Unravelling the Voice of Willem Frederik Hermans: an Oral-History Indexing Case Study

Roeland Ordelman, Marijn Huijbregts and Franciska de Jong

July 11, 2006

1 Introduction

With the 10th anniversary of the death of the Dutch novelist Willem Frederik Hermans (1921–1994), the Willem Frederik Hermans Institute initiated the set-up of a Willem Frederik Hermans portal. Here, all available information related to the Dutch novelist and his work can be consulted. A part of this portal was planned to be dedicated to a collection of spoken audio material. This report describes the search functionality that was attached to this collection by the Human Media Interaction Group of the University of Twente. This project (further refered to as the WFH project) can be regarded a case-study of the disclosure of an oral-history spoken word archive using audio mining technology.

2 Audio mining

The number of digital spoken-word collections is growing rapidly. Due to the ever declining costs of recording audio and video, and due to improved preservation technology huge data sets are created, both by professionals at various types of organisations and non-professionals at home and underway. Partly because of initiatives for retrospective digitisation, data-growth is also a trend in historical archives. These archives deserve special attention because they represent cultural heritage: a type of content which is rich in terms of cultural value, but has a less obvious economical value. Spoken-word archives belong to the domain of what is often called oral history: recordings of spoken interviews and testimonies on diverging topics such as retrospective narratives, eye witness reports, historical site descriptions, and modern variants such as ‘Podcasts’ and so-called amateur (audio/video) news[1].

Where the growth of storage capacity is in accordance with widely acknowledged predictions, the possibilities to index and access the archives created is lagging behind though [1]. Individuals and many organisations, often do not have the resources to apply even some basic form of archiving. Spoken word collections may become the stepchild of an archive—minimally managed, poorly preserved, and hardly accessible. The potentially rich content in these collections risk to remain inaccessible.

For ‘MyLifeBits’ chronicles collected by non-professionals under uncontrolled conditions [2] the resemblance with shoe-box photo collections (i.e., little annotation and structure) may be acceptable. But for audio collections with a potential impact that is not limited to the individual who happened to do the recording, there is a serious need for disclosure technology. Tools for presenting and browsing such collections and to search for fragments could support the information need of various different types of users, including archivists, information analysts, researchers, producers of new content, general public, etc.

The observation that audio mining technology can contribute to the disclosure of spoken word archives has been made many times [3], and several initiatives have been undertaken to develop this technology for audio collections in the cultural heritage domain. Worthwhile mentioning are projects such as ECHO (European CHronicles Online), that focused on the development of speech recognition for historical film archives for a number of European languages [4], and MALACH, applying ASR and NLP for the disclosure of holocaust testimonies [5]. But the high expenses required to process historical material in combination

[1] ‘Podcasts’ are home-brew radio shows covering personal interest items and can be viewed as the audio variant of a ‘blog’ which is basically a journal that is made available on the web. Amateur news is news compiled by amateurs and broadcasted via the web.

with the expected limited financial return on investment have prohibited real successes. A break through for the application of audio mining outside standard domains (typically: English news) is still pending.

3 Audio mining in the WFH project

The overall goal of the WFH project was to enable content based searching of audio documents in the collection. Not only the selection of relevant audio/video documents given some user query, but especially the selection of relevant audio/video *document parts* was targeted. A spoken document retrieval approach was choosen: full, time-coded text transcriptions of the speech that is encountered in the audio collection are generated automatically by means of state-of-the-art speech recognition technology. The transcriptions in turn are then used as an index for searching the documents.

3.1 Collection

Audio and video material

The initial collection (april-2005 collection) to be disclosed consisted of some 10 to 15 hours of audio material: lectures, book-reading excerpts and interviews featuring Willem Frederik Hermans (WFH). More data is expected to become available at a later stage. There was also one video document but only the audio stream was used for indexing (no video analysis was applied).

Although WFH is not the only speaker present in the material, his voice dominates the larger part of the collection. The lectures were recorded at different locations with different reverberation characteristics. Lectures may have applause, laughter, coughing and questions from the audience that –even for a human listener– sometimes are hard to recognise. Parts of the interviews are quite informal and recorded in a home environment on celluloid tape.

Metadata

For every audio/video item in the collection, descriptive metadata was available, encompassing a titel, production data, and a short description of the contents. This metadata came available at a later stage and was only used for presentation purposes. In Appendix the april-2005 collection and metadata is listed.

In domain text data

A number of textual data sources from the domain were made available for language modeling purposes. These sources contain comtemplative work, a short novel and interviews with the novelist. Some of the text data had been digitized using OCR and contained errors.

