Opening Up IIHS Video with SpokenMedia

Brandon Muramatsu[1], Andrew McKinney[1] and Peter Wilkins[1]

[1]Office of Educational Innovation, Massachusetts Institute of Technology

mura@mit.edu, mckinney@mit.edu, pwilkins@mit.edu

Abstract

The Indian Institute for Human Settlements (IIHS, www.iihs.co.in) and the SpokenMedia (spoken-media.mit.edu) team from the MIT Office of Educational Innovation and Technology (OEIT) have been discussing how SpokenMedia technologies might be used by IIHS to provide cost effective ways of making video/audio course materials accessible to the diversity of students expected by IIHS.

The SpokenMedia project is developing a set of tools and services to enable OpenCourseWare and Open Educational Resource providers to automatically transcribe lecture video for search and to enable innovative interactions through rich media notebooks. The SpokenMedia tools take audio from lectures and processes it to create a time-coded transcript. The transcript can be used for search—extending search from a few words of metadata to the full text of a video. The transcript can also facilitate improved accessibility—the text can be used as a replacement for the audio or for translation into regional languages in India in IIHS' case. The transcript is used in a video player that enables learners to interact with lecture video in more educationally relevant ways than are currently available.

This paper presents a case study of the proof-of-concept demonstration SpokenMedia developed for IIHS.

Keywords

SpokenMedia, automatic lecture transcription, rich media notebooks

1 Introduction

SpokenMedia prepared a demonstration for the Indian Institute for Human Settlements January 2010 Curriculum Conference in Bangalore, India to show how video linked with transcripts can provide an innovative learning interface to enable IIHS to effectively use video as part of its curriculum.

1.1 Participants

1.1.1 Indian Institute for Human Settlements

The Indian Institute for Human Settlements (IIHS, www.iihs.co.in) is a nascent university that "will create India's first independent National Innovation University focused on the challenges and opportunities of its urbanisation." IIHS is committed to "recasting the role of the university education in light of an open world presents a value proposition far more profound than the free dissemination of educational tools and resources – it allows [IIHS] to *proactively construct new preferred learning.*" With 23 official languages in India and IIHS' goal to use rich media throughout its curriculum, IIHS "will fail or scale based on language." (IIHS, 2010) IIHS intends to provide its curriculum in the breadth of languages in India. A particular challenge is providing rich media, such as video, in the wide range of languages in India cost-effectively.

1.1.2 SpokenMedia Project

SpokenMedia is a MIT OEIT project that is exploring the development and use of rich media notebooks for teaching and learning. The SpokenMedia project is guided by two questions:

- How can SpokenMedia better support learners searching for and finding relevant video content?
- How can SpokenMedia provide innovative tools enabling learners to better use and interact with video content?

SpokenMedia builds upon research into automatic lecture transcription by Jim Glass and his Spoken Language Systems group in the Computer Science and Artificial Intelligence Laboratory at MIT. Partially funded by the iCampus MIT/ Microsoft Alliance, in the SpokenLecture project, Jim Glass and his researchers developed the technology that enables the project to automatically transcribe lecture video. The researchers have focused on the unique aspects of automatic speech recognition on lecture video to develop a system capable of as high as 85% accuracy.

2 Presenting IIHS Video through SpokenMedia

As a proof-of-concept, SpokenMedia automatically generated transcripts in a video player for 24 IIHS video segments (spokenmedia.mit.edu/demo/iihs/). The automatically generated transcripts were used as the basis for two 99% accurate transcripts and the Hindi translations for those transcripts. Each video was approximately 5 minutes in duration, and was spoken in English by speakers from a wide range of backgrounds (regional dialects and accents).

2.1 Transcribing and Translating the Videos

SpokenMedia, building on the software transferred from the Spoken Lecture project, processed the IIHS videos to automatically create lecture transcripts and make them available via a player. Figure 1 shows the overall workflow for automatically transcribing lecture video. The highlighted boxes are the ones used as part of the proof-of-concept.

