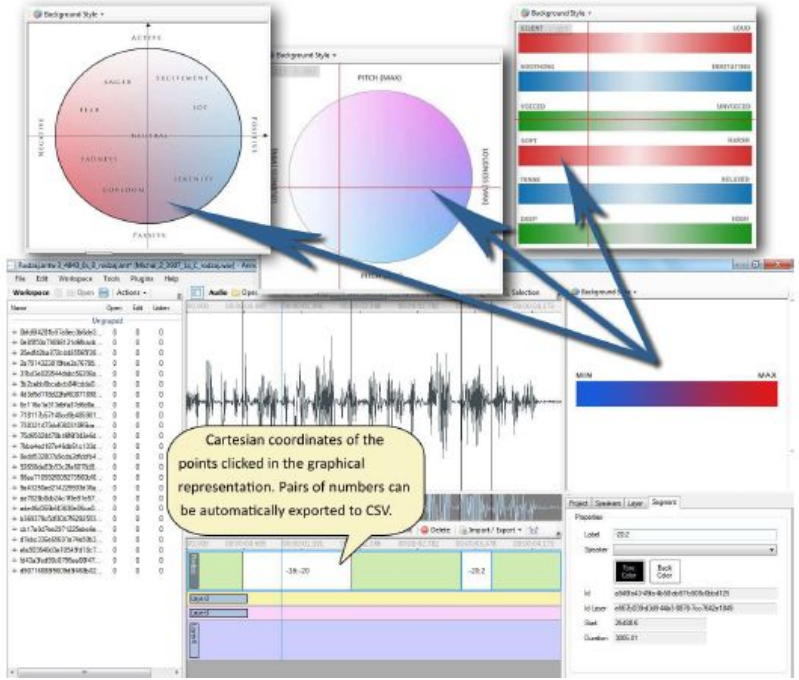


Perception test mode - options:

- Use any type of rating scales;
- Show / hide file names;
- Randomize file order;
- Collect participant information;
- Save information about participant activities (listen, open, close) in annotation layers & additionally in special report file.

Graphical representations of the feature space





- A set of built-in graphical representations & a possibility to create one's own pictures as .JPG or .PNG files
- Listeners click on the pictures -> results are saved as *x, y* coordinates of the points clicked (Cartesian coordinate system)

Perception of speaking rate

- Set-up: 6 speakers, 2 intended speech tempi, 23 listeners, continuous rating scale (*min-max* tempo).
- Are there two or more perceived rates?
- Do speakers tend to differentiate more between slow and normal rates than between normal and fast rates?

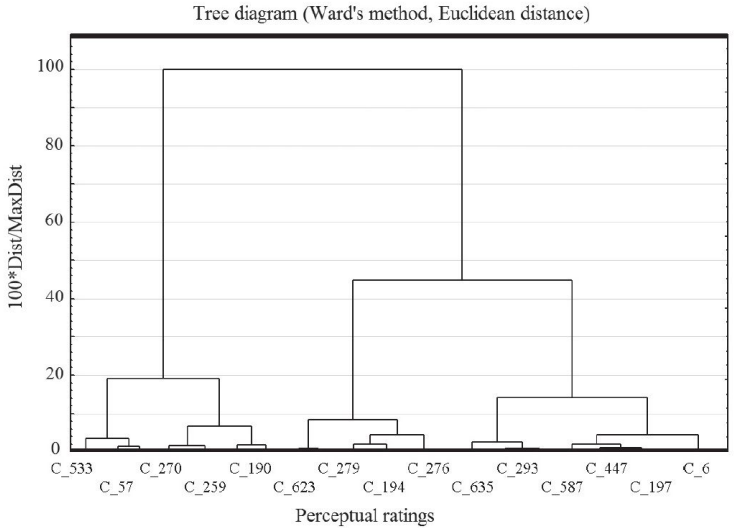


Figure 3: Cluster analysis of the perception test results: agglomeration tree diagram

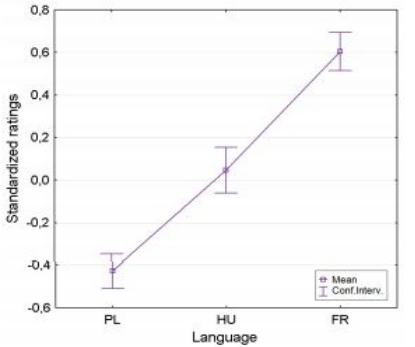
Mean for cl.1	Mean for cl.2	Mean for cl.3
23.5961	53.5356	78.5576

Gibbon et al. (2014).