4 Speech recognition

Information retrieval research that uses the spoken audio parts of documents for retrieval is commonly referred to as spoken document retrieval (SDR) or alternatively, speech-based retrieval. 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 domains such as the broadcast news domain which is very general and makes data collection for system training relatively easy. For the broadcast news domain speech transcripts therefore approximate the quality of manual transcripts for several languages and spoken document retrieval in the American-English broadcast news (BN) domain was even declared "solved" with the NIST sponsored TREC SDR track in 2000 [6]. In other domains than broadcast news, a similar recognition performance is usually harder to obtain due a lack of domain specific training data, in addition to a large variability in audio quality, speech characteristics and topics that are addressed. This applies to the oral-history domain in particular.

The most obvious approach in spoken document retrieval is the word-by-word translation of the encountered speech using a large vocabulary continuous speech recognition (LVCSR) system. Having generated a textual representation (full text transcription) of an audio or video document, the document can be searched as if it were a text document. As mentioned above, the time-labels provided by the speech recognition system and the segmentation of a large document into sub-documents, provide additional means for structuring the document.

In this project we used the tools and resources collected and developed earlier for a broadcast news (BN) LVCSR system as a starting point. This system showed adequate performance in a Dutch spoken document retrieval task in the news domain [7]. As the performance of a BN system in the oral history domain was expected to be poor, the goal was to port the BN system to a WFH specific system. Similar BN systems are available in many labs, so the conversion of the BN system and tuning to a collection from the oral history domain might be a case of a more general interest for research groups that want to pursue applications for their ASR tools for similar purposes.

4.1 Development and training data

One of the lectures and a television documentary with a number of interviews were manually annotated at word level, encompassing 130 minutes of speech, with WFH speaking approximately 85% of the time. This subset of the audio recordings was divided in a training set, a test set and an evaluation set. The training set (78 minutes) was used for training the acoustic models. The test set was used to evaluate both the acoustic models and the language models during development. The evaluation set was used for the final evaluation of the system.

For training BN language models we used a large Dutch news related text corpus of in total some $400M$ words [7]. Two other text collections were available for domain adaptation. A number of written interviews with WFH and one of his short novels made up the first collection (further referred to as WFH-text) containing one and a half million words. Word-level transcripts of general conversational speech from the Spoken Dutch Corpus [8] formed the other text collection. This collection consists of $1.65M$ words. Both text collections were used for adaptation of the language model.

4.2 Broadcast news system

Two broadcast news ASR systems were available as a starting point. The first system (UT-BN2002) is based on hybrid RNN/HMM acoustic models, a $65K$ vocabulary and a statistical trigram language model, created using a news corpus. The acoustic models are created out of approximately 20 hours of broadcast news speech.

The other system (UT-BN2005), is based on a recogniser which is developed at the University of Colorado (CSLR) and is freely available for research purposes. Its acoustic models are decision-tree state-clustered Hidden Markov Models. [9].

Twenty-two hours of broadcast news recordings from the Spoken Dutch Corpus ([8]) were used to port the gender independent acoustic models from the English system to Dutch and to train new broadcast news acoustic models. The ARPA language model created for UT-BN2002 was reused in the UT-BN2005 system. The main advantage of UT-BN2005 over the UT-BN2002 system is that it allows to apply adaptation methods to the acoustic models, which is highly relevant given the mismatch between the target collection and the BN training material.

4.3 WFH system

The WFH transcription system is based on components of both BN systems. The BN language model and acoustic models were adapted to the target domain as will be described in this section.

4.3.1 Language model adaptation

Two domain specific trigram language models were trained. These models both use a $30K$ vocabulary containing domain specific words.

The most occurring words from the WFH-text described above were complemented with the most occurring words from the newspaper corpus. Vocabularies of different sizes were created and the out-of-vocabulary (OOV) rates were computed on a preliminary test set. The OOV of a $60K$ vocabulary ($50K$ from newspaper data and $10K$ from WFH-text) was only marginally better than the $30K$ vocabulary ($20K$ newspaper words with $10K$ from WFH-text). As a smaller vocabulary will result in less acoustic confusability, the $30K$ vocabulary was chosen to be used for the WFH-specific language models.

From each of the two domain specific text collections described in section 4.1, a language model was created. The language model created from the WFH-text contains $178K$ trigrams and $335K$ bigrams. The conversational speech text collection contains $153K$ trigrams and $306K$ bigrams. A third model was created using the newspaper corpus and the $30K$ vocabulary. From these three models, a mixture language model was created. Mixture weights were computed using the transcripts of the acoustic training set.

4.3.2 Acoustic model adaptation

In the annotated speech material, WFH is speaking 110 out of the 130 minutes. It is therefore reasonable to expect that the recognition rate will improve when a WFH-dependent acoustic model is used instead of the broadcast news acoustic model. Two new acoustic models were trained. Both models were evaluated using the mixture language model described in the previous section.