With the transcripts automatically created for the IIHS videos, IIHS and OEIT examined them for accuracy. The recognition on men and women, native and non-native English speakers with backgrounds in the United States, India and the United Kingdom ranged from 40-60% accuracy. It is important to keep in mind that the process used the generic acoustic model (which is optimized for relatively un-accented Americans speaking English) and the generic domain model, the resulting low accuracy was not a surprise. Without editing, this low accuracy is probably not acceptable for search, accessibility or to facilitate translation.

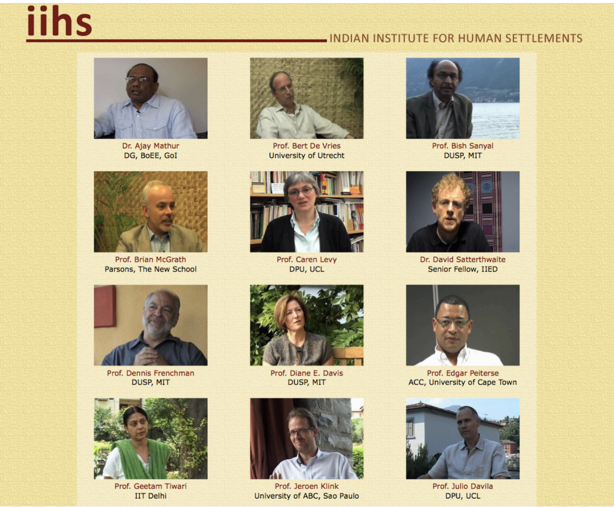


Figure 2. IIHS Video Page

OEIT worked with IIHS to edit two of the transcript files (Professors Bish Sanyal and Geetam Tiwari) for 99% accuracy and then translate the resulting text into Hindi. During the editing process it was important to keep the relationship between the words and the start time to enable the player to properly display the word at the time it is spoken. This requirement, coupled with our lack of specialized tools, made the task of editing the transcripts especially challenging. The low accuracy and lack of tools meant that in creating the 99% accurate transcript it was easier to start over and then manually align the words with time codes. IIHS staff also translated the 99% accurate English transcript into Hindi.

OEIT presented all of the transcribed videos (see Figure 2) for review by IIHS faculty advisors and consultants at the January 2010 Curriculum Conference in Bangalore, India.

2.2 Viewing the Videos

The SpokenMedia project developed a video player with the following capabilities:

- Playback of Flash/QuickTime videos.
- Transcript text linked to time code and video playback.

