

Speech for Multimedia Information Retrieval

Alexander G. Hauptmann, Michael J. Witbrock, Alexander I. Rudnicky
and Stephen Reed

Carnegie Mellon University, School of Computer Science
5000 Forbes Avenue, Pittsburgh, PA 15213-3890
(412)268-1448: alex@cs.cmu.edu

ABSTRACT

We describe the Informediatm News-on-Demand system. News-on-Demand is an innovative example of indexing and searching broadcast video and audio material by text content. The fully-automatic system monitors TV news and allows selective retrieval of news items based on spoken queries. The user then plays the appropriate video "paragraph". The system runs on a Pentium PC using MPEG-I video compression and the Sphinx-II continuous speech recognition system [6].

KEYWORDS: video information retrieval, speech recognition, News-On-Demand, multimedia indexing and search, Informedia }

GOALS

The Informedia Digital Video Library Project [4] at Carnegie Mellon University is creating a digital library of text, images, videos and audio data available for full content retrieval. Through the integration of technologies from the fields of natural language understanding, image processing, speech recognition and video compression, the Informedia System [1] allows a user to explore multi-media data in depth as well as in breadth. An overview of the structure of the Informedia system is shown in Figure 1. The Informedia system for video libraries goes far beyond the current paradigm of video-on-demand, retrieving and displaying short video paragraphs in response to the user's query. As a result, a large body of video material can be searched with very little effort.

While our work is centered around processing news stories from TV broadcasts, the system exemplifies an approach that can make any video, audio or text data accessible. The same methods can be used to index and search other streamed multi-media data by content. *News on Demand*. One particularly significant application of the Informedia Digital Video Library deals with television and radio news.

The current limitations. Currently, the TV and radio

news is broadcast at a particular time, and if a person is not in front of a TV or radio at that time, the information becomes virtually inaccessible. There is simply not enough time to scan through tapes of yesterday's news for relevant stories. In addition, the viewer must watch all stories in a news show, without the ability to select which stories to skip and which to pursue in more detail.

Figure 1: Overview of the Informedia Digital Video Library

Furthermore, a person can only attend to one news channel at a time. Related information broadcast on another news channel at the same time cannot be viewed.

The solution. The solution is to compress and digitally store news broadcasts on computer. All information is made accessible through interactive queries. These queries allow the user to retrieve relevant segments from all the programs that carried stories on the topic of interest.

SPEECH RESEARCH ISSUES

We can distinguish two distinct phases in the Informedia News-On-Demand process: library creation and library exploration. Library creation deals with the accumulation of information, transcription, segmentation and indexing.

Library exploration concerns the interaction with the user trying to retrieve selections in the database.

Speech recognition for library creation. During library creation, speech recognition helps create a time-aligned transcript of the spoken words as well as segmenting the broadcast into video paragraphs. A portion of the news stories also have close-captioning text available. However, the close-captioned data, if available, may lag behind the actual words spoken. These problems, as well as inaccuracies in transcription are especially glaring when the broadcast is "live". In News-on-Demand we use speech recognition in conjunction with close-captioning to improve the time-alignment of the transcripts and to correct for gross errors.

For the news broadcasts that are not close-captioned, we need a transcript generated exclusively by the speech recognition system. The vocabulary and language model used here approximate a "general American news" language model. It was based on a large corpus of North American business news from 1987 to 1994 [5].

Spoken queries for library exploration. During library exploration, the Sphinx-II [7] speech recognition allows a user to query the system by voice, simplifying the interface by making the interaction more direct. The integration of speech with the interface enhances access to the stored video data by allowing more immediate and direct entry of queries. The language model used during exploration is similar to [5] but emphasizes key phrases frequently used in queries such as "How about", "Tell me about", "Is there anything about", etc.

THE NEWS-ON-DEMAND SYSTEM

Unlike the other Informedia prototype [2, 3] which is designed as an educational testbeds created with human assistance, this system is fully automated. Of course, a fully automated system is likely to contain errors. We distinguish 5 types of errors, all of which are areas of active research:

- False segmentation of stories. This happens when we incorrectly identify the beginning and end of a video paragraph associated with a single news story. Incorrect segmentation is usually due to inaccurate transcription by speech recognition or close-captioning.
- False words in transcripts. Errors in the transcript are either the result of faulty speech recognition or errors in close-captioned text.
- False synchronization. These errors are retrieved words that were actually spoken elsewhere in the video as a result of faulty alignment to the close-captioned text.
- Incorrectly recognized query. This is the result of an incorrect speech recognition during the library exploration.

The user can correct misrecognitions through typing, repeating or rephrasing the query.

- Incorrect stories returned for a query. This type of error indicates lack of information recall or precision of the search module.

CONCLUSIONS

Despite the drawbacks of a fully automated system, the benefits are very dramatic. We can finally navigate the complex information space of news stories, without the linear access constraint that normally makes this process so time consuming. Thus Informedia News-on-Demand provides a new dimension in information access to video and audio material.

REFERENCES

1. Stevens, S., Christel, M., and Wactlar, H. "Informedia: Improving Access to Digital Video.". *Interactions 1*, 4 (October 1994), 67-71.
2. Christel, M., Stevens, S., and Wactlar, H. Informedia Digital Video Library. Proceedings of the Second ACM International Conference on Multimedia, New York, October, 1994, pp. 480-481. Video Program.
3. Christel, M., Kanade, T., Mauldin, M., Reddy, R., Sirbu, M., Stevens, S., and Wactlar, H. "Informedia Digital Video Library". *Communications of the ACM 38*, 4 (April 1994), 57-58.
4. The CMU Informedia Project.
<http://www.contrib.andrew.cmu.edu/usr/hw2b/informedia/informedia.html>.
5. Rudnicky, A.I. Language modeling with limited domain data. Proceedings of the 1995 ARPA Workshop on Spoken Language Technology, 1995. in press.
6. The CMU Speech Project.
<http://www.cs.cmu.edu/afs/cs.cmu.edu/user/air/WWW/SpeechGroup/Home.html>.
7. Hwang, M.-Y., Rosenfeld, R., Thayer, E., Mosur, R., Chase, L., Weide, R., Huang, X. and Alleva, F. Improving Speech Recognition Performance via Phone-Dependent VQ Codebooks and Adaptive Language Models in SPHINX-II. Proceedings of ICASSP-94, 1994, pp. 549 - 552.

