User requirements for access to Dutch spoken audio archives

Willemijn Heeren
Human Media Interaction
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
w.f.l.heeren@ewi.utwente.nl

Abstract

Searching archived audiovisual collections will change in the near future. Instead of sifting through kilometers of analog tapes in archives' deposits, end users will be able to explore the collections from behind a personal computer, either at an archive or at their home. A first step in the development of search technology and user interfaces suitable for supporting such access is finding out what users want and expect from the technology. Therefore, this report present a requirements analysis conducted within the CHoral project, which is part of the NWO-CATCH program.

1 Introduction

The number of spoken word collections that can be searched via the Web or via internal networks is growing, on the one hand due to retrospective digitization, as for instance in radio archives, and on the other hand due to the increasing amount of digital-born audiovisual documents with a speech track, as for example in meeting and interview collections. These collections are being exploited for all kinds of purposes, e.g., in corporate environments (meeting recordings), by content producers (TV and radio archives), in research settings (oral history), for teaching (lecture recordings) and by the general public.

In the cultural heritage (CH) domain, however, audiovisual (A/V) collections in general are at risk of becoming inaccessible, because both the analog data carriers they are stored on are deteriorating and corresponding playback devices are becoming obsolete, and the documents are insufficiently disclosed to allow fast and easy access, see e.g., [3]. The preservation issue has been taken up in retrospective digitization projects for historic A/V collections, large-scale examples of which are the EU IST PrestoSpace[1] project and the Dutch 'Beelden Voor De Toekomst' ('images for the future')[2]. The issue of improving disclosure of spoken word documents has been taken up in research projects. The general approach is to increase the granularity of annotations by automatically processing digitized A/V documents for index generation, by applying techniques

[1] http://www.prestospace.org/
[2] http://www.beeldenvoordetoekomst.nl

from information retrieval to those indexes, and by supporting search through innovative user interfaces.

In this paper we will focus on accessibility of spoken word audio collections, i.e. on collections where the speech is the main information layer. Research initiatives aiming to improve access to spoken word CH collections are e.g., Multilingual Access to Large Spoken Archives (MALACH) that addressed the issue of automatic indexing and access technology for the Shoah Visual History foundation interview collection [7], the National Gallery of the Spoken Word the goal of which was to make American historical voices searchable online [8], and the CHoral project on spoken document retrieval (SDR) for Dutch historical collections [19]. In addition to the contribution these projects make to large vocabulary speech recognition (LVCSR) and SDR, the study of collection usage and user needs received attention in the MALACH project [7, 12, 22] and is receiving attention in the CHoral project and the Multimatch project, which aims to develop a multilingual search engine for CH content, [1]. In the development of access technology for spoken heritage collections, it is important to gain insight into how the collections are being used and for which purposes.

An earlier requirements analysis of A/V archiving in the CH domain was undertaken as part of the EU Prestospace project[3]. The goal of PrestoSpace is to study and facilitate digital preservation of Europe's audiovisual heritage. A survey undertaken as part of that project aimed at providing "an overview of potential users' functional requirements for the PrestoSpace factory tools and services", [5]. Three types of end users were identified: producers (58%), cultural institutes (21%), and researchers and private persons (21%). As for access to the collections, remote access was reported to be virtually nonexistent and on-line access was very limited. Catalogs could generally be consulted on-line by querying the metadata, but one out of every two archives was not satisfied with their system's performance. The participating audiovisual archives were convinced that online viewing and listening access would increase their sales, but that rights management would remain a problem. Another problem for online listening is that the data are not stored on servers. Digitization currently means that A/V data is transferred from analog carriers to digital ones instead of onto Digital Mass Storage Systems (DMSSs), as proposed by Prestospace. The reasons for using DMSSs are that they support fast and easy access to audiovisual data, and they can be used to check the integrity of the data in order to prevent data loss. Moreover, the exchangeability of materials is complicated by the current lack of standards for documentation. One of the challenges in changing access therefore lies with the management of collections at the archives themselves.

As part of the MALACH project studies were conducted on the relevance criteria that users employ to judge search results for audio, [12, 15]. The results of [12]indicated that generally topics and summaries are helpful, but that also, information on the genre of the audio (e.g., interview, debate, report) was judged relevant, as well as time information: both the time frame of the audio and its recency.

In this paper we will focus on how to improve access to collections from the CH domain from an A/V archive user's perspective. Since studies into usage of A/V heritage collections are scarce, the first step towards gathering information

[3] http://www.prestospace.org

on the needs of users of audiovisual CH collections was undertaken in the form of a requirements analysis. This was conducted to gain insight into user needs and the current state of disclosure and access of Dutch audiovisual documents from the CH domain. The results will be used to formulate recommendations for further research and development in section 3.

