
TRANSCRIPTION PORTAL – A ZERO-CONFIGURATION WORKBENCH FOR TRANSCRIBING SPOKEN LANGUAGE RECORDINGS

Christoph Draxler, Julian Pömp

¹*Institute of Phonetics and Speech Processing, LMU Munich*

<mailto:{draxler|j.poemp}@phonetik.uni-muenchen.de>

Abstract: The Transcription Portal is a web application that implements a semi-automatic workflow for orthographic transcription of spoken language recordings. The user drops audio files on the browser window and activates the tasks to perform, e. g. automatic speech recognition and manual correction. The files are processed automatically by the portal, and a graphical user interface displays the current status. After processing, the transcripts are downloaded to the local computer for further analysis and annotation.

1 Background

Many research fields rely on transcriptions of spoken language recordings. In a collaboration of sociolinguists, oral historians, phoneticians and speech technologists, we have developed a graphical web portal implementing a semi-automatic transcription workflow. This transcription portal (clarin.phonetik.uni-muenchen.de/apps/TranscriptionPortal/ [1]) allows users to upload audio files, choose an automatic speech recognition (ASR) service, perform manual correction, execute post-processing steps such as word alignment, and export transcripts.

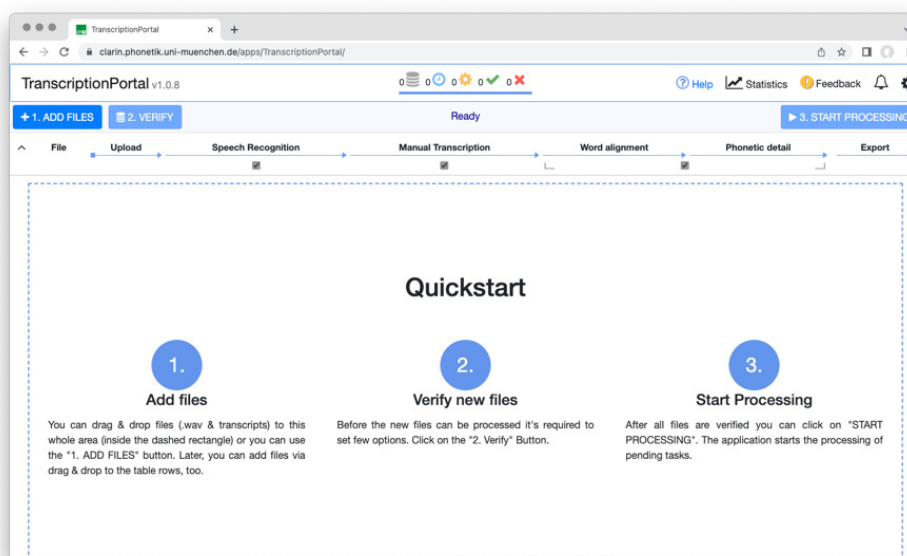


Figure 1 – Transcription Portal main page

2 Method

The transcription portal is a web application, i. e. it runs in a browser window (Figure 1). The client communicates with a backend system, which in turn calls the web services of the Bavarian Archive for Speech Signals. The BAS web services provide access to a number of ASR services, e. g. Fraunhofer, Google Speech Cloud, IBM Watson, and Radboud University's CSLT. Furthermore, for word alignment and phonetic details, the Transcription Portal accesses

the BAS segmentation and labeling service WebMAUS [2] and the Emu WebApp [3] for an in-depth phonetic annotation of the audio signal.

The graphical user interface of the transcription portal is organized as a table. The columns correspond to the workflow tasks: upload, verification, speech recognition, manual transcription, word alignment and phonetic detail, export. A file is associated with a row (Figure 2).

File	Upload	Speech Recognition	Manual Transcription	Word alignment	Phonetic detail	Export
RG-50.719.0002.01.05_0.wav	✓	⚙️	⚙️	⊙	⊙	
RG-50.719.0002.03.05_0.wav	✓	✓	✍️	⊙	⊙	📄
RG-50.719.0002.02.05_0.wav	✓	⚙️		⊙	⊙	

Figure 2 – Workflow tasks and progress visualisation for the audio files

To start processing, the user drags one or more audio files to the portal page (Figure 1). The portal checks the format of the audio files and performs basic preprocessing, e. g. splitting stereo files into separate mono channels.

