

KA3: Speech Analytics for Oral History and the Language Sciences

by Joachim Köhler (Fraunhofer IAIS), Nikolaus P. Himmelmann (Universität zu Köln) and Almut Leh (FernUniversität in Hagen)

In the project KA3 (Cologne Centre for Analysis and Archiving of Audio-Visual Data), advanced speech technologies are developed and provided to enhance the process of indexing and analysis of speech recordings from the oral history domain and the language sciences. These technologies will be provided as a central service to support researchers in the digital humanities to exploit spoken content.

Audio-visual and multimodal data are of increasing interest in digital humanities research. Currently most of the data analytics tools in digital humanities are purely text-based. In the context of a German research program to establish centres for digital humanities research, the focus of the three year BMBF project KA3 [L1] is to investigate, develop and provide tools to analyse huge amounts of audio-visual data resources using advanced speech technologies. These tools should enhance the process of transcribing audio-visual recordings semi-automatically and to provide additional speech related analysis features. In current humanities research most of the work is performed completely manually by labelling and annotating speech recordings. The huge effort in terms of time and human resources required to do this severely limits the possibilities of properly including multimedia data in humanities research.

In KA3 two challenging application scenarios are defined. First, in the interaction scenario, linguists investigate the structure of conversational interactions. This includes the analysis of turn taking, back channelling and other aspects of the coordination involved in smooth conversation. A particular focus is on cross-linguistic comparison in order to delimit universal infrastructure for communication from language-specific aspects which are defined by cultural norms. The second scenario is targeting the oral history domain. This research direction uses extended interviews to investigate historical, social and cultural issues. After the recording process the interviews are transcribed manually turning the oral source into a text document for further analysis [1] [L2].

To apply new approaches to analyse audio-visual recordings in these digital humanities research scenarios, Fraunhofer IAIS provides tools and expertise for automatic speech recognition and speech analytics [2] [L3]. For the automatic segmentation and transcription of speech recordings, the Fraunhofer IAIS Audio Mining System is applied and adapted. The following figure shows the user interface of a processed oral history recording.

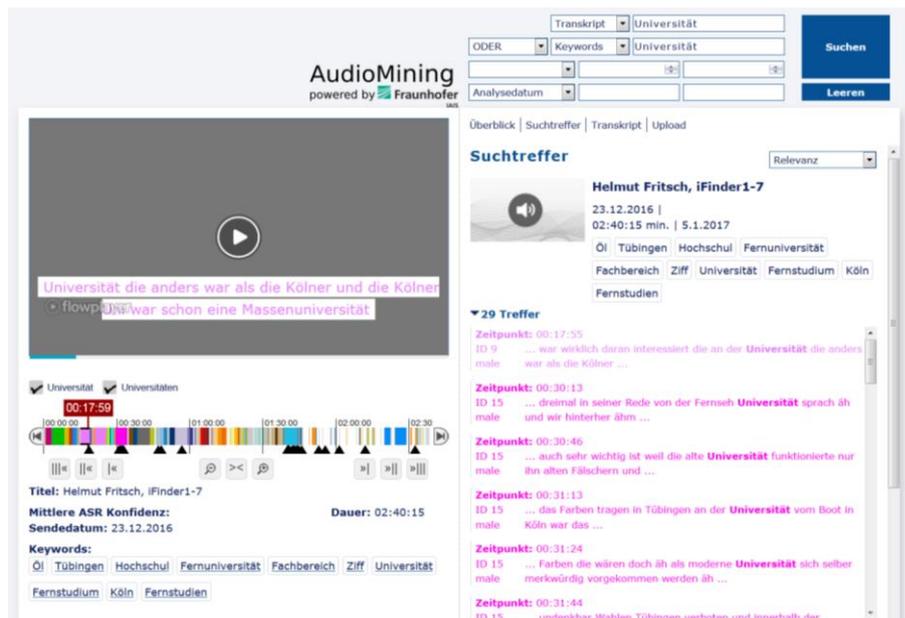


Figure: Search and Retrieval Interface of the Fraunhofer IAIS Audio Mining System for Oral History

The recording is transcribed with a large vocabulary speech recognition system based on Kaldi technology. All recognised words and corresponding time codes are indexed in the Solr search engine. Hence, the user can search for relevant query items and directly acquire the snippets of the recognised phrases and the entry points (black triangles) to jump to the position where this item was spoken. Additionally the application provides a list of relevant key words which are calculated by a modified tf-idf algorithm to roughly describe the interview. All metadata is stored in the MPEG-7 format which increases the interoperability with other metadata applications. Mappings to the ELAN format are realised to perform an additional annotation with the ELAN tool. The system is able to process long recordings of oral history interviews (up to three hours). Depending on the recording quality, the initial recognition rate is up to 75 percent. For recordings with low audio quality the recognition process is still quite error-prone. Besides the speech recognition aspect, other speech analytics algorithms are applied. For the interaction scenario, the detection of back-channel effects and overlapping speech segments are carried out.

The selected research scenarios raise challenging new research questions for speech analytics technology. Short back-channel effects are still hard to recognise. On the other hand, speech analytics tools already provide added value in processing huge amounts of new oral history data, thus improving retrieval and interpretation. The collaboration between speech technology scientists and digital humanities researchers is an important aspect of the KA3 project.

Links:

[L1] <http://dch.phil-fak.uni-koeln.de/ka3.html>

[L2] <http://ohda.matrix.msu.edu/>

[L3] <https://www.iais.fraunhofer.de/en/research/multimedia-pattern-recognition.html>

References:

[1] D. Oard: "Can automatic speech recognition replace manual transcription?", in D. Boyd, S. Cohen, B. Rakerd, & D. Rehberger (Eds.), *Oral history in the digital age*, Institute of Library and Museum Services, 2012.

[2] C. Schmidt, M. Stadtschnitzer and J. Köhler: *“The Fraunhofer IAIS Audio Mining System: Current State and Future Directions”*, ITG Fachtagung Speech Communication, Paderborn, Germany, September 2016.

Please contact:

Joachim Köhler, Fraunhofer IAIS, Germany

+49 (0) 2241 141900

joachim.koehler@iais.fraunhofer.de