Native or foreign?

- NeuroPerKog corpus of infant-directed and adult-directed speech.
- Purpose: the study of infant speech perception and development
- Perception tests with adult participants

French, Hungarian, Polish
56 syllables, 2 repetitions, random order. 6 speakers (2 per language) 16 Polish native listeners.
Continuous rating scale: Polish-foreign as shown in the Annotation Pro screenshot (right column of the poster). The middle of the scale = lack of any certainty. Listeners could re-play sound signals.

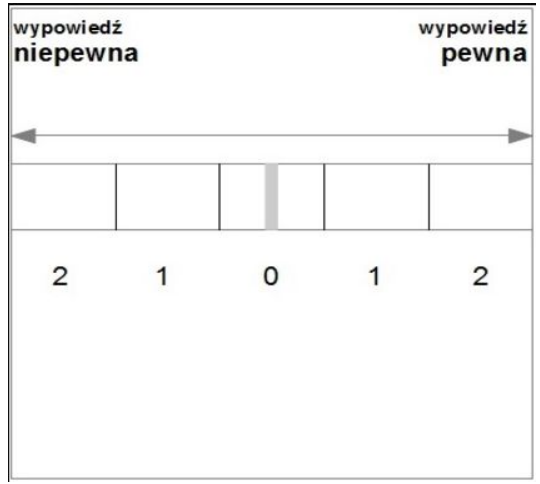


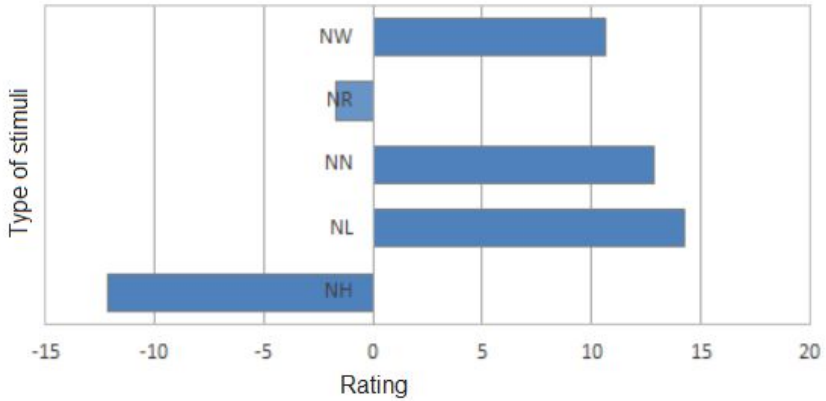
Polish, Hungarian, French syllables: PL or foreign? Means of ratings categorized by language

Klessa, Karpiński & Czoska (2015).

Prosody of (un)certainty

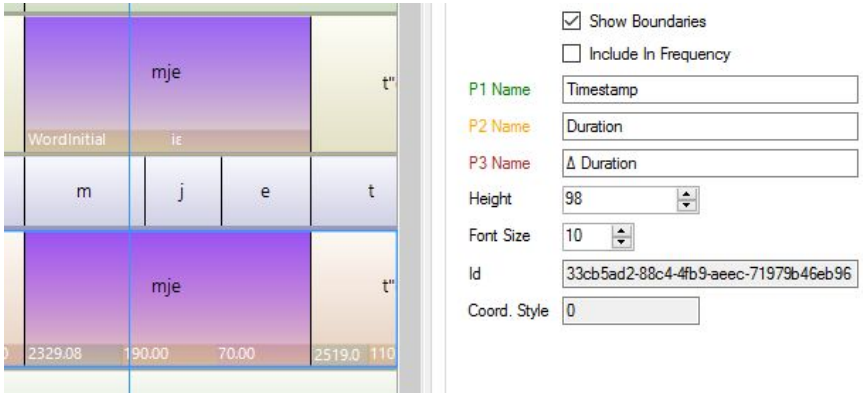
- Do global prosodic parameters of an utterance influence the assessment of the degree of uncertainty of that utterance?
- Set-up: 70 signals, neutral utterances, manipulated f0 range, F0 level, speaking rate
- Smaller differentiation in pitch contours (NR) and utterances with higher f0 (NH) are perceived as less certain.





