Never the twain shall meet?
Collaborations between speech technologists and oral historians in the Netherlands

Norah Karrouche (Vrije Universiteit Amsterdam/Erasmus Universiteit Rotterdam)
Liliana Melgar (Universiteit van Amsterdam/Netherlands Institute for Sound and Vision)
Roeland Ordelman (Universiteit Twente/Netherlands Institute for Sound and Vision)

In the past decade, speech technologists and oral historians in the Netherlands have collaborated on a number of projects, seeking to increase the usability of automatic speech recognition (ASR) for historians working with interview data. Of particular interest is CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities)[1], the latest example of such a collaboration. In this paper, we detail the nature of collaboration in CLARIAH and identify the ways in which oral historians can profit from closer collaboration with speech recognition technologists.

CLARIAH aims to increase access to cultural heritage and research data and develop innovative digital tools for humanities and social science scholars in the Netherlands. In order to accomplish these goals, universities and scholarly research centres in the Netherlands collaborate extensively with museums, archives and libraries.[2] One of the sub-projects in CLARIAH focuses on the accessibility of audio(visual) data, via the so-called MediaSuite.[3] Research environments like the CLARIAH MediaSuite allow scholars to conduct cross-medial analyses on a plethora of subjects. The CLARIAH MediaSuite is primarily designed for media scholars and oral historians, and aims at opening up archival collections owned by several institutions in the Netherlands that would otherwise remain difficult to access, with a primary interest in audio(visual) archives, including some 70 oral history collections. The MediaSuite also provides scholars with tools, enabling them to search through vast amounts of audio, construct corpora and conduct source criticism and analysis in the same digital environment. Speech recognition is but one of these tools or services, as the MediaSuite also allows scholars to, for instance, perform explorative and comparative searches, bookmark their search results, add manual annotations and save these in a personal workspace.

In this paper, we seek to analyze the results of this most recent collaboration between speech technologists and oral historians, with an emphasis on ASR-output in the MediaSuite. We will do so by briefly demonstrating how the MediaSuite is currently used by oral historians. From a speech

---

[1] https://www.clariah.nl
[2] See CLARIAH, https://www.clariah.nl/over/geschiedenis/voorstel#samenvatting, consulted October 7 2018.
[3] See http://mediasuite.clariah.nl/, and Ordelman et al. 2018.

recognition perspective, ASR caters to the two scholarly primitives (Unsworth 2000) of *discovery* and *annotation* (Ordelman & van Hessen 2018). The ASR-output available in the MediaSuite is viewed as an annotation that can additionally facilitate the discovery of oral history data. This has not been fully recognized by oral historians, who view transcription as a means of data preparation, not annotation. Hence, as also recently stated by Ordelman & van Hessen (2018), the usability of ASR is contested by oral historians. In this paper we therefore identify and explain marked differences of opinion between speech technologists and oral historians on what exactly constitutes quality in a transcript and more importantly, to what end transcripts should be used in scholarly research (compare Scagliola & de Jong 2014). Issues addressed in the paper are: oral historians' continued preference of manually created (or corrected) transcripts over plain, possibly erroneous ASR-output, and verbatim transcripts over indexed ones.

We henceforth argue that continued collaboration with speech technologists can significantly alter the oral historian's digital workflow, specifically with regard to the oral historian's consideration of AV-material as primary source and ASR-output as an annotation and as a means by which they can discover primary sources. User documentation on the quality of ASR-output can accordingly encourage oral historians to reconsider the use of ASR in the research process. On the other hand, oral historians' recognition of ASR-output as a means to data discovery may change ruling opinions about the usability of ASR tools. This is particularly relevant in view of recent trends in historical research methodology and source criticism, where AV-material is considered more authentic than a verbatim transcript of that material.

R. Ordelman, A. van Hessen. 2018. Speech Recognition and Scholarly Research: Usability and Sustainability. In *CLARIN Annual Conference 2018*, Pisa.

R. Ordelman, et al. 2018. Challenges in enabling mixed media scholarly research with multi-media data in a sustainable infrastructure. In *Digital Humanities 2018 (DH2018)*, Mexico City, Mexico.

S. Scagliola, F. de Jong. 2014. Clio's Talkative Daughter Goes Digital. The Interplay between Technology and Oral Accounts as Historical Data', R. Bod, J. Maat, T. Weststeijn (eds.) *The Making of the Humanities, Volume III: The Modern Humanities*, Amsterdam, 511-526.

J. Unsworth. Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? Humanities Computing: formal methods, experimental practice, May 13, 2000, London, UK.