NEWS TUNER: A SIMPLE INTERFACE FOR SEARCHING AND BROWSING RADIO ARCHIVES

Jon Marston, Gavin MacCarthy[*]

WBUR
890 Commonwealth Avenue
Boston, MA 02215, USA

Beth Logan, Pedro Moreno, JM Van Thong[†]

HP Labs CRL
One Cambridge Center
Cambridge, MA 02142, USA

ABSTRACT

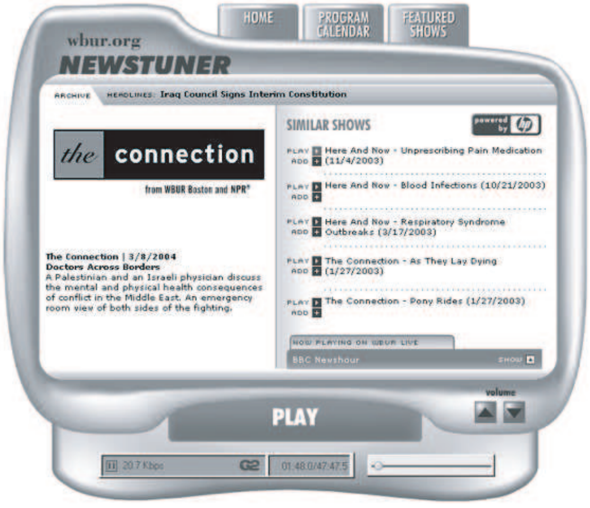
We present in this paper a new web-based application, called the News Tuner, for searching and browsing large radio archives. While popular search engines provide means for finding text and images, our approach combines semantic and acoustic search for efficient retrieval of audio documents. Semantic search allows the user to retrieve stories for a given concept, while acoustic search allows random access within stored audio files. Our experiments on over 1,700 programs show that our method is effective at quickly retrieving stories that would be difficult to find otherwise. The News Tuner paradigm is intended primarily for news and talk radio programs, however it may be applied to browsing and searching any spoken word audio content.

1. INTRODUCTION

Until recently radio existed only in the moment; once a program had finished its on-air broadcast it ceased to exist. Recording radio broadcasts and offering these archives over the Internet has created a second life for such programs. Unfortunately, the end-user experience of wading through archives looking for interesting programs can be a daunting task. Clearly there is a need for a means to search and browse archived audio programs. However, traditional search engines are often limited to text and image indexing, thus many multimedia documents, video and audio, are excluded from such retrieval systems.

The radio station WBUR[1] produces 30 hours of original programming per week which it permanently archives on its public website. Despite the wealth of topics covered in nearly 4,000 hours of archived programming, WBUR website visitors spend the vast majority of their time listening to the live broadcast at an 8 to 1 ratio[2]. The archives represent therefore an underutilized resource and improving their access would greatly enhance the end-users' experiences.

In this paper, we describe a new web audio player interface that provides efficient access to archived material. This player, called the *News Tuner* (see figure 1), combines live broadcasts, archived audio, chat, and studio cameras into a single application. Its simple and intuitive user interface is designed for browsing and searching large recordings of topical spoken-word audio. It takes advantage of semantic similaries among news items to suggest potential related stories to the listener. It does this without the need of selecting or entering any queries. The user is only asked to click a potential story.



HEADLINES: Iraq Council Signs Interim Constitution

The Connection | 3/8/2004
Doctors Across Borders
A Palestinian and an Israeli physician discuss the mental and physical health consequences of conflict in the Middle East. An emergency room view of both sides of the fighting.

Here And Now - Unprescribing Pain Medication (11/4/2003)
Here And Now - Blood Infections (10/21/2003)
Here And Now - Respiratory Syndrome Outbreaks (3/17/2003)
The Connection - As They Lay Dying (1/27/2003)
The Connection - Pony Rides (1/27/2003)

BBC Newshour

Fig. 1. The News Tuner interface.

In this paper, we focus on the indexing and searching aspects of our player. News retrieval has attracted much

[*] jmarston@wbur.bu.edu, gmaccart@wbur.bu.edu

[†] beth.logan@hp.com, pedro.moreno@hp.com, jm.vanthong@hp.com

[1] WBUR is a National Public Radio (NPR) affiliate based in Boston, online at *http://www.wbur.org*. The station's broadcast programming ranges from current events to classical literature and art.