2 Requirements analysis

If SDR systems would already be in use for retrieval of information from A/V archives, actual users could be asked for their opinions on current systems and users' search actions could be logged. To my knowledge, there are only two Dutch A/V collections that can be searched and accessed online: Academia[4], access to which is restricted to paying clients from educational institutes, and the Media and Communication section of the Memory of the Netherlands, an online showcase repository of digital CH content[5]. Search logs for A/V collections are not available[6].

In order to gather information on user needs and current usage of Dutch CH collections our starting point was a requirements analysis with collection keepers, who deal with requests from users in their daily practice.

2.1 Methodology

This requirements analysis was done in the form of semi-structured interviews with maintainers of audiovisual collections from the Dutch cultural heritage domain. Its objective was to meet the following three goals:

1. Gain insight into the current practice of disclosure of and the realization of access to historically relevant audio collections;

2. Determine information needs for a variety of users of audiovisual collections;

3. Estimate the acceptance of automatic indexing technology from the point of view of collection maintainers.

Maintainers of different Dutch spoken word collections participated: two archivists from the Rotterdam Municipal Archives (GAR)[7], an archivist and an editor-in-chief from the regional radio station Radio Rijnmond (RTV)[8], a maintainer of an oral history collection at the Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV)[9], a radio documentalist from the Netherlands Institute for Sound and Vision (SV)[10], and the general affairs director and the audio expert of the Meertens Institute for Research and Documentation of the Dutch Language (MI)[11]. In this selection, keepers of both

[4] http://www.academia.nl
[5] http://www.geheugenvannederland.nl
[6] After the analysis reported in this paper was completed, some of the logs from the Netherlands Institute for Sound and Vision have become available for research.
[7] http://www.gemeentearchief.rotterdam.nl
[8] http://www.rijnmond.nl
[9] http://www.kitlv.nl
[10] http://www.beeldengeluid.nl
[11] http://www.meertens.knaw.nl

broadcast collections and collections gathered for research purposes were included.

The group of participants is far from exhaustive, though. Other CH spoken word collections in the Netherlands include the KomMissieMemoires[12], het Geheugen van Nederland ('memory of the Netherlands')[13], the collection of interviews in Dutch of the Shoah Visual History Foundation[14], Groningen dialect speech[15], and (radio) podcasts[16].

The questionnaire that was used for the semi-structured interview consisted of questions concerning:

- the audio collections and their maintenance;
- disclosure of the collections;
- searchability and accessibility of the collections;
- the types of users and their uses;
- acceptance of automatic indexing technology in the archiving workflow.

The full list of questions can be found in the appendix. As for information gathered on the types of users and the current practice of disclosure and access there is some overlap with the PrestoSpace study. The current study furthermore aimed at gathering information on e.g., requests from users and the frequency of requests, practices at smaller archives, and maintenance of non-broadcast collections to get a more complete impression of audio-archiving in Dutch CH.

The result of the interviews will be presented in the following subsections. First, the collections and their maintenance will be described. This is followed by a discussion on the disclosure of those collections, and on how users can gain access to the content of the collections at present. Thirdly, the different user groups and their information needs will be presented. Finally, the role that audio indexing can play in the disclosure and accessibility of such collections will be discussed.

2.2 Audio collections and their maintenance

The types of collections maintained by the institutes that contributed to this analysis are broadcast collections and collections developed for research purposes, either in the oral history domain or for linguistic research.

Oral history collections consist of eyewitness reports on a certain event or period in history. The field of oral history is concerned with the relation between the history that has been recorded in books and the memories of individuals [16]. An example is the 'Memories of the East' collection on the end of the Netherlands colonial presence in Asia [23]. Collections that are quite similar with respect to their audio and speech characteristics are other interview collections, for instance gathered to make documentaries, such as the 'In mei, Rotterdam 1940' collection on the bombardment of the city of Rotterdam during World War

[12] http://www.ru.nl/kdc/beeld_en_geluid/kommissiememoires/
[13] http://www.geheugenvannederland.nl/index_en.html
[14] http://www.jhm.nl/default.aspx
[15] http://www.gava.nl/
[16] e.g., http://www.radiocast.nl/ and http://www.podfeed.nl

II. Other collections that are mainly being used for research are maintained at the Meertens Institute. The approximately 5.000 hours of audio maintained there – 80% of which has been digitized – are being used to investigate Dutch language variation.