Karpiński & Klessa (2015).

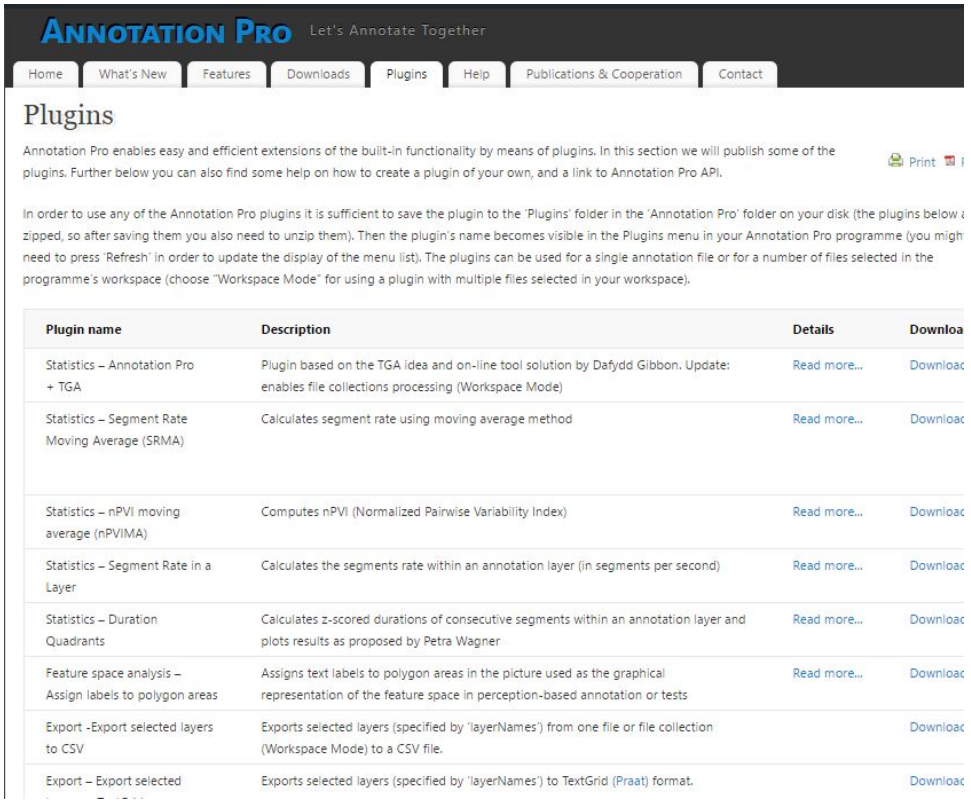
Annotation mining & interoperability



- Extracting information from the annotation, data processing:
  - export layers, files or file collections to spreadsheets;
  - automatic annotation mining with plugins (using annotation labels as input for calculations or label processing).
- Import/export formats: possibility to analyse data produced with other tools

Manual & automatized annotation mining

- Visual inspection of annotation layers (colours, chart displays), easy editing (layer names, ordering...)
- Workspace = file collection
- Plugins: bulk operations on individual files and for the whole workspace - a number of plugins available for downloads & modifications (and *Plugins* menu) http://annotationpro.org/plugins/



Temporal convergence: Paralingua - DiaGest 2

- Two corpora initially annotated in different tools (Annotation Pro. ANT native files, **ELAN .EAF format**).
- Factors: mutual visibility, speaker gender, task symmetry.
- Stronger convergence for more homogeneous stretches of speech in mutual visibility condition, weaker effects for