The first model was trained solely on the part of the training set in which WFH is speaking, in total 78 minutes of speech data. The second model was created by adapting the broadcast news acoustic model (UT-BN2005) to the training data using the Structured Maximum a Posterior Linear Regression (SMAPLR) adaptation algorithm [10]. In [9] and [10] it is shown that SMAPLR performs very good, even when little adaptation data is available.

4.3.3 Dictionary

A pronunciation dictionary was created from the $30K$ vocabulary using a large background pronunciation lexicon provided by Van Dale Lexicography and a grapheme-to-phoneme (G2P) converter trained on the same pronunciation lexicon [7].

4.4 Experimental results

4.4.1 BN system performance

On the BN test set (4 hours from the Spoken Dutch Corpus), the UT-BN2005 system outperforms the BN-2002 system with word error rates (WERs) of 30% and 35% respectively. On the WFH test set both BN systems have a comparable performance of above 80% WER. The word error rate of the UT-BN2005 system is 80.4%. On the parts in which only WFH is speaking a 81.6% is scored, on other speakers we obtain a WER of 67.2%.

4.4.2 WFH system performance

To evaluate the performance of the WFH system, three evaluation runs were performed. First, the performance of the WFH language models will be reported. Next, the performance of the acoustic models will be discussed and finally, the performance of the combination of the best LM and AM will be reported.

4.4.3 Language model

Table 1 shows the perplexities of the four created language models (see session 4.3.1), along with the word error rates on the test set of those models obtained using the BN2002 acoustic model. All language models use the same $30K$ vocabulary.

All language models perform better in terms of WER than the original news model. Merging all models into a single mixture model gave the best results on the WFH test set. Note that mixtures of two of the three language models did not improve results.

Name	PP	%WER
Newspaper	245	77.2
Conversational speech	274	78.8
Domain	235	77.1
Mixture	195	75.4

Table 1: *The perplexity (PP) and the word error rate (WER) of the four language models. All language models use the same $30K$ vocabulary. The WERs were calculated the 2002 broadcast news acoustic model.*

4.4.4 Acoustic model

The word error rates of the broadcast news acoustic model and the adapted acoustic models are shown in Table 2. In order to make a fair comparison with the broadcast news system, the $65K$ broadcast news language model was used during these recognition runs. Table 2 shows three word error rates for each model. The first WER is of the part of the audio in which WFH is speaking, the second one is based on speech from other people and the third is the overall word error rate.

Both adapted models perform better than the broadcast news model. Although the broadcast news model performs best on the small subset with various speakers (15% of the total amount of speech), the adapted models show improved WERs on the part of the data in which WFH is speaking. The SMAPLR adapted model (66.9% WER) outperforms the speaker dependent model (76.6% WER). The 78 minutes of speech used for training the speaker dependent model does not contain enough data for building a robust acoustic model.

	%WER	%WER	%WER
AM	WFH	other	total
UT-BN2005	81.6	67.2	80.4
WFH	76.0	83.2	76.6
BN/WFH	66.7	77.1	67.5

Table 2: *WERs of three acoustic models: the 2005 broadcast model, the WFH model and the SMAPLR adapted model. The second column shows the WER of the part in which WFH is speaking, the third the WER on the other speech parts of the evaluation set. The last column shows the total WER.*

The speaker adaptation we employed here, is a so-called ‘supervised adaptation.’ The segmentation into speakers (WFH and non-WFH) has been performed manually, and the acoustic models have been adapted to the speaker WFH using transcribed text. Both the segmentation and the speaker adaptation can in principle be performed automatically, or ‘unsupervised.’ For speaker segmentation an acoustical segmentation/clustering algorithm can partition the audio stream in segments belonging to the same speaker [**?**]. Using a first speech recognition pass, an automatic transcript can be generated. For each cluster of audio-segments spoken by the same speaker, this automatic transcript can be used to adapt the speaker-independent acoustic models to cluster-dependent models. These models can then be employed for a second speech recognition pass for the speech segments in that cluster, in order to obtain more accurate transcripts. Our supervised SMAPLR experiments give an impression of the maximum achievable performance increase, reducing the WER from 81.6 % to 66.7 %, for speaker WFH.

4.5 Overall results

The $30K$ mixture language model and acoustic models described in the previous sections were combined and tested on the evaluation set. Table 3 shows the word error rates of these combined systems.

The combination of the 30K mixture language model and the SMAPLR adapted acoustic model results in the best system performance: 66.9% WER.

Having a diarisation system available that could produce a series of time marks with associated speaker labels (“who spoke when”), the SMAPLR adapted system could be deployed for decoding the speech of

WFH and the broadcast news system for recognition of other speech. Given a diarisation system that would produce our manual annotated segmentation, the word error rate would improve to 66.6%.