The largest repository of historical audio in the Netherlands is the SV, where the radio and TV archives of Dutch national television are being kept: it consists of over 700.000 hours of multimedia archives. Smaller broadcast collections, i.e. from regional radio and television stations, are mostly being maintained at municipal archives and at the broadcasters themselves. The Rotterdam Municipal archives for instance maintain over 2.000 hours of regional radio broadcasts from RTV Rijnmond that have been digitized and disclosed, and a multiple of that amount which remains undisclosed, mainly on analog data carriers.

Digitization is a trend in the cultural heritage domain as analog carriers are decaying and their playback devices are running out of fashion. Archivists and maintainers, however, tend to prefer the use of those data carriers for which the durability, reliability and quality are well known: analog carriers. When analog data are being translated to the digital realm, the content of analog carriers is in the Netherlands most often transferred to digital data carriers, such as CD(-rom)s or DVDs. At the moment, this is the case at all institutes that participated in this study. Even though this may guarantee the collections' preservation for a while, it does not aid their accessibility. For that to be the case, digital mass storage systems (DMSSs) should be used, see http://www.prestospace.org. The Dutch maintainers recognize the importance and the added value of such a system, but not all institutes actually have plans to start using DMSSs in the near future (for instance due to lack of funding and expertise).

In addition to digital materials arising from the conversion of existing archives from analog to digital formats, radio and TV broadcasts and also oral history initiatives are nowadays being recorded digitally. The absence of digital standards is illustrated by an example given by a born-digital collection of the KITLV. Its 'Memories of the East' collection was recorded on mini-disc, but a new medium is needed since minidisc recorders are getting out of fashion. As digital archiving standards are still under development, archivists may tend to postpone the digitization process until interoperability, uniformity and quality can be guaranteed.

2.3 Disclosure of collections

There are basically two types of descriptors to disclose AV documents: content annotations and context annotations. Content descriptors are for instance summaries, full transcripts and keywords. They describe what the document is about. Context descriptors concern production date, data carrier etc., i.e. the technical details of the document. With respect to the content, descriptions are found at the level of tracks, i.e. coherent chunks of several minutes each, such as at GAR and at the KITLV, and at the level of programs with a resolution of several minutes to an entire hour or more, such as at RTV Rijnmond and SV. These differences in the granularity of description directly impact the granularity with which access to the collection can be provided.

The amount of effort put into the descriptions greatly differs within as well as between collections types. Research collections, for instance, are generally more elaborately documented than broadcast collections. In the practice of radio

archiving, differences are found between annotation practices at RTV and those at SV. This is mainly due to the different company goals, i.e. broadcasting and archiving, respectively. At RTV only a few keywords and sometimes a short content description were used to annotate a program until 1995. As a result, over a decade of material is relatively badly disclosed (the channel started broadcasting in 1983). After 1995, RTV used program scripts and written news items to make more elaborate descriptions of broadcast news, but the content of interviews, for instance, was not elaborately disclosed. Radio broadcasts at SV are being annotated much more elaborately. These is a protocol that prescribes the amount of time the description of a certain type of program may cost, the vocabulary to be used and the types of information to preserve. Not all radio broadcasts are being annotated, however; news and sports are fully described, but only a selection of the other programs is being annotated. If available, program scripts and other collateral documents are used for manual disclosure and description.

The SV has documentalists specialized in archiving broadcast materials, but at RTV archiving is the responsibility of the program makers. The result is that descriptions may not be fully accurate, and even that programs may not become archived at all. A simple example of inaccuracy are spelling mistakes that may effectively result in irretrievability of documents – at least until more advanced search systems are implemented.

Research collections are generally described more elaborately as is the case with the KITLVs Memories of the East collection[17]. The metadata per audio document has been entered into a database that – in addition to an elaborate summary of the content – contains fields for all kinds of information, such as personal data (name, birth date, family constitution, place of residence, occupation, etc.), and information on the interview itself (such as date and duration). Each audio document is partitioned into a number of 10-minute tracks, which enables the retrieval of relatively short fragments.

There have been initiatives to make the usually lengthy description process more efficient.For instance, fairly recently, the SV started using a new catalog system with an on-line search interface, iMMix, and descriptions in the new system are structured such that program characteristics are annotated at one level only, preventing double inputs. In some cases, metadata databases may be delivered to archives together with the audiovisual data and incorporating such digital metadata into the archival descriptions may save time. It is, however, not without problems, since in addition to format conversions the sets of descriptors may differ between the metadata database and the archiving standards, and the (controlled) vocabularies used may also differ substantially.