[2] For the representative week of March 1st, 2004, Arbitron reported that of WBUR.org's 87,394 hours of streamed audio, only 14% of the online listening was to archived audio.

attention in the literature (e.g. [1], [2], [3], [4]) and commercial systems are now available. Most systems combine one or more automatic content analysis techniques for segmenting and annotating the documents with manually generated annotations. Media documents may then be searched and browsed like text, allowing the user to directly access the relevant parts of the content. In our system we use a combination of acoustic and semantic indexing techniques to provide a flexible interface to the audio archives.

2. THE NEWS TUNER SEARCHING AND BROWSING USER INTERFACE

The News Tuner offers two ways for users to find an audio file: keyword and similarity searches. Keyword search allows users to hunt through transcripts or production metadata of a collection of audio stories looking for an exact word match. Similarity search returns a list of audio files that contain similar content to a selected story according to its semantic closeness as described in section 3.2. Keyword searching is effective for users who want to find a specific piece of content while similarity matching is good for users browsing from topic to topic without a specific goal in mind.

The News Tuner is implemented as an HTML page with two embedded controls: a Flash user interface and a Real Player. A back-end server provides services for managing document processing and indexing. It is the primary entity for submitting media processing requests and subsequently analyzing content for indexing and topic classification. This component is a highly distributed application involving multiple processes running on large server farms [11]. A front-end document retrieval service processes requests from the News Tuner and extracts relevant documents from the index and the metadata repository. The different components of the front-end and back-end are implemented as web services, i.e. remote procedure calls made via SOAP-formatted messages over HTTP.

3. INDEXING RADIO PROGRAMS

Authored metadata, such as titles, abstracts, categorizations, and air dates, can be used to index radio program archives. In the past, WBUR has taken a hierarchical approach to categorizing groups of content, referred to as *stories*. In consultation with the story producers WBUR developed a list of 13 main categories and a set of 15 sub categories for each main grouping. Examples of main categories are *Arts and Leisure*, *Business and Economics*, and *Education*. Examples of sub categories for the *Arts and Leisure* main category are *Film*, *Theater*, *Visual Arts*, and so on. The purpose of classifying the content in this manner was to allow website visitors to browse content by topic and to offer suggestions for related material.

In practice, a rigid and static taxonomy proved to be an ineffective means of grouping similar content for browsing. First, stories frequently covered multiple topics but could not be assigned to multiple categories. Second, producers often disagreed on how to categorize stories. Different producers made different choices leading to inconsistency in the data. Third, website visitors were confused by topic headings and unable to find stories they were looking for. And last but not least, the fixed hierarchy was not flexible enough to respond to emerging events, even with updates and additions. To overcome these problems, our new approach relies upon automatic grouping content on an acoustic and semantic basis, as detailed below.

3.1. Acoustic indexing

A straightforward approach to implement content-based audio indexing consists of generating the transcription automatically using a large vocabulary speech recognition system, and then using information retrieval algorithms to index the corresponding textual documents. The index can then be used to retrieve relevant portions of the audio documents using standard word query terms. HP's audio search engine, *SpeechBot*[3], is an example of such a system [5]. Although speech recognition technology is inherently inaccurate, particularly when the audio quality is degraded due to poor recording conditions and compression schemes, it has been shown that satisfactory retrieval accuracy can be achieved [6].

Searching the acoustic space has the advantage of returning the precise location within audio documents where the words have been uttered. However, this method fails if the query terms have not been spoken, were misrecognized, or are out of vocabulary (OOV). Studies show that a large number of words occur only on a single day over a long period of time [7]. These words are often proper nouns, and may represent a significant proportion of the query terms [6]. The problem can be surmounted with various techniques ranging from vocabulary adaptation, subword indexing, to acoustically similar query term matching, and query expansion. Alternatively topic similarity methods can be used to retrieve and browse semantically related audio or text documents as detailed in the next section.

