

## Enabling the IIHS Vision, Part 1

Brandon Muramatsu

Andrew McKinney

Peter Wilkins—Our colleague at MIT at 0° C

January 2010

## 2 Demos For January 2010

### ■ SpokenMedia

- Video/audio transcription, enabling translation
- Process and tools
- "Access to high-quality learning must be open to all"

### ■ Open IHS Experience

- Course/activity design; student interaction
- "Make curriculum openly available"

*“The IHS Website is our  
commitment to a different  
way of looking at things.”*

*– Aromar Revi  
5 January 2010*

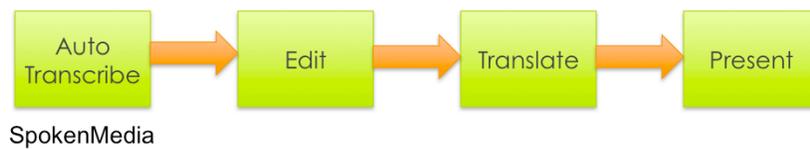
*“The Institution will fail or scale  
based on language.”*

*– Aromar Revi  
5 January 2010*

## Our Goals with this Demo

- **Demonstrate transcripts and translations of IHS videos**
- **Describe the process and our experiences**
  - Transcribe -> Edit -> Translate -> Present

## What did we do?



Four step process (simple)

First we used the SpokenMedia to do automatic transcription.

Next we did hand edit and translation steps.

Then we created a player for the presentation of the video and transcripts...

iihs INDIAN INSTITUTE FOR HUMAN SET

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

Search transcript



00:15 07:36

Prof. Bish Sanyal

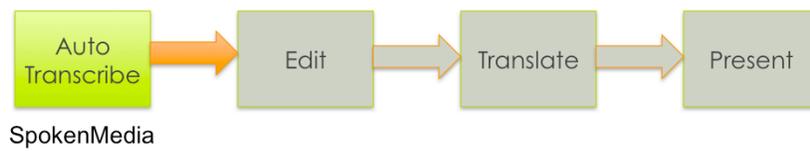
I think one central challenge of planning is legitimacy of government as a key institution that should be involved in bringing the change. The legitimacy of the government is under attack in many different parts of the world so that it has to be reestablished as a major planning institution. I think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

Audio English (US) Transcript English (US)

oeIT

Demo

## How did we do it?



First we used the SpokenMedia to do automatic transcription.

# How do we do it?

## Lecture Transcription

James Glass  
glass@mit.edu



- Spoken Lecture: research project
- Speech recognition & automated transcription of lectures
- Why **lectures**?
  - Conversational, spontaneous, starts/stops
  - Different from broadcast news, other types of speech recognition
  - Specialized vocabularies

### Lecture Transcription

- Jim Glass and his group have years of research experience for spoken languages
- Lectures are a different type of spoken language
  - Much of the speech recognition research has focused on real time transcription of news broadcasts, or interactive voice response systems (telephone)
  - Broadcast news has something like 300 unique words in an hour long broadcast
  - Broadcast news is well structured, prepared copy (in studio via teleprompters), clear transitions between speakers, etc.
  - Lectures are conversational and spontaneous
  - Can use highly specialized vocabularies, engineering, physical sciences, mathematics

# Spoken Lecture Project

James Glass  
glass@mit.edu



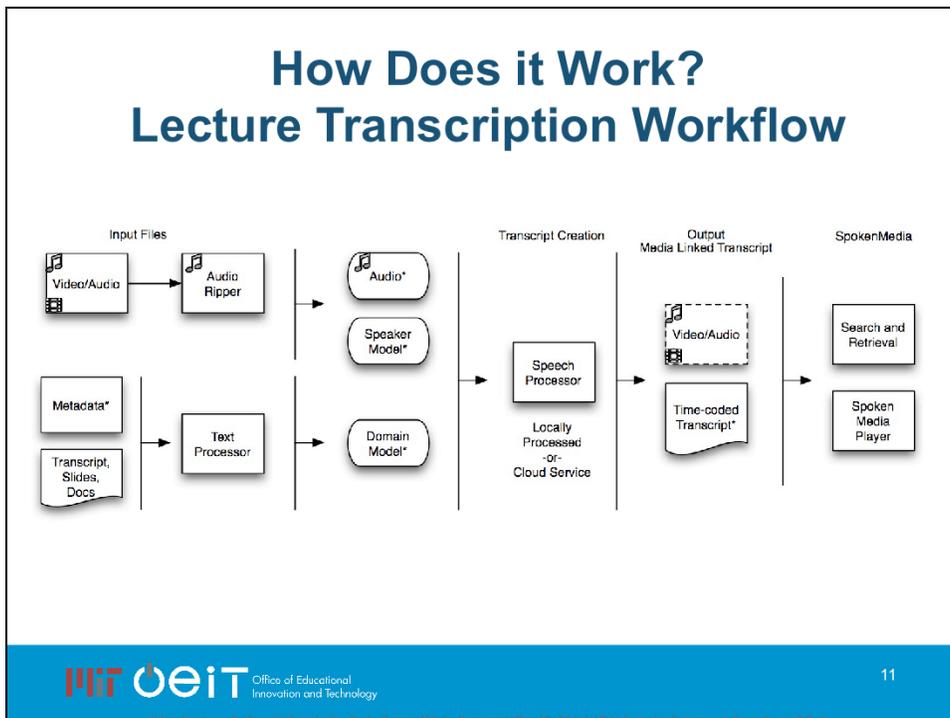
- **Processor, browser, workflow**
- **Prototyped with lecture & seminar video**
  - MIT OCW (~300 hours, lectures)
  - MIT World (~80 hours, seminar speakers)