Not all Dutch audiovisual CH collections are being disclosed. Factors that contribute in the decision of whether materials should be disclosed are the (historic) importance of the recording, the cost of maintenance and materials, copyright issues and the collections uniqueness. Apart from conscious decisions on whether or not a collection should be disclosed, lack of time, man power and resources further prevent collections from becoming annotated.

[17] http://www.kitlv.nl/smgi.php

2.4 Searchability and accessibility of collections

This section will discuss the different collections' current accessibility and searchability. Access and search will first be discussed given the scenario that the user is at the institute where the collection is maintained, and second in the case that the user searches from home using the Web.

If an archive is open to the general public (the MI, for instance, is not), its catalog can generally be searched through some user interface at the institute's reading room. Standard search options allow keyword search and free text search, and more specific options may be available. Keywords are inherently restricted to those terms that appear in a controlled vocabulary (which often remains unknown to the user). As a result, the user cannot decide how a query could be improved if no results or unsatisfying results are found. Even though there are methods in information retrieval to work around such problems, e.g., thesauri and spelling correction/suggestion, these do not seem to be widely used in search engines providing access to databases on historical audio collections. Often, collection keepers are being consulted. Since they know the contents of the collections, they are able to find fragments that could not be found by a relatively naive searcher or that have not been documented as such in the database. Moreover, those specialists can give detailed information on data formats, conversion possibilities, copyright issues etc.

If a catalog search has a number of results, it lists (a part of) the descriptions, but does not give a link to the audio itself. This is in contrast with search actions for photos, maps, texts or other 2D media that are can often be shown directly. So with respect to audio or video documents, the user ends up with several IDs that refer to the actual audio documents. At the GAR, users can then listen to the documents in the audiovisual self-service room, where copies on CD or VHS-tape are available for exploration. If the user wants a copy of the materials for his/her own use, it can be requested from the service desk. As for the oral history collection of the KITLV, access to the audio is similarly organized: after identifying relevant audio descriptions, the audio documents are requested from a counter clerk. In contrast with the audio maintained at the municipal archives, however, this collection can only be listened to and studied at the KITLV. Copies are not distributed to prevent privacy breaches and out of context presentation of the, sometimes sensitive, materials. Another, very practical reason that was given for not linking the audio directly to the search results is the fact that this would require much memory capacity of the institute's network.

The numbers of requests for searches in spoken word collections are generally small. At RTV there are several requests per day, the KITLV oral history collection receives about one request per day, at the Rotterdam municipal archives and at the Meertens Institute there are only a few per month. From SV the numbers of requests are unknown to the author, but one of its documentalist mentioned that there were few requests for radio broadcasts specifically.

In many cases, users nowadays do not initiate a search by visiting an archive, but start at home behind their personal computer and search either the Web using one of the well-known search engines, or a particular institute's website. From those institutes' websites the catalogs can usually be searched online. Search options are generally the same as from within the institute, and the audio documents that search results refer to cannot be listened to. A visit to the maintainer to request access or a copy is inevitable.

According to the interviewees the possible improvements on the current situation relate to the descriptions, the user interface and the use of a DMSS server. Firstly, maintainers remarked that the more detail is being described, the easier it is to fulfill requests. In the case of archiving broadcasts, annotating quotes and remarkable background sounds such as barrel organs have proven to be valuable. Moreover, maintainers learn from experience which topics they encounter in user requests; this can make them adapt the way in which they disclose new materials. Secondly, search interfaces could in some instances be more user-friendly. One way of realizing this – according to the participants – is by offering a standard and an advanced search screen; the user can either enter a number of search terms into the general search field, or he can specify the nature of a number of search terms to retrieve documents more precisely. Thirdly, direct access to audio and video documents is thought to improve on the current situation. It would significantly reduce the time that lapses between entering a request and listening to actual results. This would be very beneficial in the case of producing news items in response to unexpected events. Moreover, users could assess the use of the materials faster and new user groups could be reached. The current situation of audio collections on digital data carriers cannot support this scenario: therefore, DMSSs should be used. In the case of most collections, however, on-line access should be carefully designed to prevent copyright violations and misuse.

2.5 Users and uses of historical audio collections

Users of historical audio archives can be divided into two main groups: professional users and the general public. At GAR about 75% of the users are professionals (e.g., makers of new content, researchers[18]). Also SV is mainly being searched by professional users. The collection of the MI is exclusively for research purposes. The KITLV collection's users are mainly researchers, students and content producers, but it is also being consulted by the general public.