3.2. Semantic indexing

In addition to keyword searching, the News Tuner allows users to browse by topic. Because of problems with manually generated categories, as previously described, we use an automatic technique to derive topic similarity information. Our approach is based on Probabilistic Latent Semantic Analysis (PLSA) [8] which we have previously found

[3] see www.speechbot.com

works well, even on speech recognition transcripts [9], [10].

The PLSA model provides a way of transforming documents represented in a very high-dimension ‘bag-of-words’ space, where each dimension represents a word, to a much lower dimension topic space. Each dimension in this new space represents a concept in the training data. For the current application, the system learnt 128 topics, roughly matching the number of subcategories initially used by WBUR. In effect the technique automatically finds clusters that map to natural concepts as shown in table 1.

Topic 0	**Topic 1**
health, care, hospitals, medical, federal, money, million, hospital, medicaid, hughes, services, people, program, government, department, system	africa, african, congo, nigeria, rwanda, west, diamond, country, rebels, kabila, power, station, people, two, pipeline, ouganda, years, south

Table 1. Examples of the most likely words for 2 topics.

Given these learnt topic models, we can convert the metadata for each indexed story to a 128-dimension topic vector. If the stories are annotated by producers, we use this as our metadata, otherwise we use speech recognition transcripts. Each dimension in the vector gives the probability of the story belonging to each of the trained topics. We then use the L1 distance to compute the distance between bags of words in this new space. This distance reflects likely semantic closeness between the stories, or between queries and stories.

Semantic indexing overcomes some of the problems of acoustic search by returning related documents even if the exact query terms are not in the story summary, nor spoken in the audio document. However, it is less effective for short queries and there is no guarantee that the terms will be in the returned documents. Approaches that combine both acoustic and semantic search have shown improvements [10] and are the subject of ongoing work.

4. EXPERIMENTAL RESULTS

The accuracy of the topic similarity algorithm was evaluated using the WBUR category metadata. The subcategory information was not considered, because it would have introduced too much noise in the measure. We used humans to derive a mapping between the 128 machine-generated topics and the 13 WBUR categories. Each topic was characterized by its 50 most likely words learnt by the topic training algorithm. For example, topic 0 is characterized by *”health care hospitals medical federal”*, as shown in table 1. The training set for learning topics consists of 95,000 *New York Times* text articles from 1998, 1999 and the first 5 months of 2000. This provides us clean and well segmented data to train the models. The topic model dictionary is around 250,000 words and does not contain stop words.

Five users examined the top 50 words of each topic and independently assigned to it one of the 13 WBUR categories. For example, all users assigned the category *Health* to this topic. A total of 1,711 annotated stories from shows such as *The Connection* and *Here and Now* were then automatically categorized. For each story, we computed a topic probability vector using a bag of words composed of the story title and description. This 128-dimension topic vector was then collapsed into a 13 dimension category vector using the mapping defined by our users. For 119 topics, the majority of users selected the same category so we used this for the mapping. For the remaining 9 topics where there was no clear winner, we arbitrarily chose the selection of one of the users. We then examined the probability of each category, comparing it to the categories assigned by WBUR.

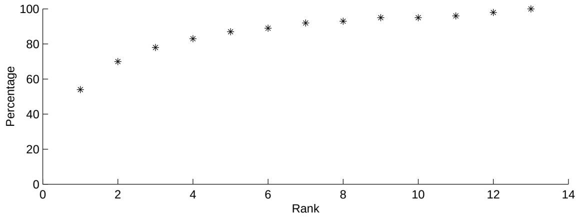


Fig. 2. Fraction of times the rank of the WBUR chosen category was less or equal to n, with n varying from 1 to 13.

Figure 2 shows the percentage of documents for which the most probable top n categories match that assigned by WBUR where n varies from 1 to 13. The results show that for 70% of the documents, the correct category is assigned rank 1 or 2. The average rank of the WBUR chosen category is 2.71. We see then that although there is not full agreement, our topic models are on average capable of automatically assigning reasonable categories and are therefore useful for similarity browsing.