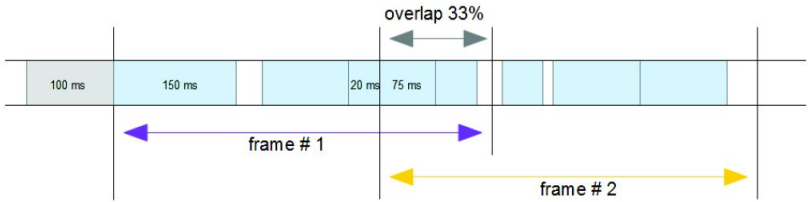


Figure 1. Segment rate moving average scheme

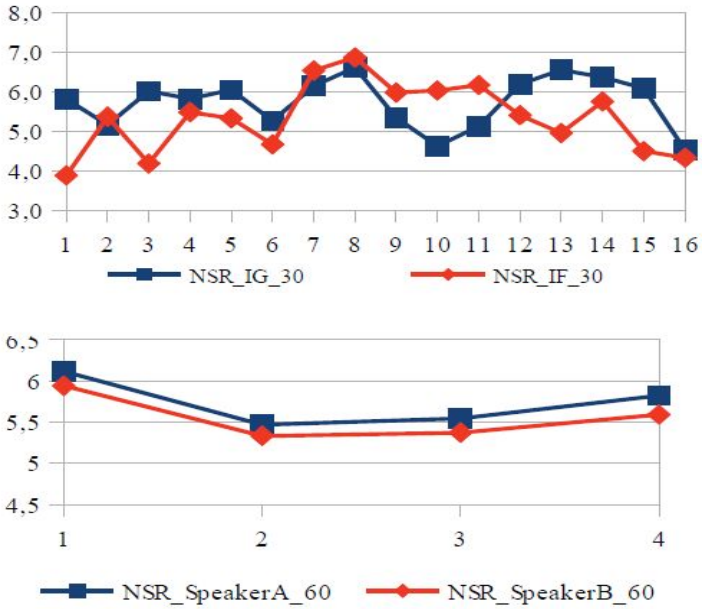
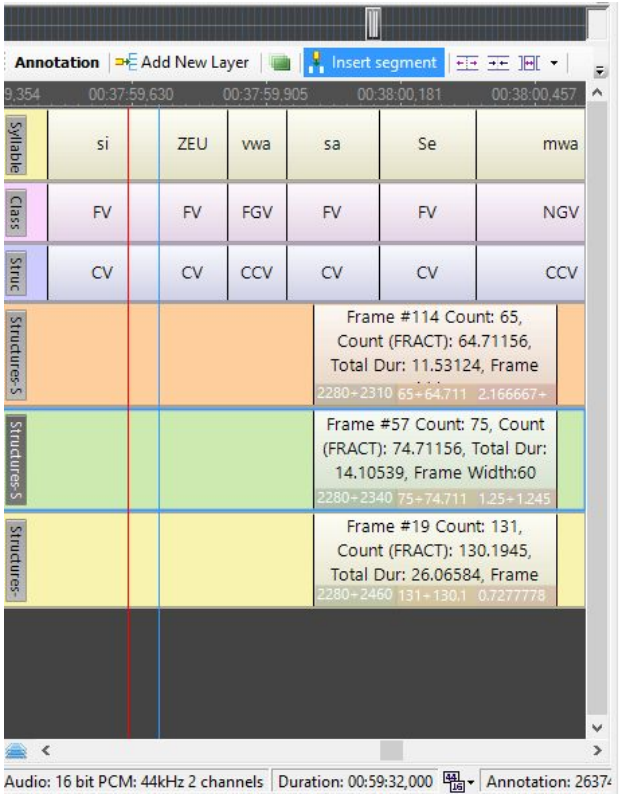