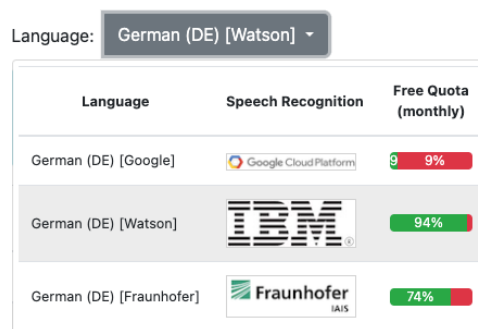


Figure 3 – Selection of ASR services. A click on the provider name opens the end user license agreement for the service

In the second step, the user selects a ASR provider and the language (Figure 3). Then, the user starts the workflow. A cogwheel icon indicates work in progress, a green checkmark stands for a successful completion of the task, a red X marks an error (Figure 2). A click on the icon reveals information on the current process and allows a preview of the current state of the transcription.

For manual correction, the transcription portal opens the Octra editor [4] within the portal window (Figure 4). A click on ‘Send transcription’ returns to the workflow for further processing.

The final task is to download the transcripts to the local computer. A large range of common file formats are supported, e. g. .txt, AnnotJSON, BAS partitur format, .TextGrid, .eaf, .ctm, .srt, .vtt). Transcripts can be downloaded individually by clicking on the download button in the corresponding row. A click on the column head will download all files in this column.

3 Summary and Outlook

The Transcription Portal is a first step into the direction of an automatic workflow for transcribing spoken language recordings. However, the hurdles for non-technical users are still high: 1) Video cannot be processed, and audio recordings often need to be converted to a suitable format outside the portal. 2) Currently available automatic speech recognition *smoothens* the transcript by eliminating hesitations, repairs and word fragments – and thus removes information which is of great interest to scholars and scientists. This is an issue that cannot be solved by the portal. 3) Finally, due to the current limited inter-process communication, status information on external processes is not available, reducing the responsiveness of the portal.

Future work will explore the potential of installing a feedback loop with the external service providers – the manual correction of ASR output may be used to improve the ASR service by providing gold standard reference data, the adjustment of phonetic segment boundaries may contribute to improving MAUS segmentation and labeling.

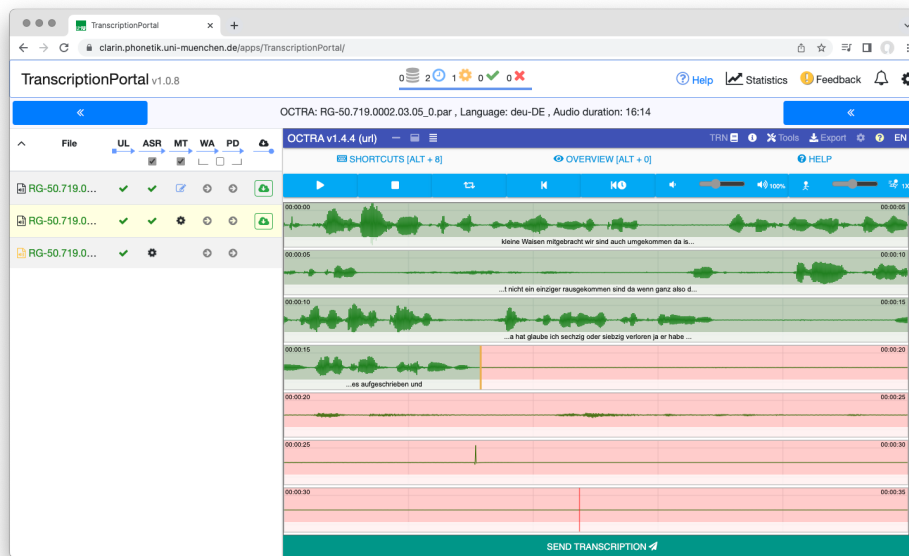


Figure 4 – Octra editor window for manual correction of ASR output within the Transcription Portal

References

- [1] DRAXLER, CHRISTOPH; VAN DEN HEUVEL, HENK; VAN HESSEN, ARJAN; CALAMAI, SILVIA; CORTI, LOUISE; SCAGLIOLA, STEFANIA: *A CLARIN Transcription Portal for Interview Data*. In: *Proceedings of the 12th Conference on Language Resources and Evaluation*, pg. 3353-3359, Marseille, 2020
- [2] KISLER, THOMAS; SCHIEL, FLORIAN; SLOETJES, HAN: *Signal Processing Via Web Services: The Use Case WebMAUS*. In: *Proceedings Digital Humanities*, pg. 30-34, Hamburg, 2012
- [3] WINKELMANN, RAPHAEL; RAESS, GEORG: *Introducing a Web Application for Labeling, Visualizing Speech and Correcting Derived Speech Signals*. In: *Proceedings of the 9th Conference on Language Resources and Evaluation*, Reykjavik, 2014
- [4] PÖMP, JULIAN; DRAXLER, CHRISTOPH: *OCTRA – A Configurable Browser-based Editor for Orthographic Transcription*. In: *Tagungsband Phonetik und Phonologie*, S. 145 – 148. Berlin, 2017.