	%WER	%WER	%WER
AM	WFH	other	total
WFH	73.8	83.6	74.6
BN/WFH	66.4	72.5	66.9

Table 3: *The word error rates (WER) of the two adapted acoustic models combined with the 30K mixture language model. The first row contains the AM trained on WFH solely. The second row contains the SMAPLR adapted acoustic model.*

By creating a mixture LM and a speaker dependent AM the word error rate was reduced with 13.5% (16.8% relative). To determine possible further improvements, a brief error analysis was conducted as described in the next section.

4.6 Error analysis

To investigate to what extent different audio conditions influence the word error rate, speech segments were classified into five classes: clean speech (F_0), speech with audible echo (F_1), speech containing background music (F_2), speech with background noise (F_3) and overlapping speech (speech interrupted by other speakers, F_4).

Table 3 shows the word error rates in each of the conditions. Two third of the segments in the WFH evaluation set are classified as 'clean'. One third contains echo, music, noise or overlapping speech. Although all speech in the music class is clearly understandable for human listeners music increases WER substantially, in the WFH task by more than 10%, which is comparable with the statistics reported in [11].

	%WER	%WER	%WER
Class	WFH	other	total
F_0	63.9	61.8	63.8
F_1	75.4	100	76.5
F_2	76.1	86.1	78.6
F_3	82.4	83.3	82.4
F_4	100	100	100

Table 4: *The word error rates of the five manually classified parts of the WFH evaluation set.*

As the goal of this project is to automatically create speech transcripts that can be used for indexing and retrieval, a stop-word list was used to filter out function words, hesitations and other words that are not helpful during search from the speech recognition transcripts of the development set. Only 25% of all words remain after removing stop words (60% of all unique words). The out-of-vocabulary rate of the filtered text is higher than the original OOV (9.1%). It was hypothesised that these OOV words are typically domain specific words that might be learned from additional domain specific text material. Going manually through the OOV list though, revealed that only a small number of words and names can be regarded as domain specific words. The majority of the OOVs appeared to be either misspelled words, highly infrequent words or words that can not directly be related to the task domain.

5 Search functionality

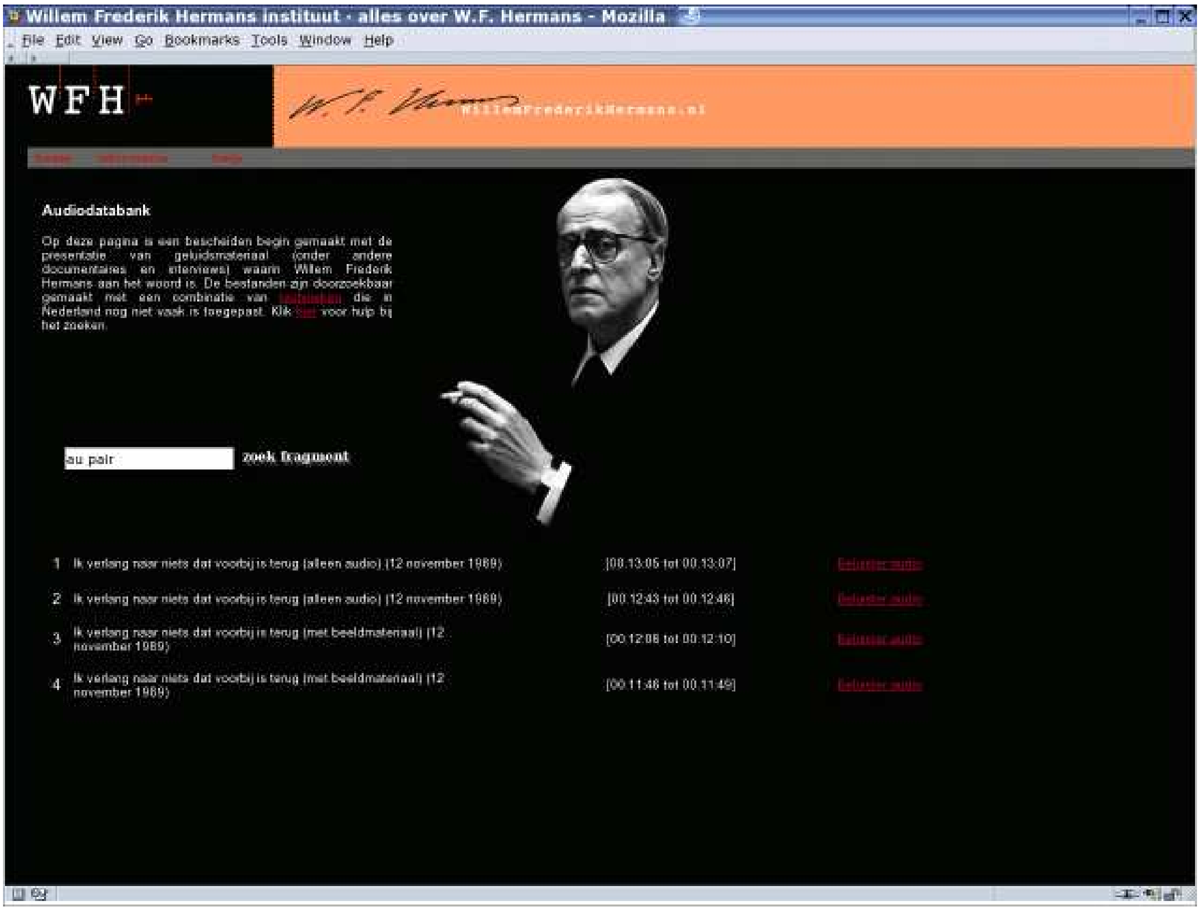
Automatically obtaining transcriptions of the audio files was the main goal of this project. The *search functionality* and the *document presentation* of the SDR system are implemented using standard software packages.