5. CONCLUSIONS AND FUTURE WORK

We presented a new web-based application for browsing radio program archives. This approach is unique in its kind, combining semantic and acoustic searching on post-produc/tion metadata and textual spoken transcriptions. Our early experiments show that the News Tuner is intuitive and easy to use, helping the user to discover rapidly and accurately stories of interest. Semantic search allows retrieval of stories about the same concept, while acoustic search allows users random access within an audio stream. The semantic search has some significant advantages. First the models are easier to adapt; Adding or removing new important words and semantic clusters reflects the changing nature of world news. Subsequently the use of larger dictionaries alleviates the problem of out-of-vocabulary query terms. Second, the text of existing stories is used as queries for the semantic similarity engine, freeing the user from selecting and entering query terms, making the search of related documents simple, intuitive and effective. And last, there is no need for static taxonomies.

We plan to improve the current system in several directions. First, we will continue to improve the current semantic search by topic similarity. The topic models need to be incrementally trained on a daily basis with a source of data closely related to radio broadcasts. Updating the dictionary for topic models is much easier than for a speech recognizer. Moreover, the story topic probability vectors can be easily recomputed with new models, whereas it is not realistic to re-process every audio recording in the archives to generate new spoken transcriptions. This approach will help to alleviate the out-of-vocabulary problem, as well as providing better similarity browsing. Second, we will look at alternate methods to improve the speed of a semantic query (i.e. when we cannot pre-compute the topic probability). The current implementation is a straightforward comparison of distances with every stored story. Although fast, it will not scale to very large repositories. Finally, we will explore methods for combining acoustic and semantics indexes. This should allow to further improve information retrieval accuracy.

6. ACKNOWLEDGMENTS

Experiments on a large scale (over 3,000 radio shows) have been made possible thanks to the collaboration between a well established content provider (WBUR) and an industrial research laboratory (HP CRL) with a long experience with search engines (e.g. *Digital AltaVista,* and *Digital-Compaq-HP SpeechBot*). The design of the WBUR News Tuner is inspired by the work of Geir Gaseidnes, Laura Russell, and others at NetNumina Consulting. We would like to thank Robin Lubbock of WBUR for investing the time and energy in supporting long term development projects like the News Tuner. We also thank Scott Blackwell, Chris Weikart of HP, and Will Smith of WBUR for providing us with their talented engineering support. They have been principal contributors to the design and implementation of the system.

7. REFERENCES

[1] H. D. Wactlar, A. G. Hauptmann, and M. J. Witbrock, "Informedia: News-on-demand experiments in speech recognition," in *DARPA Speech Recognition Workshop*, 1996.

[2] S. E. Johnson, P. Jourlin, G. L. Moore, K. S. Jones, and P. C. Woodland, "The Cambridge University spoken document retrieval system," in *Proc. IEEE ICASSP*, 1999.

[3] D. Abberley, G. Cook, S. Renals, and T. Robinson, "Retrieval of broadcast news documents with the THISL system," in *Proc. TREC-8*, 1999.

[4] P. C. Woodland, T. Hain, S. E. Johnson, T. R. Nielser, A. Tuerk, and S. J. Young, "Experiments in broadcast news transcription," in *Proc. IEEE ICASSP*, May 1998.

[5] JM. Van Thong, D. Goddeau, A. Litvinova, B. Logan, P. Moreno, and M. Swain, "Speechbot: a speech recognition based audio indexing system for the web," in *Proceedings RIAO*, 2000.

[6] B. Logan, P. Moreno, JM. Van Thong, and E. Whittaker, "An experimental study of an audio indexing system for the Web," in *Proc. ICSLP*, 2000.

[7] E. W. D. Whittaker, "Temporal adaptation of language models," in *Proc. of the 2001 ISCA Workshop on Adaptation Methods for Speech Recognition*, 2001.

[8] T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR*, 1999.

[9] D. Blei and P. J. Moreno, "Topic segmentation with an aspect hidden markov model," in *SIGIR*, 2001.

[10] B. T. Logan, P. Prasangsit, and P. J. Moreno, "Fusion of semantic and acoustic approaches for spoken document retrieval," in *Proc. ISCA Workshop on Multilingual Spoken Document Retrieval*, 2003.

[11] H. Mandviwala, S. Blackwell, C. Weikart, and JM Van Thong, "Multimedia content analysis and indexing: Evaluation of a distributed and scalable architecture," in *SPIE's International Symposium on ITCom 2003*, 2003.