Professional users: An important user group for audiovisual (broadcast) archives are makers of new content. This group is very diverse: e.g., exhibition makers, event organizers, companies, makers of films or documentaries, artists and both local and national broadcasters. Moreover, the keepers of the archives themselves also function as makers of new content, especially in the case of the SV and RTV. As a whole, this user group has two main uses for the materials: (i) research during the preparation of a production, and (ii) content for a production. Mainly in the second case, time pressure may be high as news producers want to react as soon as possible to sudden events such as accidents and disasters. Content producers tend to search audio collections looking for (combinations of) events, keywords, locations and persons. In some cases, they look for sound impressions, such as the sound of a harbor or city, but these may be very difficult to find as they are normally not being annotated.

Another group of professional users are researchers, students and teachers. These users usually pose very specific research questions in comparison with

[18] In contrast to the Prestospace survey we include researchers amongst the group of professional users.

content producers, see also [14]. For the KITLV collection, free text search in the elaborate summaries works relatively well. For most of the other collections, keyword search is the most promising according to the collection keepers. Researchers may be interested in all kinds of subjects and will search for (combinations of) names, keywords, events, periods and locations in order to find results to incorporate into their research, writing and teaching. The MI's collection that was specifically built for linguistic research purposes is somewhat different. When studying language structure the question of *how* things are being said (e.g., pronunciation, grammatical structure, intonation) is often much more important than the question of *what* is being said. The users of this collection often make their own, time-costly annotations.

General public: The second type of users are the general public. They typically search for audio documents that fit their personal interests, such as a hobby or their family history. Their requests are mostly for names of persons, companies, locations and/or events in the case of both the broadcast and the oral history collections.

2.6 A role for automatic indexing

The cultural heritage domain has been characterized as somewhat hesitant towards technological development concerning the automatic indexing of collections. This is understandable, for instance because automatically generated transcripts are certainly not error-free (as opposed to manually checked descriptions). Still, given the vast size of several audio collections that have not been disclosed, and the manual labor of one to ten times real time that disclosure would cost, archivists and collection maintainers understand the added value of automatic indexing. They furthermore suggested another use. Automatic indexing could be employed to provide archivists with an impression of a collection on the basis of which a selection for full disclosure can be made. A number of participants explained that their deposits held tapes for which the exact content was unknown.

According to the interviewees there are certain restrictions on what can be expected from automatic indexing. First, since human interpretation is lacking, certain abstractions, i.e. higher-level annotations, cannot easily be made. As a result it was expected that users looking for journalistic content (i.e. facts) would have less problems retrieving relevant documents that were automatically annotated than users looking for artistic content (i.e. sound impressions). Moreover, there is the (partial) mismatch between the words that are being spoken and the more abstract topic that is being talked about. Second, when collections are to be used for certain types of research, automatic transcription may not be suitable at all, since researchers need manually checked indexes at layers of information that may abstract from the words (e.g., communicative acts, prosody). To generate a first version of an index, however, speech technology might be employed to reduce the amount of work (which is exactly what has been proposed in the Prestospace project). Thirdly, searchers are unfamiliar with a situation in which the annotations on which search is based are not manually checked. They must therefore receive instructions on the probabilistic nature of automatically generated annotations. Firstly, metadata models must

be able to incorporate the different annotation types, manual and automatic, preferably at multiple levels of abstraction.

3 Discussion

In this report we set out with three goals in mind: (i) gain insight into the current practice of disclosure and the realization of accessibility of Dutch historical audio collections, (ii) gather information on the users of audiovisual collections and their needs, and (iii) to receive feedback from collection maintainers on the potential of automatic indexing technology in the audiovisual archiving workflow. Together these goals aimed at gathering user requirements for spoken document retrieval systems in CH. Those requirements, in turn, will be used to determine a research agenda for improving automatic disclosure and access for spoken word collections from CH. In the rest of this section we will discuss the main findings of our requirements analysis, and also how those requirements can be met or addressed in future research.

Now that archives increasingly acknowledge the need to make their collections accessible to end users – in addition to the traditional tasks of describing and maintaining collections – the question of how to make collections available to the general public is being addressed. We found that access to Dutch audiovisual collections – in line with its European counterparts, [5] – is relatively slow and cumbersome at present: e.g., direct, on-line access to audiovisual content is basically nonexistent, short content descriptions seem insufficient to meet the wide variety in users' information needs, and many collections have not yet been annotated which makes them almost inaccessible. Problems in disclosure, which is a prerequisite for access, are mainly caused by the costliness – both in time and in man-power – of producing elaborate, high-quality annotations. Access is furthermore complicated by the fact that the digital infrastructure in archives is in many cases not yet ready for on-line presentation of audiovisual content (provided that IPR issues etc. enable publication).