A *mySQL* server is used to store the transcriptions. Each segment (sub-document) is stored in a single record containing start- and end time, its document ID and the text itself. For the text field, full-text search is enabled. For each document (entire audio file) the title and a short description are stored.

This basic database provides for full-text search on the transcriptions. The time codes of each segment are returned to the user, so that the *document presentation* system can link the query results to the actual audio fragments.

6 Presentation

The search page of the web-portal contains a single input text field which can be used for quering the database. The result of a query, a list of sub-documents, is printed to the screen (see figure 1). Each item in the list contains the title of the document to which the sub-document belongs, the length of the audio fragment and a link to the audio fragment itself.


Willem Frederik Hermans instituut - alles over W.F. Hermans - Mozilla
File Edit View Go Bookmarks Tools Window Help
WFH
WillemFrederikHermans.nl
zoek fragment

Audiodatabank

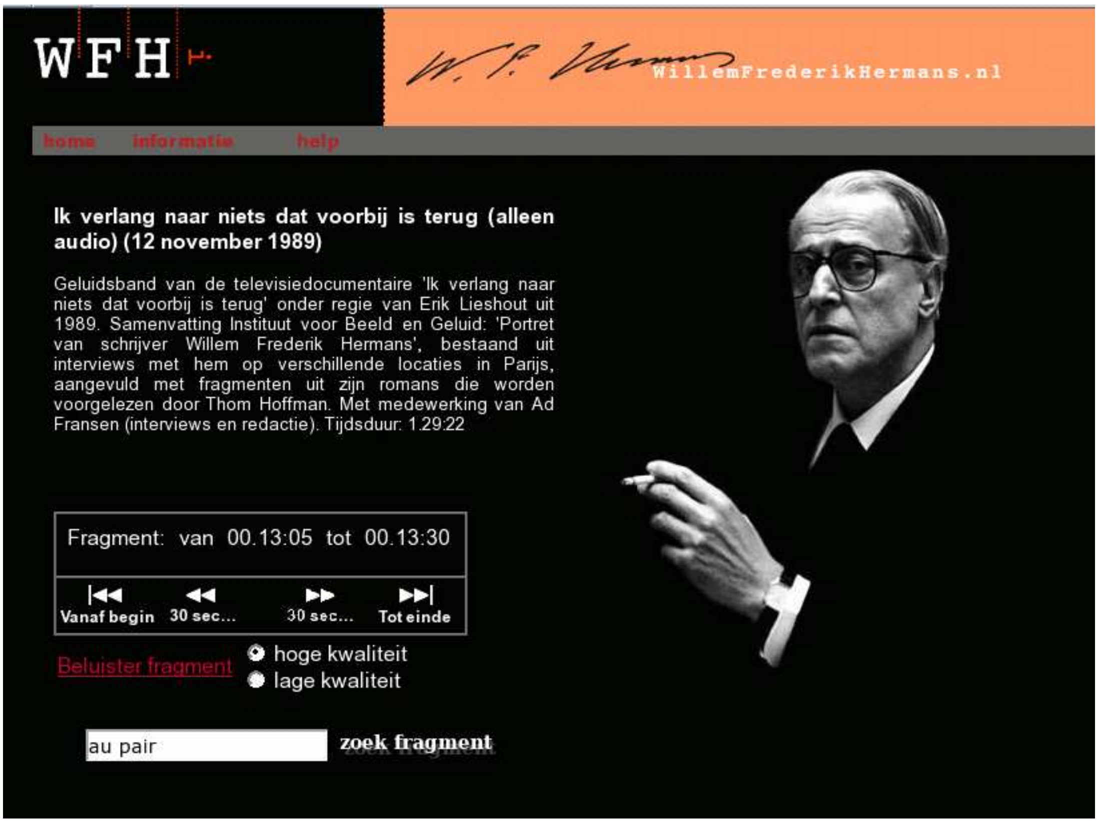
Op deze pagina is een bescheiden begin gemaakt met de presentatie van geluidsmateriaal (onder andere documentaires en interviews) waarin Willem Frederik Hermans aan het woord is. De bestanden zijn doorzoekbaar gemaakt met een combinatie van technieken die in Nederland nog niet vaak is toegepast. Klik hier voor hulp bij het zoeken.

