Interspeech 2022
18-22 September 2022, Incheon, Korea

OCTRA – An Innovative Approach to Orthographic Transcription

Christoph Draxler, Julian Pömp

Institute of Phonetics and Speech Processing
Ludwig Maximilian University Munich
Germany

[draxler|j.poemp]@phonetik.uni-muenchen.de

Abstract

Octra is an open-source editor for web-based orthographic transcription. It runs in a web browser on any platform, and operates in an online and a local mode. In the local mode, files are opened via drag & drop on the browser window, in the online mode the transcriber logs in and selects the transcription project; the next file to transcribe is opened immediately.

Octra features three different viewers: dictaphone, linear and 2D, and it may access external web services, e. g. for ASR or time-aligning the transcript and the signal. It supports various common file formats for interoperability, for import and export.

In Octra, a transcriber places boundaries in the signal and then transcribes these transcription units. The length of the transcription units is determined by the personal preferences of the transcriber, and may vary with the type of recordings.

Index Terms: orthographic transcription, automatic speech recognition, forced-alignment, interoperability

1. Introduction

Generating verbatim orthographic transcripts for audio files is a basic task in spoken language processing, e. g. to create high-quality training material for speech recognition.

The WER of human transcribers is between 2% and 10% for a wide range of signal qualities, domains of discourse and speaking styles. The quality of ASR is still not sufficient for the generation of verbatim transcripts, especially for less resourced languages or dialects, historic recordings, unlimited domains, or colloquial speaking styles. [1] report WERs between 10.6% and 25.0% for contemporary clean broadcast recordings in German, and more than 45.3% for oral history recordings. In 2021, a pilot study by the first author on three different ASR services resulted in WERs of 23.2%, 58.7%, and 68.9% for recorded student presentations, compared to 3.7% for human transcribers.

For the generation of verbatim orthographic transcripts there are two basic workflows: a) transcribe from scratch, or b) pre-process the recording with ASR and manually correct the transcript. As a rule of thumb, for WERs above 10% transcription from scratch is more efficient than the manual correction of ASR-generated transcripts. Thus, for the majority of cases, transcription from scratch is the method of choice, and for this, optimised tools are necessary.

2. Octra features

From the start, Octra was designed to optimise transcription efficiency and ease of use [2]. To achieve these goals, a number of innovative features were implemented:

- Web-based operation for full platform independence.
- Local mode for individual transcription, and online mode for automated large-scale transcription projects.

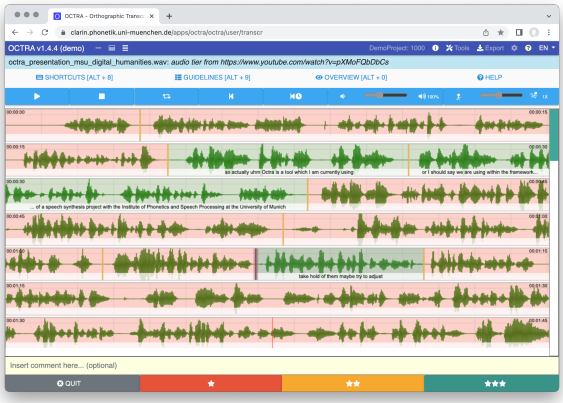

Figure 1: *Octra 2D viewer displaying 1:45 minutes of audio signal. Transcribed units show a green background.*

- Different viewers for the signal data and the transcripts
- Automatic lexical validation of the transcripts.
- Access to speech processing services such as ASR or time-alignment of transcripts and recordings.
- Support of several import and output formats.

These features will be described in the following sections.

2.1. Web-based operation

To use Octra, one simply opens the Octra web page[1]. The application runs in a browser window; once loaded, the Internet connection is no longer needed. No local software installation is necessary, and the transcriber automatically works with the latest version of the software.

Octra is an open-source project[2]. It is implemented in TypeScript using the Angular framework.

2.2. Local and online modes

In the local mode, Octra works with files from the host computer or the local network. The transcriber simply drags an audio file to the browser window and clicks on the open button. To save his or her work, the user exports the transcript in any of the supported formats (see section 2.6).

In the online mode, Octra automatically selects and opens the next available audio file for transcription. The selection can be sequential, or in random order. When the transcriber has

[1] https://clarin.phonetik.uni-muenchen.de/apps/octra
[2] https://github.com/IPS-LMU/octra

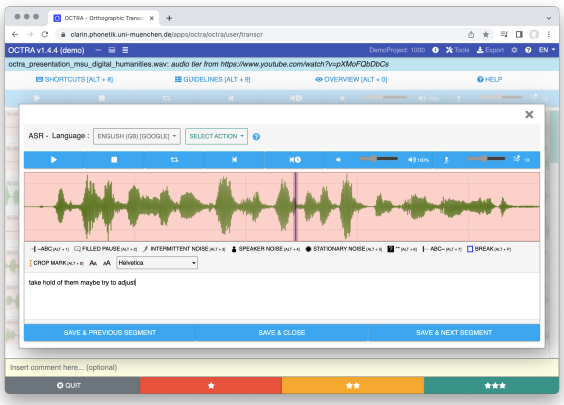

Figure 2: *Detail editor with a fragment of the signal in a transcription unit.*

finished the transcription of the current file, he or she submits the transcript to the server, and the next audio file is fetched and opened automatically.

