Audio Indexing Technology for the Exploration of Audiovisual Heritage Collections
extended abstract[1]

Roeland Ordelman Franciska de Jong Willemijn Heeren
Arjan van Hessen

Human Media Interaction, University of Twente, The Netherlands

1 Introduction

A number of techniques from the AI-realm have proven to have added value for spoken document retrieval. Browsing tools for audio and/or video archives not only benefit from speech recognition, but also from techniques such as clustering, topic detection, speaker classification and segmentation. This paper will discuss audio indexing tools that have been implemented for the disclosure of Dutch audiovisual cultural heritage collections, and will analyze the specific requirements imposed by the nature and formats of the collections from a technological point of view. Moreover, the paper argues that research is needed to cope with the varying information needs for different types of users.

The number of digital audio collections in the cultural heritage domain is growing rapidly. Whereas the growth of storage capacity is in accordance with widely acknowledged predictions, the possibilities to index and access these archives is lagging behind. As a result, particular information may only be accessible via manual browsing of a collection of files, which is extremely time-consuming. Recent years have shown that automatic speech recognition can successfully be deployed for equipping spoken-word collections with search functionality. This is especially the case in the broadcast news domain. For that domain speech transcripts approximate the quality of manual transcripts for several languages. In other domains, a similar recognition performance is usually harder to obtain due to (i) a lack of domain-specific training data, in addition to (ii) a large variability in audio quality, speech characteristics and topics that are addressed. This applies to historical, audio(visual) data in particular. The application of audio indexing to Dutch historical audio collections, however, may greatly improve their accessibility.

2 Audio indexing for Dutch oral history collections

Full text transcription of historical audio data from a number of national audiovisual archives showed that search technology based on speech recognition might easily collapse. This is due to shockingly high word error rates caused by the typical characteristics of historical material (a wide variety in audio quality, background noise, overlapping and spontaneous speech, topics that are unknown beforehand, etc.). Therefore, requirements for successful tuning and improvement of available tools for indexing the heterogeneous A/V collections from the cultural heritage domain are reviewed through a number of pilot projects. Both the adaptation of language models to historical settings, and the adaptation of acoustic models for a homogeneous audio collection are discussed. Audio indexing for the historical domain is complicated, however, by the fact that speech training databases for that particular domain are only minimally available.

For homogeneous oral history collections, speaker adaptation of acoustic models is capable of reducing the word error rate considerably as is illustrated by the W.F. Hermans project, [2]. To overcome the mismatch of statistical language models based on contemporary text with the old-fashioned language and unknown

[1]The full paper appeared in the Proceedings of the First European Workshop on Intelligent Technologies for Cultural Heritage Exploitation at ECAI 2006, [1].

words in the task domain, historical in-domain text data are needed - preferably in large amounts. At least some information on the topics of the particular documents is wanted, e.g. to reduce the numbers of out-of-vocabulary words. In case collateral text data is available for an audio collection, it is worthwhile to investigate whether synchronization of text data with the audiovisual data using alignment techniques is an option. This will result in a time-aligned index suitable for subtitling, search and cross-media browsing (by linking semantic representations from different media). In the Radio Oranje project a number of speeches from Queen Wilhelmina (1880-1962), broadcast from England and addressed to the Dutch people during World War II, were thus aligned with their written versions.

A number of techniques described above have been implemented in two separate demos that illustrate how the concept of cross-media browsing for a multimedia archive can be realized. The first demonstrator, the cross-media news browser, was initially a demonstrator for on-line access to an archive of Dutch news broadcasts (NOS 8 uur Journaal). It shows how either available collateral data sources (subtitling information for the hearing-impaired) or full-text speech recognition transcripts can be used as linguistic content for the generation of time-coded indexes for searching within audio archives. Next to the broadcast news browser, at TNO a news browser for heterogeneous media archives has been developed: Novalist, [3]. It aims to facilitate the work of information analysts by (i) clustering related news stories to create dossiers, (ii) analysing and annotating dossiers with several types of metadata, and (iii) providing a browsing screen with multiple views on the dossiers and their metadata. These technologies can also be deployed for the disclosure of audio archives from the cultural heritage domain.

3 Variance in information needs

Multiple levels of annotation will become available and collection fragments can be linked to internal or external multimedia sources via cross-media linking. Different types of users (e.g. archivists, information analysts, researchers, teachers and the general public) are expected to have varying information needs with respect to those annotation levels. In research, for instance, questions can be asked that apply to any of the levels of metadata information and particularly new insights emerging from combining a multitude of views on the data will be interesting. The general public, however, is more likely to search for information related to their personal interests. These user needs in the cultural heritage domain must therefore be investigated more elaborately. Moreover, professional users of audiovisual archives could contribute knowledge, for example in a personalized peer-to-peer set-up that stimulates the exchange of content.

Given the greatly varying information needs from different types of users, interface requirements for search options and data presentation are also likely to differ between user groups. For eventual successful deployment of the tools to be built, the development of methodologies for the use of historical multimedia collections is a prerequisite. In the ideal case, tools for adequate navigation and selection may even unfold new information.

4 Conclusion

The challenge is to facilitate access to heterogeneous audio collections from the cultural heritage domain. For such collections, extraction of speech transcripts and metadata calls for robust audio indexing technology that performs well irrespective of speaker, bandwidth or audio quality. In addition, user research is needed to gain insight into users' information needs and system requirements that will optimally serve those needs.

References

[1] R. Ordelman, F. de Jong, and W. Heeren. Exploration of audiovisual heritage using audio indexing technology. In *Proceedings of the First European Workshop on Intelligent Technologies for Cultural Heritage Exploitation at ECAI 2006*, pages 36–39, 2006.

[2] M.A.H. Huijbregts, R.J.F. Ordelman, and F.M.G. de Jong. A spoken document retrieval application in the oral history domain. In *Proceedings of the 10th International Conference Speech and Computer*, pages 699–702, 2005.

[3] W. Kraaij. Novalist: the multimedia news browser. http://twentyone.tpd.tno.nl/druid/folders/novalist.pdf.