au pair

1	Ik verlang naar niets dat voorbij is terug (alleen audio) (12 november 1989)	[00.13:05 tot 00.13:07]	Beluister audio
2	Ik verlang naar niets dat voorbij is terug (alleen audio) (12 november 1989)	[00.12:43 tot 00.12:46]	Beluister audio
3	Ik verlang naar niets dat voorbij is terug (met beeldmateriaal) (12 november 1989)	[00.12:08 tot 00.12:10]	Beluister audio
4	Ik verlang naar niets dat voorbij is terug (met beeldmateriaal) (12 november 1989)	[00.11:46 tot 00.11:49]	Beluister audio

Figure 1: An example of a query result. All sub-documents are shown in a single list.

The audio and video is stored on the *surfnet* server (www.surfnet.nl) and can be streamed to the web-portal. It is possible to stream the content in low or high quality. Because of technical reasons the begin and end time of the streamed fragment need to be given before the streaming is started. Therefore, we have created a web page with some basic controls that can make the audio or video fragment a bit shorter or longer before actually playing the fragment. In figure 2, part of this page is shown. To make the interface as intuitive as possible (despide the technical restrictions), well known symbols for rewinding and forwarding audio are used. As can be seen in figure 2, for each fragment a short description of the entire document is provided.



Ik verlang naar niets dat voorbij is terug (alleen audio) (12 november 1989)

Geluidsband van de televisiedocumentaire 'Ik verlang naar niets dat voorbij is terug' onder regie van Erik Lieshout uit 1989. Samenvatting Instituut voor Beeld en Geluid: 'Portret van schrijver Willem Frederik Hermans', bestaand uit interviews met hem op verschillende locaties in Parijs, aangevuld met fragmenten uit zijn romans die worden voorgelezen door Thom Hoffman. Met medewerking van Ad Fransen (interviews en redactie). Tijdsduur: 1.29:22

Fragment: van 00.13:05 tot 00.13:30

● hoge kwaliteit
○ lage kwaliteit

au pair

Figure 2: The controls for adjusting some basic settings before playing an audio or video stream. The user can change the quality (bandwith) of the stream and the start and end time.

7 Discussion

The Willem Frederik Hermans case study revealed that simply deploying the broadcast news system for transcribing oral history data resulted in high error rates of around 80% WER. In order to improve on the BN system, several adaptation schemes were applied on the acoustic level and on the language model level that indicated that a maximum achievable performance increase, reducing the WER from 81.6 % to 66.7 % for speaker WFH, is achievable. The overall word error rate was 66.9%, a 13.5% absolute improvement on the baseline BN system. Although near perfect transcripts are not required for a successful application of spoken document retrieval, error rates in this range are clearly below threshold.

It has already been noted that the large variability in audio quality, speech characteristics and topics are typical for the oral history domain and make the successful application of speech recognition technology difficult. A descriptive study on the characteristics of a certain collection is an important minimal prerequisite for identifying useful developments strategies.

What makes a successful application even more difficult is the fact that for the oral history domains we have seen until now, related audio and text sources that could be used for adapting the speech recognition components to these domain characteristics are usually only minimally available. This can be due to the fact that oral-history archives have limited resources so that links to useful metadata are simply missing, or to the historical nature of the collections. For example, in the ECHO collection we could only use contemporary newspaper texts to model the ancient, out-dated speech of the Dutch Queen Wilhelmina

(1880-1962), as there were no example text data digitally available that could be used to model this type of speech. An attempt to apply OCR techniques on related historical text data failed because of the low quality of the paper copies. Next to text (language model) related problems, ancient or dialectic speech that does not or only minimally occur in contemporary speech training databases, impose additional constraints to the effort to obtain an adequate speech recognition performance. A first step towards the successful application of the automatic disclosure of oral-history collections, should therefore be to collect (from a speech recognition developer point of view) or make available (from a content provider point of view) as much related data sources as possible for fine-tuning the system. A strategy to deal with the lack of acoustic training data is deploying (partly) unsupervised training strategies.

Other topics that need to be addressed are related to the retrieval functionality proper. Dependent on the users that are expected to search the collections, this functionality may need to be adapted. For example, users may use highly selective, domain specific-words in their queries that are infrequent in the material. Because of their low frequency, such words have a low chance of being selected for the speech recognition vocabulary and thus become so called 'query out-of-vocabulary'. Applying a word spotting approach to search for such words would then be an option. Another example concerns the presentation of the retrieval results. Presenting a user with short excerpts from the collection that contain query words may not be very informative. Instead, providing coherent fragments that are structured according to speaker or topic may be preferred.