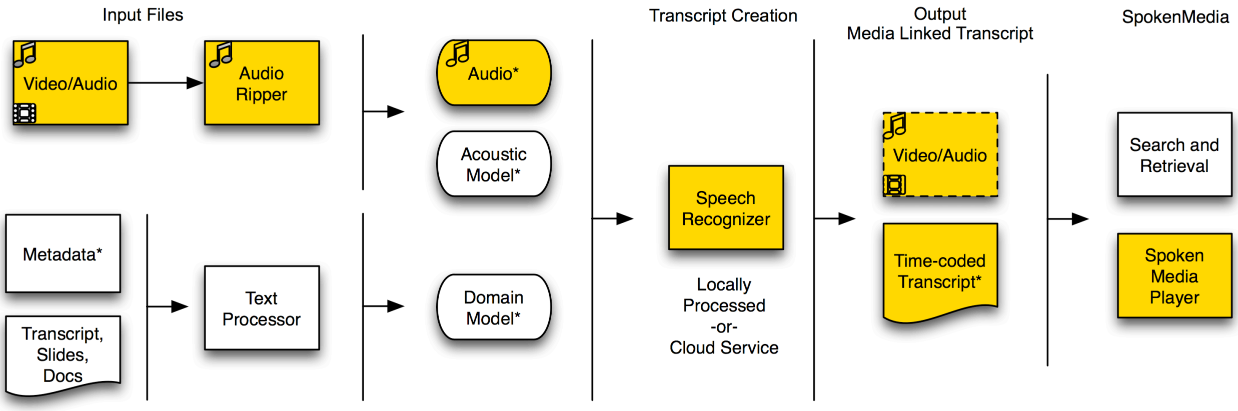
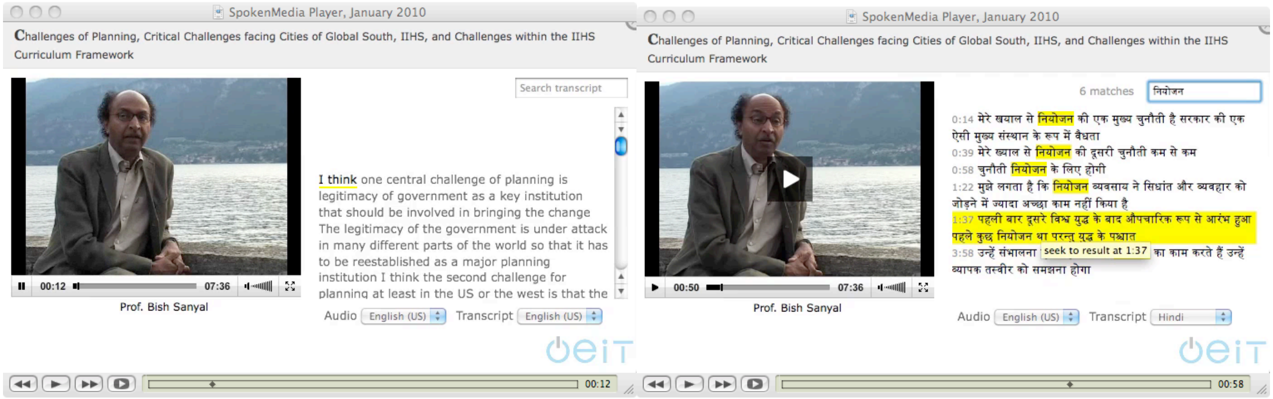


Figure 1. SpokenMedia Workflow—Highlighted Items Indicate the Steps used in the IIHS Proof-of-Concept



Challenges of Planning, Critical Challenges facing Cities of Global South, IIHS, and Challenges within the IIHS Curriculum Framework

I think one central challenge of planning is legitimacy of government as a key institution that should be involved in bringing the change The legitimacy of the government is under attack in many different parts of the world so that it has to be reestablished as a major planning institution I think the second challenge for planning at least in the US or the west is that the

Audio: English (US) Transcript: English (US)

Challenges of Planning, Critical Challenges facing Cities of Global South, IIHS, and Challenges within the IIHS Curriculum Framework

Search: नियोजन

0:14 मेरे खयाल से नियोजन की एक मुख्य चुनौती है सरकार की एक ऐसी मुख्य संस्थान के रूप में वैधता
0:39 मेरे ख्याल से नियोजन की दूसरी चुनौती कम से कम
0:58 चुनौती नियोजन के लिए होगी
1:22 मुझे लगता है कि नियोजन व्यवसाय ने सिधांत और व्यवहार को जोड़ने में ज्यादा अच्छा काम नहीं किया है
1:37 पहली बार दूसरे विश्व युद्ध के बाद औपचारिक रूप से आरंभ हुआ पहले कुछ नियोजन था परन्तु युद्ध के पश्चात
3:58 उन्हें संभालना का काम करते हैं उन्हें व्यापक तस्वीर को समझना होगा

Audio: English (US) Transcript: Hindi

Figure 3. SpokenMedia Player—English Transcript with "Bouncing Ball" (left) and Hindi Search and Playback from Search Result (right)

- "Bouncing ball" to highlight the text in the transcript for a given time segment.
- Ability to click on any word/phrase and play the video from that word.
- Transcript search and playback from the search results.
- Support for multiple transcript languages.
- Placeholder for multiple audio tracks.

Figure 3 shows the video player with Professor Bish Sanyal. The video includes a 99% accurate hand-transcribed English transcript of the original audio and a hand-translated transcript in Hindi. The player allows seamless switching between transcript languages. The left image shows the "bouncing ball" highlight on the English transcript. The right image shows search in the Hindi transcript along with selecting the search result for playback ("seek to result at 1:37"). (Muramatsu, et. al, 2010) The player has provision for multiple audio tracks, but a Hindi audio track to match the Hindi transcript was not included due to time constraints.

