# THE PARADOX OF "TEXTUALIZING" DIGITIZED AUDIOVISUAL SOURCES
## Automatic speech recognition in digital television collections

*Tom Slootweg (Utrecht University, the Netherlands)*

## Abstract

This short paper will explore the new possibilities that have recently emerged for analyzing *audio*visual sources at the intersection of digital humanities and media history. The growing importance of digital sources and research infrastructures, including their tools and other computational affordances, have aroused various media historians to re-evaluate their established research practices, but also the methodological and epistemological underpinnings of the discipline (cf. Aasman et al. 2019; Masson 2018; Wijfjes 2017). To a considerable degree, these re-evaluations have taken place within the context of the gaining influence of digital humanities in media studies and closely associated fields (cf. Sayers 2018).

Wevers and Smits (2019) recently proposed a "visual digital turn" in media history. This turn, they claim, is both relevant and necessary due to the still pervasive reliance on the "textual" in much research at the crossroads of media history and digital humanities. Interested in the analysis of various types of historical images, they propose to deploy computer vision and convolutional neural networks (CNN) to analyze the various aspects of the pictorial dimensions of visual datasets. This paper, however, will go a step further and urges to explore the possibilities for an "*audio*visual digital turn". For media historians, in particular those working with audiovisual sources — i.e. historical broadcast television datasets — such a turn should be welcomed in order to seek new computational modes of analysis that take into consideration the idiosyncrasies and multimodalities of audiovisual data. As of now, various projects are investigating computational methods for object and facial recognition, but also automatic shot-boundary detection, and analyzing movement and color (see, among others, the SEMIA Project at the University of Amsterdam, Barbara Flueckiger's Color Film project at the University of Zürich).

When it concerns the auditive dimensions of *audio*visual digital sources, other new computational possibilities loom on the horizon. In particular automatic speech recognition (ASR) holds much promise. Currently, the Dutch CLARIAH Media Suite (cf. Ordelman et al. 2018) is integrating automatic speech recognition algorithms as part of its research infrastructure. The inclusion of ASR allows for the automated transfer of spoken Dutch in (historical) television programs into time-based, textual metadata. This opens up many new possibilities for media historians. TV-collections can now be queried on the basis of the spoken in addition to the traditional metadata provided by archival institutions. Moreover, from a methodological perspective, *audio*visual collections are rendered more suitable for various qualitative and quantitative methods such as *topic modeling*, and *frame* and *discourse analysis*.

This "textualization" of the spoken in audiovisual sources, however, also results in a striking paradox. While textualization indeed offers many new possibilities, in addition to the advancements in various forms of image recognition, the idiosyncrasies of the auditive in audiovisual sources will be lost in translation. By textualizing the spoken word, a truly

"*audio*visual digital turn", for now, seems to lie beyond our grasp. The latter will be highlighted by means of a use case from the CLARIAH Media Suite; showing the challenges that still await us in striving for an "*audio*visual digital turn" in media history.

**Keywords**: digital humanities, digital archives, ASR, audiovisual sources

**References:**
Aasman, Susan, Tom Slootweg, Liliana Melgar Estrada and Rob Wegter, 'Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History,' *VIEW: Journal for Television History and Culture*, 8, 14, 2019.

Masson, Eef, 'Humanistic Data Research: An Encounter Between Epistemic Traditions,' in Mirko Tobias Schäfer and Karin van Es, eds, Datafied Society: Studying Culture through Data, Amsterdam University Press, 2017, 25-37.

Ordelman, Roeland, Liliana Melgar, Carlos Martinez-Ortiz, Julia Noordegraaf, 'Media Suite: Unlocking Archives for Mixed Media Scholarly Research', Inguna Skadina and Maria Eskevich, eds, *CLARIN Annual Conference Proceedings 2018*, 21-25.

Sayers, Jentery, ed, *The Routledge Companion to Media Studies and Digital Humanities*, Routledge, 2018.

Wevers, Melvin and Thomas Smits, 'The Visual Digital Turn: Using Neural Networks to Study Historical Images', *Digital Scholarship in the Humanities*, fqy085, 2019, 1-14.

Wijfjes, Huub, 'Digital Humanities and Media History: A Challenge for Historical Newspaper Research,' *Tijdschrift voor Mediageschiedenis*, 20, 1, 2017, 4-24.

**Contact information:**
Tom Slootweg
Postdoctoral researcher
Utrecht University, Department of Media and Culture Studies
Muntstraat 2a
3512 EV Utrecht
The Netherlands
t.slootweg@uu.nl

**Biography:**
*Tom Slootweg* is a postdoctoral researcher at the Department of Media and Culture Studies at Utrecht University and the *MediaNow* project. He is a media historian and wrote a doctoral thesis on the arrival of electronic video in the Netherlands between the 1960s and 1990s. His current research focuses on exploring digital tools for teaching and research in media studies in general, and media/television history in particular.