It can be concluded that applying audio mining techniques for the disclosure of oral-history collections is a promising approach. Proof-of-concept has already been provided in other domains. However, due to the typical characteristics of the oral history domain, substantially more effort must be directed towards obtaining speech transcripts that can be used adequately for indexing. Research aiming at the optimisation of presentation strategies, of interest for spoken-word collections in general, could also boost the usability of audio mining considerably.

8 Appendix

1. Over Age Bijkaart. Een gesprek met Willem Frederik Hermans (omstreeks 1980) Freddy de Vree in gesprek met Willem Frederik Hermans over zijn alter ego Age Bijkaar t en over het verschijnsel columnist. In dit gesprek leest Hermans ook het artikel 'D okter doet zijn best' van Age Bijkaart voor. Tijdsduur: 33:00 (hermansA)

2. Waarom schrijven? (lezing op Nijenrode, 1 augustus 1983) Opname van de voordracht 'Waarom schrijven', gehouden op 1 augustus 1983. Ter gelegen heid van de jaarwisseling 1983-1984 werd een bekorte versie van deze lezing uitgegeve n als nieuwjaarsgeschenk van Uitgeverij De Harmonie. Hermans draagt onder andere drie gedichten van Hendrik de Vries voor, die hij vergelijkt met novellen. De lezing loop t tot omstreeks 38:40; daarna beantwoordt Hermans vragen uit de zaal. Tijdsduur: 1.11 :08 (hermansB)

3. Eerste zinnen van romans (18 januari 1987) De rede 'Eerste zinnen van romans' werd door Willem Frederik Hermans uitgesproken op 18 januari 1987 bij de feestelijke heropening van het theater De Balie te Amsterdam, op initiatief van de Stichting Literaire Activiteiten Amsterdam (SLAA). Tijdsduur: 55 :35 (hermansC)

4. Het boek der boeken, bij uitstek (8 maart 1986) De rede 'Het boek der boeken, bij uitstek' werd door Willem Frederik Hermans uitgespr oken op 8 maart 1986 tijdens de presentatie van Winkler Prins Lexicon van de Nederlan dse letterkunde te Antwerpen. Deze opname dateert van een dag voor de presentatie, to en Hermans de tekst thuis nog een keer repeteerde. De tekst werd onder dezelfde titel uitgegeven door de Uitgeverij De Bezige Bij (1986). Tijdsduur 18:50 (hermansD)

5. Over onder andere Richard Simmillion en 'Een veelbelovende jongeman'. Een gesprek met Willem Frederik Hermans (datum onbekend) Een gesprek met Willem Frederik Hermans over onder ander het autobiografische persona ge Richard Simmillion, en ook over de personages Osewoudt en Dorbeck. Voorts over het verhaal 'Een veelbelovende jongeman' (uit Een landingspoging op New

Foundland), en h et titelverhaal uit Paranoia. Voorts onder andere over Maurice Gilliams en Charles B. Timmer. Tijdsduur: 15:50 (hermansE)

6. Over Nooit meer slapen. Een gesprek met Willem Frederik Hermans (datum onbekend) Freddy de Vree in gesprek met Willem Frederik Hermans over diens roman Nooit meer sla pen. Wanneer en waar dit gesprek heeft plaatsgevonden is niet bekend. Tijdsduur 26:18 (hermansF)

7. Over Au pair. Een gesprek met Willem Frederik Hermans (omstreeks 1989) Freddy de Vree in gesprek met Willem Frederik Hermans over diens roman Au pair. Tijds duur: 37:29 (hermansG)

8. Over Wittgenstein. Een gesprek met Willem Frederik Hermans (15 november 1990) Freddy de Vree in gesprek met Willem Frederik Hermans over de filosoof Ludwig Wittgen stein, naar aanleiding van de toekenning van een eredoctoraat aan Willem Frederik Her mans door de Universiteit Luik. Tijdsduur: 18:32 (hermansI)

9. Over Nederlands-Indië en koloniale literatuur. Een gesprek met Willem Frederik H ermans (omstreeks 1978) Freddy de Vree in gesprek met Willem Frederik Hermans over de Nederlandse bemoeienis met Indonesië, en over een aantal belangrijke en spraakmakende Nederlandse auteu rs die romans en beschouwingen schreven over Nederlandsch-Indië. Ter sprake kome n onder anderen Multatuli, E. du Perron, Herman Neubronner van der Tuuk, M.H. Szekely-Lulofs, P.A. Daum ('Maurits') en Franz Wilhelm Junghuhn. Tijdsduur 2.02:17 (hermansJ)