The first requirement is to make access faster. This is expected to be realizable (i) by presenting content online, and (ii) by making search results more focused, i.e. by retrieving pointers to relevant locations within documents instead of to entire documents. For online presentation, audio sources should be digitally available and linked to the results from catalog search. If necessary, this could be arranged via a log-in procedure to prevent IPR violations and/or misuse. Online presentation moreover entails the development of an infrastructure that supports data management, and also retrieval and presentation of both metadata and time-labeled spoken content. Most of these developments fall outside the scope of the CHoral project, but are being taken up at the CH institutes themselves and in other research projects.

The second way of making access faster is being researched in the CHoral project, and has been addressed in other spoken document retrieval projects such as MALACH and The National Gallery of the Spoken word. Automatic content indexing and audio processing tools are being developed to increase the time-resolution of search results through the addition of time labels to the spoken content, or to highlights within the documents. Since current descriptions may lack much detail, such technology has the potential to increase the numbers of user requests that can be fulfilled, and the accuracy of content delivery.

There are a number of restrictions on what can currently be expected from automatic annotation, mainly caused by the statistical nature of the techniques employed. Content-based annotation of spoken word documents is usually done using automatic speech recognition (ASR). The Word Error Rates on documents with spontaneous conversational speech lie in the range of 40-60% for a number of languages, see e.g., [3, 8, 11]. Only correctly recognized words can in principle be successfully retrieved. Correct recognition depends on the suitability of the acoustic and language models for the recognition task at hand.

Audio pre-processing tools such as Speech Activity Detection (SAD) and speaker segmentation are generally referred to as audio diarization. It aims at determining which audio intervals contain speech and of which type, so that the ASR engine is only fed speech, and not music, and models can be adapted to the data. Performance on broadcast news audio is high, with miss and false alarm rates around 1% (see [24] for an overview), but the SAD error rate is significantly higher, i.e. around 11%, on more heterogenous data sets [10].

In line with findings from the PrestoSpace requirements study, Dutch audiovisual collections are mostly being used by professionals, both content producers and researchers/students[19]. The second requirement therefore is to support search by these user groups. A logical first step in further research would be to study how these users are supported by the current state-of-the-art in spoken document retrieval in comparison with running demonstrator systems, such as the Radio Oranje demonstrator, [25]. A next step would then be to make adaptations to the user interfaces given the users' preferences and feedback.

Another requirement pertains to the types of information users want to find. They mainly search for (combinations of) events, keywords, locations and names of persons/companies. A significant proportion of requests therefore contains named entities that are – however – not straightforwardly extracted from spoken content automatically. Both approaches to named entity recognition and optimal use of information on named entities present in manual metadata should be investigated further in order to support searchers. As solutions to enhanced recognition of named entities, or Out-Of-Vocabulary (OOV) terms in general, several approaches have been forwarded, e.g., multi-pass recognition, e.g., [6], query and document expansion, e.g., [27], and subword approaches, e.g., [2, 17]. As part of the CHoral project, search in phoneme lattices derived from word lattices will be investigated further with the goal of improving retrieval of named entities and OOVs in Dutch. The current analysis further showed that users do not often search for time-related concepts such as dates or periods, and if they do, it is mainly in combination with other types of terms.

A seemingly obvious requirement is that search interfaces should be made more user-friendly. In both the PrestoSpace survey and the present requirements analysis search interfaces for exploring an archive's catalog were often judged as insufficiently user-friendly by collection keepers. To improve on the current situation different tools that have demonstrated added functionality in research systems should be tested on systems in public use. Moreover, additional tools to support searching, browsing, and selection of audiovisual documents are be needed. For instance, thanks to the increased granularity of automatic indexing time-aligned metadata can be presented while the A/V documents are being

[19] Given the popularity of websites such as YouTube (http://www.youtube.com/), however, the interest of the general public in online interactivity with AV documents may increase in the near future.

played, [4, 9, 13], and there are also other ways of helping users during the assessment of A/V documents by representing the contents textually, [18, 20], and/or visually, [21, 26].

Many user support tools developed so far were intended for use on specific collections with specific document types. Audiovisual archives typically hold a large variation in document types, which complicates the situation for the user. Retrieved results should not only be on-topic, but their genres should for instance also be available to searchers. For as far as we know automatic detection of genre in spoken word documents is an open research question, as well as the way such information should be presented in the user interface. First steps towards such a classification could be made by using diarization technology to estimate the numbers and turn patterns of speakers.