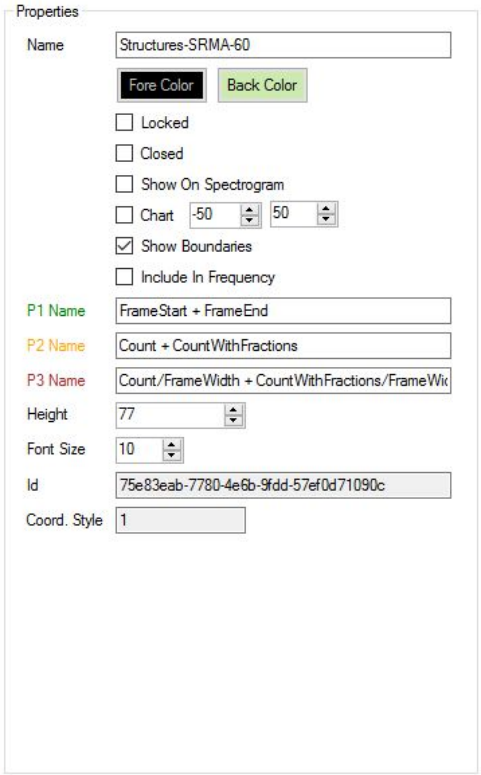
Figure 3. *NSR* changes in dialogue 6 from DiaGest2 corpus (30 second window; upper panel) and in dialogue 7 from Paralingua (60 second window; lower panel)

Karpiński et al. (2014).

Temporal convergence: CID (on-going)

- Annotations automatically **imported from Praat TextGrids**
- Spontaneous dialogs by 16 native speakers of French (CID-DISP, 2015).
- Approx 1h per dialog.
- SRMA measurement.
- Is there convergence?
- Friends tend to converge less than other interlocutors

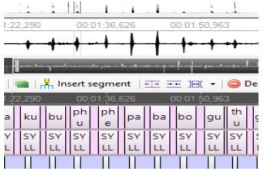




Building NeuroPerCog stimuli corpus (on-going)

E-prime output file (.CSV table format with labels & time stamps)

Wave file (a list of isolated utterances in IDS or ADS register)

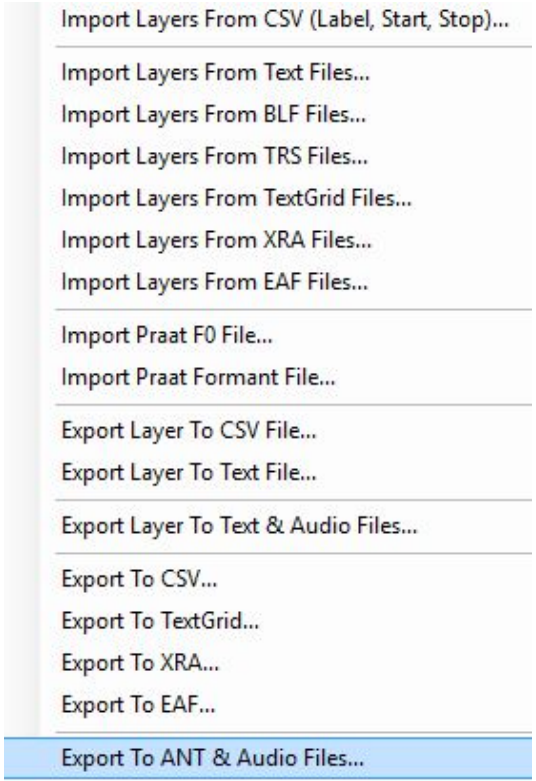


Import to Annotation Pro .ANT layers & adjust annotations

WAVE + ANT export of stimuli with meta data based on annotations

Photo: courtesy of Magdalena Oleśkowicz-Popiel

Export / import formats



- Import / export to / from the major speech annotation tools:
  - **Praat**
  - **ELAN**
  - **Wavesurfer**
  - **Transcriber**
  - **SPPAS** (automatic time-alignment, PL beta)
- Text or CSV formats;
- Text annotation software (TypeCraft);
- Possibility of integrated use of Salian (Szymański & Groocholewski, 2005) & Polphone (Demenko et al., 2003).