10. Uit talloos veel miljoenen (fragment, voorgelezen op 29 mei 1981) Willem Frederik Hermans leest een fragment voor uit zijn pas verschenen roman Uit tal loos veel miljoenen, 29 mei 1981, plaats onbekend. Na een kleine tien minuten voorlez en, beantwoordt Hermans nog enige vragen uit de zaal; het gaat daarbij onder andere o ver het legendarische pornografische tijdschrift The Pearl (1879-1880). Tijdsduur 11: 29 (hermansB)

11. Over De laatste roker. Een gesprek met Willem Frederik Hermans (juli 1991) Een gesprek met Willem Frederik Hermans over de verhalenbundel De laatste roker, en o ver het belang van non-fictie en naslagwerken voor de schrijver. Tijdsduur: 43:11 (hermansL)

12. Over de Tractatus logico-philosophicus van Ludwig Wittgenstein. Een gesprek met Wille m Frederik Hermans (omstreeks 1975) Naar aanleiding van het verschijnen van Willem Frederik Hermans' vertaling van Ludwig Wittgensteins Tractatus logico-philosophicus spreekt Freddy de Vree ruim een uur lan g met de vertaler over Wittgenstein. Daarbij komen ook andere filosofen, zoals Kant, Schopenhauer en Russell ter sprake, en tevens het eigen werk van Hermans, zoals het D e God Denkbaar Denkbaar de God en Het evangelie van O. Dapper Dapper. Hermans leest v anaf 57:50 een fragment uit zijn vertaling voor. Tijdsduur: 1.15:19 (hermansN)

13. Over De tranen der acacia's. Een gesprek met Willem Frederik Hermans (omstreeks 1976) Gesprek met Willem Frederik Hermans over zijn roman De tranen der acacia's uit 1949. Onder andere over de hoofdpersoon Arthur Muttah, de zoektocht naar een uitgever en de kritieken op het boek. Tijdsduur: 25:18 (hermansO)

14. Over het onderscheid tussen wetenschap en literatuur. Een gesprek met Willem Frederik Hermans (19 maart 1992) Raymond Benders en Hugo Bousset interviewen Willem Frederik Hermans over Uit talloos veel miljoenen en Herinneringen van een engelbewaarder, en over het onderscheid tusse n wetenschap en literatuur. Voorts onder andere over psychologie, geschiedenis, Lou d e Jong, A.L. Sötemann, Adriaan Morriën, Hugo Claus en Harry Mulisch. Aan he t slot vragen uit de zaal. Geluidsband van een bijeenkomst aan de Katholieke Universi teit Brussel, 19 maart 1992. Tijdsduur: 1.54:46 (hermansP)

15. Over De raadselachtige Multatuli. Een gesprek met Willem Frederik Hermans (1976) Een gesprek met Willem Frederik Hermans over het leven van Multatuli, kort vóór het verschijnen van De raadselachtige Multatuli in 1976. Tijdsduur 45:06 (hermansQ)

References

[1] K. W. Church, “Speech and Language Processing: Where Have We Been and Where Are We Going?” in *Eurospeech-2003*, Genève, Switzerland, September 2003.

[2] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C.Wong, “Mylifebits: fulfilling the memex vision,” in *ACM Multimedia*, 2002, pp. 235–238.

[3] J. Goldman, “Report of the EU/NSF working group on Spoken Word Audio Archives,” http://www.ercim.org/publication/ws-proceedings/Delos-NSF/SpokenWord.pd%f, 2003.

[4] “ECHO:http://pc-erato2.iei.pi.cnr.it/echo/.”

[5] W. Byrne, D. Doermann, and M. Franz, “Automatic Recognition of Spontaneous Speech for Access to Multilingual Oral History Archives,” *IEEE Transactions on Speech and Audio Processing, Special Issue on Spontaneous Speech Processing*, July 2004.

[6] J. Garofolo, C. Auzanne, and E. Voorhees, “The TREC SDR Track: A Success Story,” in *Eighth Text Retrieval Conference*, Washington, 2000, pp. 107–129.

[7] R. Ordelman, “Dutch Speech Recognition in Multimedia Information Retrieval,” Ph.D. dissertation, University of Twente, The Netherlands, October 2003.

[8] I. Schuurman, M. Schouppe, H. Hoekstra, and T. van der Wouden, “CGN, an Annotated Corpus of Spoken Dutch,” in *In Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, 2003.

[9] B. Pellom and K. Hacioglu, “Recent Improvements in the CU Sonic ASR system for Noisy Speech: The SPINE Task,” in *Proc. ICASSP*, 2003.

[10] O. Siohan, T. Myrvoll, and C. Lee, “Structural Maximum a Posteriori Linear Regression for Fast HMM Adaptation,” in *Computer, Speech and Language, 16*, 2002, pp. 5–24.

[11] B. Raj, V. N. Parikh, and R. M. Stern, “The Effects Of Background Music On Speech Recognition Accuracy,” in *Proc. of the ICASSP, Munich, Germany*, 1997.