An earlier report that investigated the attitude of archivists towards audio indexing technology showed that archivist are ambivalent towards the technology, [28]. On the one hand they acknowledged the potential added value. On the other hand, confronted for example with imperfect speech transcripts, archivist may become skeptical about the usefulness of the automatically generated metadata. The present analysis showed that most participants welcomed automatic solutions for annotations, and they also forwarded a new ideas for applying the technology. Keepers of research collections, however, were understandably more hesitant, since their collections need exact annotations for scientific analysis that in most cases cannot (yet) be generated automatically.

4 Conclusion

The findings of this study provide more specific instructions on which lines of research to pursue in order to improve search in A/V archives:

- Focus on two target groups in user and usability studies; producers and researchers;
- Further automatic indexing technology for spoken documents, i.e. diarization and speech recognition;
- Research and develop ways to deal with OOV queries, mainly with named entities;
- Develop classification tools for generating higher-level semantics, e.g., topic classification;
- Optimally use the index for content representation in the user interface;
- Test the usability of the UI and its components in ecologically valid settings.

Most of these issues have been taken up in the CHoral project.

Acknowledgements This paper is based on research funded by the CATCH program (http://www.nwo.nl/catch) of the Netherlands Organisation for Scientific Research.

Appendix The questions that were used during the semi-structured interviews with collection keepers were:

1. In 1999, Kooijman published an inventory of the audiovisual collections in Dutch archives. Which changes/additions can you report for your institute?
2. Which types of collections do you maintain?
    - radio/TV
    - oral history
    - other types of interviews
    - speeches/monologues
    - other
3. Does you archive contain digital audio or video materials? If so, are the files kept on hard drives or on data carriers?
4. How have the materials been disclosed? Which type of metadata/description of these materials is available?
5. How can a user get access to the materials from within the archive?
6. How can a user get access to the materials from outside of the archive?
7. How often do you receive requests for audio documents?
8. Which groups of users do those requests come from?
9. Which types of queries do users have? Are there any topics that are asked for regularly?
10. What type of information do users search for?
    - names of persons or places
    - dates or periods
    - events
    - keywords
    - a particular topic
    - speaker profile
    - other
11. What do searchers want to use the information for?
12. Could disclosure and access of the collection(s) you maintain be improved? If so, how?
13. What is your opinion on developments in speech and language technology for spoken document retrieval? (This question was always preceded by a short explanation of the state-of-the-art in SDR)
14. Do you have any further comments?

References

[1] Giuseppe Amato, Juan Cigarrán, Julio Gonzalo, and carol Peters. Multimatch - multilingual/multimedia access to cultural heritage. In *Proceedings of the 2nd Italian Research Conference on Digital Library Management Systems*, 2006.

[2] M. G. Brown, J. T. Foote, Gareth J. F. Jones, Karen Sparck Jones, and S. J. Young. Open-vocabulary speech indexing for voice and video mail retrieval. In *ACM Multimedia*, pages 307–316, 1996.

[3] W. Byrne, D.Doermann, M. Franz, S. Gustman, J. Hajic, D. Oard, M. Picheny, J. Psutka, B. Ramabhadran, D. Soergel, T. Ward, and W-J. Zhu. Automatic recognition of spontaneous speech for access to multilingual oral history archives. *IEEE Trans. Speech Audio Proc.*, 12(4), 2004.

[4] M.G. Christel. Evaluation and user studies with respect to video summarization and browsing. In *Proceedings of IS&T/SPIE Symposium on Electronic Imaging*, 2006. San Jose, CA.

[5] B. Delaney and B. Hoomans. Prestospace deliverable 2.1 User Requirements Final Report, 2004.

[6] P. Geutner, M. Finke, and A. Waibel. Selection criteria for hypothesis driven lexical adaptation. In *ICASSP '99: Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference*, pages 617–620, Washington, DC, USA, 1999. IEEE Computer Society.

[7] S. Gustman, D. Soergel, D. Oard, W. Byrne, M. Picheny, B. Ramabhadran, and D. Greenberg. Supporting access to large digital oral history archives. page 18.

[8] J.H.L. Hansen, R. Huang, B. Zhou, M. Deadle, J.R. Deller, A. R. Gurijala, M. Kurimo, and P. Angkititrakul. Speechfind: Advances in spoken document retrieval for a national gallery of the spoken word. *IEEE Transactions on Speech and Audio Processing*, 13(5):712, 2005.