2.3. Different transcription viewers

Octra features different viewers for the generation of transcripts. The user may select the preferred viewer, and change between the viewers without loss of data.

The most prominent visual feature of Octra is the 2D viewer. It displays large parts of the signal without visual compression – the transcriber has a good overview of the general structure of the signal, and may visually place boundaries to define the transcription units (Fig. 1). This allows a very rapid chunking of the signal according to personal preferences. Pressing the return key opens the detail editor of this transcription unit (Fig. 2). In the course of the transcription, boundaries may be shifted, deleted or inserted.

The background color of the transcription units reflects their status: red for pending, and green for done. Where available, the transcript text is shown below the signal.

2.4. Automatic validation of transcripts

Octra allows an easy specification of allowed markers and symbols for transcription and automatically performs a lexical validation while typing to prevent malformatted transcripts.

2.5. Access to external services

To speed up transcription, and to increase the joy of use, Octra accesses external services such as ASR and time alignment of transcript and audio. The ASR providers include Google Speech, IBM Watson, European Media Lab ASR, Radboud University's CLST, and Fraunhofer IAIS. For aligning the transcript and the audio, the MAUS system is used [3].

To use these services, authentication is necessary, and the user may view the service's policy on processing the audio files and supported languages (Fig. 3).

2.6. Supported file formats

Octra supports import and export of transcripts in various formats. These include plain text with or without timestamps, a tabular format, Praat TextGrid, EAF, SRT or WebVTT subtitles,

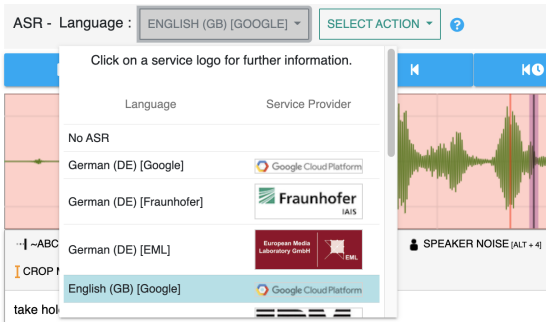

Figure 3: *Selecting ASR services for transcription.*

BAS Partitur Format, AnnotJSON, etc.

The built-in signal cutting tool may be used to split the original audio file into smaller fragments based on the transcription units. These fragments are named sequentially, and a table containing the exact signal positions of the cuts is saved along the fragments for a later reconstruction of the original file.

3. Experiences and Outlook

Octra has been used in many projects at the institute and elsewhere. In a collaboration with the Duden Verlag and the IDS Mannheim, Octra is being used to assess the quality of the recordings of more than 130,000 lexicon entries of a German pronunciation dictionary. Octra is also used for the general lingual assessment of pre-school children in Germany, for student presentations in L2 courses at the University of Modena and Reggio Emilia in Italy, for Oral History recordings in Germany, The Netherlands and Italy [4], geographically distributed transcription of social science interviews, and numerous other transcription projects.

Current developments focus on supporting different user roles, a more sophisticated data model on the server side, and increasing portability to facilitate the setup of Octra installations for private and restricted access data.

4. Acknowledgements

Octra would not have been possible without the long-term commitment of Julian Pömp and the continued support from the Bavarian Archive of Speech Signals at the Institute of Phonetics and Speech Processig at Munich University.

5. References

[1] M. Gref, J. Köhler, and A. Leh, "Improved Transcription and Indexing of Oral History Interviews for Digital Humanities Research," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, 2018, pp. pp. 3124–3131.

[2] J. Pömp and C. Draxler, "OCTRA A configurable browser-based editor for orthographic transcription," in *Proceedings Phonetik und Phonologie*, Berlin, 2017, pp. 145–148.

[3] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech and Language*, vol. 45, pp. 326–347, 2017.

[4] H. van den Heuvel, C. Draxler, A. van Hessen, L. Corti, S. Scagliola, S. Calamai, and N. Karouche, "A Transcription Portal for Oral History Research and Beyond," Utrecht, The Netherlands, Jul. 2019. [Online]. Available: https://dev.clariah.nl/files/dh2019/boa/0854.html