Conclusions

- Annotation Pro is **freely** available for research
- It can be used for speech annotation based on **both continuous and discrete** rating scales
- It makes it possible to conduct basic **perception experiments**
- Both multilayer annotations of linguistic and para/non-linguistic features as well as perception test results can be stored **within one annotation** framework
- Automatic **annotation mining** is supported by plugin architecture, and a number of plugins are available for downloads and modifications (more to come)
- The program can serve as a simple **annotation file format converter** thanks to import/export options
- Positive feedback with regard to user **interface**, comfort of use and ergonomics

Thank you

klessa@amu.edu.pl

annotationpro.org

annotationpro.org/documentation/

annotationpro.org/plugins

Related publications

- Arnold, D., Wagner, P., & Möbius, B. (2012). Obtaining prominence judgments from naïve listeners–Influence of rating scales, linguistic levels and normalisation. *Proceedings of Interspeech 2012*.
- Clore, G. L., Ortony, A. & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53*, 751-766.
- Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40*(1-2), 5–32.
- Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schröder, M. FEELTRACE: An instrument for recording perceived emotion in real time. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, Newcastle.
- Demenko, G., K. Klessa, M. Szymański, S. Breuer, W. Hess (2010). Polish unit selection speech synthesis with BOSS: extensions and speech corpora. [in:] *International Journal of Speech Technology: Volume 13, Issue 2*, Page 85.
- Francuzik (Klessa), K., Karpiński, M., & Klesta, J. (2002). A preliminary study of the intonational phrase, nuclear melody and pauses in Polish semi-spontaneous narration. In *Speech Prosody International Conference*. Aix-en-Provence.
- Karpiński, M. 2002. The Corpus of the Polish Intonational Database: Technical Specifications. *Investigationes Linguisticae, vol. VIII*.
- Karpiński, M. & Klessa, K. (2015). *Prozodia niepewności*. In Danielewiczowa M., Bilińska J., Doboszyńska-Markiewicz K., Zaucha J. (Eds.). Sens i brzmienie, z serii: *Prace Językoznawcze Instytutu Filologii Polskiej UKSW (tom 7)*, s. 49-63. Wydawnictwo Uniwersytetu Kardynała Stefana Wyszyńskiego. Warszawa. ISBN 978-83-8090-032-5.
- Karpiński, M., Klessa, K., Czoska, A. (2014). Local and global convergence in the temporal domain in Polish task-oriented dialogue, *Proceedings of the 7th Speech Prosody Conference*, 20-23 May 2014, Dublin, Ireland. ISSN: 2333-2042.

Related publications

- Klessa, K., G. Demenko (2009). Structure and Annotation of Polish LVCSR Speech Database. *Proceedings of Interspeech Conference 2009*, September 6-10 2009, Brighton, UK.
- Katarzyna Klessa, Agnieszka Czoska, Maciej Karpiński (2015). Design, structure, and preliminary analyses of a speech corpus of infant directed speech (IDS) and adult directed speech (ADS). Presented at: 48th Annual Meeting of Societas Linguistica Europea (SLE), Leiden, The Netherlands.
- Klessa, K., Nau, N., Orlovs, O. (to appear in 2017). Timing patterns variability in Latgalian read speech. In: Wim van Dommelen, Jacques Koreman (Eds.) *Nordic Prosody XII*. Peter Lang.
- Klessa, K. & Gibbon, D. (2014). Annotation Pro + TGA: automation of speech timing analysis, *Proceedings of the 9th Language Resources and Evaluation Conference*, Reykjavik, Iceland. ISBN 978-2-9517408-8-4.
- Laukka, P. (2004). *Vocal expression of emotion: discrete-emotions and dimensional accounts*. PhD Thesis. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Science, Department of Psychology, Uppsala University.
- Metallinou, A., & Narayanan, S. (2013). Annotation and processing of continuous emotional attributes: Challenges and opportunities. In Automatic Face and Gesture Recognition (FG), *10th IEEE International Conference and Workshops* (pp. 1-8).
- Smith, S. W. (1997). *The scientist and engineer's guide to digital signal processing.* California Technical Pub. Technology & Engineering.
- SPEECON Deliverable D214, see also: Fischer, V., F. Diehl, A. Kiessling and K. Marasek. (2000). Specification of Databases – Specification of annotation.
- Wagner, A., Klessa, K., & Bachan, J. (2016). Polish rhythmic database – new resources for speech timing and rhythm analysis. *Proceedings of 10th International Conference on Language Resources and Evaluation (LREC)*. 23-28 May 2016, Portorož, Slovenia.

A list of publications on the uses of Annotation Pro & cooperation credits: http://annotationpro.org/cooperation/. Thank you!