[9] W.F.L. Heeren, L.B. van der Werff, R.J.F. Ordelman, A.J. van Hessen, and F.M.G. de Jong. Radio oranje: Searching the queen's speech(es). In C.L.A. Clarke, N. Fuhr, N. Kando, W. Kraaij, and A. de Vries, editors, *Proceedings of the 30th ACM SIGIR*, pages 903–903, New York, 2007. ACM. ISBN=978-1-59593-597-7.

[10] M.A.H. Huijbregts and C. Wooters. The blame game: Performance analysis of speaker diarization system components. In *Proceedings of Interspeech 2007*, page 4, Antwerp, 2007. International Speech Communication Association. ISSN=1990-9772.

[11] Marijn Huijbregts, Roeland Ordelman, and Franciska de Jong. Annotation of heterogeneous multimedia content using automatic speech recognition. In *proceedings of SAMT*, 2007.

[12] J. Kim, D.W. Oard, and D. Soergel. Searching large collections of recorded speech: A preliminary study. In *Proceedings of the Annual Conference of the American Society for Information Science and Technology*, Long Beach, CA, 2003.

[13] S.R. Klemmer, J. Graham, G.J. Wolff, and J.A. Landay. Books with voices: paper transcripts as a tangible interface to oral histories. In *Proceedings of CHI 2003*, 2003. Ft. Lauderdale, Florida.

[14] T. Kouwenhoven. *Zoeken Navigeren Vinden. Over zoekers, zoekgedrag, zoekmachines en hun interfaces bij het zoeken naar audiovisuele content*, chapter 3, page 116.

[15] Katy Newton Lawley, Soergel Dagobert, and Xiaoli Huang. Relevance criteria used by teachers in selecting oral history materials. In *Proceedings of the 68th Annual Meeting of the American Society for Information Science and Technology (ASIST)*, 2005.

[16] S. Leydesdorff. *De mensen en de woorden*. Meulenhoff, 2004.

[17] Beth Logan, Pedro Moreno, and Om Deshmukh. Word and sub-word indexing approaches for reducing the effects of oov queries on spoken audio. In *Proceedings of the second international conference on Human Language Technology Research*, pages 31–35, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[18] C. Munteanu, R. Baecker, G. Penn, E. Toms, and D. James. The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In *Proceedings of CHI 2006*, page 493.

[19] R.J.F. Ordelman, F.M.G. de Jong, and W.F.L. Heeren. Exploration of audiovisual heritage using audio indexing technology. In L. Bordoni, A. Krueger, and M. Zancanaro, editors, *Proceedings of the first workshop on intelligent technologies for cultural heritage exploitation*, pages 36–39, Trento, 2006. Universita di Trento. ISBN=not assigned.

[20] A. Ranjan, R. Balakishnan, and M. Chignell. Searching in audio: the utility of transcripts, dichotic presentation and time-compression. In *Proceedings of CHI 2006*, 2006.

[21] Laura Slaughter, Douglas W. Oard, Vernon L. Warnick, Julie L. Harding, and Galen J. Wilkerson. A graphical interface for speech-based retrieval. In *ACM DL*, pages 305–306, 1998.

[22] D. Soergel, D. Oard, S. Gustman, L. Fraser, J. Kim, J. Meyer, E. Proffen, and T. Sartori. The many uses of digitized oral history collections: Implications for design. Malach technical report, College of Information Studies. University of Maryland, June 2002.

[23] F. Steijlen. *Memories of The East.* KITLV press, 2002.

[24] S.E. Tranter and Reynolds D.A. An overview of automatic diarization systems. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1557, 2006.

[25] L.B. van der Werff, W.F.L. Heeren, R.J.F. Ordelman, and F.M.G. de Jong. Radio oranje: Enhanced access to a historical spoken word collection. In P. Dirx, I. Schuurman, V. Vandeghinste, and F. Van Eynde, editors, *Proceedings of the 17th Meeting of Computational Linguistics in the Netherlands*, pages 207–218, Utrecht, 2007. Landelijke Onderzoekschool Taalwetenschap.

[26] Steve Whittaker, Julia Hirschberg, John Choi, Donald Hindle, Fernando C. N. Pereira, and Amit Singhal. SCAN: Designing and evaluating user interfaces to support retrieval from speech archives. In *Proceedings of SIGIR99 Conference on Research and Development in Information Retrieval*, pages 26–33, 1999.

[27] P.C. Woodland, S.E. Johnson, P. Jourlin, and K. Spärk Jones. Effects of out of vocabulary words in spoken document retrieval. In *SIGIR 2000*, 2000. Athens, Greece.

[28] E. Zuurbier. Onderzoek naar de haalbaarheid van Spoken Document Retrieval. Master's thesis, University of Twente, 2004.