3 Future Work

The IIHS proof-of-concept is just one of many steps to developing SpokenMedia as a viable toolset and services for the open education community. SpokenMedia considers its work in two general areas: cost effective creation of lecture transcripts and innovative uses of video linked with transcripts. Ultimately, time-coded transcripts are an enabler for innovative tools SpokenMedia might develop to improve the learning experience with open education video.

3.1 Lecture Transcription Service

Previously at the OpenCourseWare Consortium Global 2009 conference, SpokenMedia identified a production service to automatically generate transcripts for OCW/OER videos. While this is still a possibility, the initial accuracy of the speech recognition plays a critical factor in whether this service might be successful.

At the current accuracy levels, the automatic transcription process may not provide a useful transcript. The 40-60% accuracy, "out-of- the-box" is not sufficient for search and retrieval using the text transcript. At these low accuracy levels, there is not sufficient likelihood of even including the "unique" words (key terms) for which that learners would likely search.

Referring to Figure 1, SpokenMedia only used a portion of the toolset developed by Spoken Language Systems research. SpokenMedia continues to work with the software to enable more of the tools and techniques developed over the courses of the last twenty years in the research lab. For example, use of an acoustic model tuned for Indian-English is expected to improve the results. However, this points out a challenge in the Indian context, there may not be a single acoustic model that accounts for the richness of dialect and backgrounds of speakers (coming from the 23 "official" languages and countless local dialects). Nevertheless a generic male and female Indian-English speaker model should improve the results. From prior research, Jim Glass has shown that having 10 hours of video/audio from a single speaker can be sufficient to develop a custom acoustic model. Or, using a single 99% accurate transcript to "train" the recognizer software can also significantly increase the accuracy. Lastly, it may prove useful to develop a specialized domain model (list of terms) expected to appear in the transcripts of IIHS videos.

3.2 Editing Tools

Despite the low initial accuracy, SpokenMedia could have improved the proof-of-concept project by having better tools to edit the automatically generated transcript. When IIHS staff edited the transcripts, they found it easier to start from scratch and do a manual time code alignment—a very labor-intensive process.

SpokenMedia is in the process of developing an editing tool that works for high accuracy and low accuracy situations. In the high accuracy case, the editor clicks on a word to change only the single word/phrase. In the low accuracy version, the editor can edit the entire text in a text editor like environment. Key to the low accuracy version is being able to use the edited transcript to either train the recognizer software or to be automatically aligned with the time code after the editing is completed.

3.3 Innovative Players

SpokenMedia is actively working with developers at the Innovation Conception et Accompagnement pour la Pedagogie (ICAP) at the Université de Lyon 1 to develop innovative video players that build upon lectures with accompanying transcripts. Through technical exchanges, the teams have created requirements and mockups for a number of interfaces that will enable both learners and educators to:

- Add comments/annotations to videos.
- Create chapters and playlists of videos.
- Embed videos in a rich media notebook.
- Interact with paper-based video "prints" for low bandwidth applications.

Acknowledgements

The authors would like to acknowledge the work of the Jim Glass and the Spoken Language Systems group at MIT's Computer Science and Artificial Intelligence Laboratory. The authors also thank Aromar Revi, IIHS Executive Director, and his team at the Indian Institute for Human Settlements. Finally the authors acknowledge the support of the iCampus MIT/Microsoft Alliance for educational technology for the development of the Spoken Lecture project and providing on-going support for dissemination and diffusion through the SpokenMedia project.

References

IIHS. (2010). Indian Institute for Human Settlements: Curriculum Framework Version 3.0.

Muramatsu, B., McKinney, A., Wilkins, P. (2010). Enabling the IIHS Vision Part 1. Presented at IIHS Curriculum Conference January 2010, Bangalore, India. Retrieved on April 5, 2010 from Slideshare Website: http://www.slideshare.net/bmuramatsu/iihs-open-frameworkspoken-media