Supported with iCampus MIT/Microsoft Alliance funding

## Spoken Lecture Project

- Supported by iCampus
- Includes the browser (which was just demo'd) the processor (back end lecture transcription) and a hand workflow to do the processing
- Approximately 400 hours of video indexed

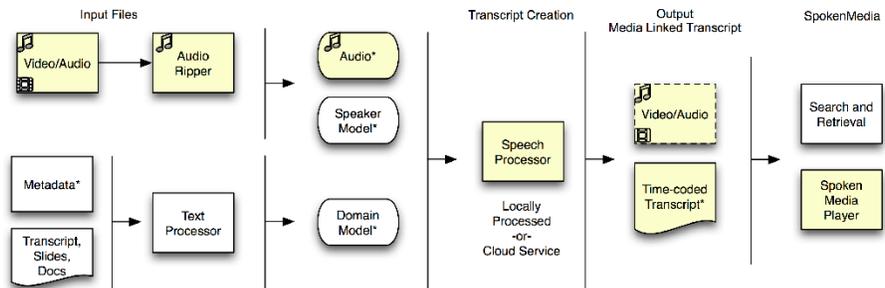
## How Does it Work? Lecture Transcription Workflow



### How does it work?

- Audio
  - System only needs audio (waveform), extracts from video
- Domain Model (base is generic domain model)
  - System needs to know what words it can expect to find in the audio
  - Syllabus, lecture notes, index from text book, research papers
  - Build library of domains
  - Separate sub-process for text for domain model
- Speaker model (base is generic speaker model)
  - If multiple lectures by the same author, best to create a speaker model
  - Separate sub-process for speaker model
- Process—With audio, domain and speaker models
- Output
  - Time coded transcript (standard formats)
  - Links media and transcript
- Applications
  - Search/retrieval
  - Player

# SpokenMedia Process



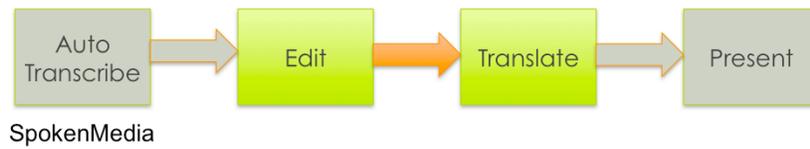
We used a portion of the SpokenMedia process for the demo

12

We only used part of the process due to time constraints.

Audio separation, speech processing, time-coded transcript, and then presentation through a SpokenMedia player.

## How did we do it?



Next we did hand edit and translation steps.

## Edit & Translate: Accuracy

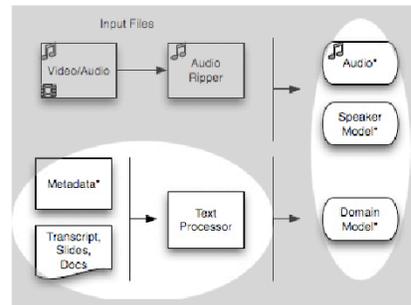
Automatic Transcription	Hand Transcription	Time Adjusted	Translated Hindi
I	I	I	मेरे खयाल से
think	think	think	
once	one	one	नयोजन की एक मुख्य चुनौती है
and	central		
so	challenge	central	
the	of		
challenger	planning	challenge of	
planning	is	planning	
nice	legitimacy	is	
legitimacy	of	legitimacy of	
of	government	government	सरकार की एक ऐसी मुख्य संस्थान के रूप में वैधता
government	as	as	

For this demo, we did computer-based automatic transcription, sent a file to IIHS for “editing” that consisted of performing a hand transcription (due to the format we sent, and the low accuracy of the automatic transcription in this case), a time alignment (though I actually feel that it’s “off” or “slow”) and then a hand translation by IIHS.

## SpokenMedia Accuracy Potential

### Accuracy

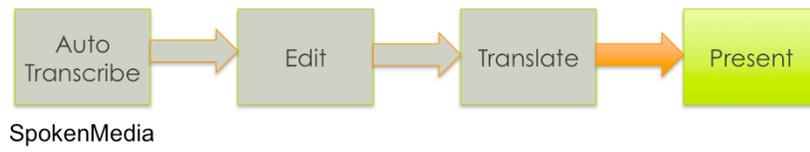
- Domain Model and Speaker Model
  - Internal validity measure
  - Seed with transcript
- Ongoing research by Jim Glass and his team @ MIT



### Recognizer Accuracy

- Base accuracy is approximately 50% (generic domain and speaker models)
- Increase accuracy with speaker model up to 80-85%, and specific domain model
  - This approach is good for courses with multiple lectures by the same speaker
  - Domain models get more useful as more relevant text documents are indexed (keyword/noun phrase extraction)
- Initial results indicate that doing one 99% accurate (by hand/manual) transcript can help immensely for additional lectures by the same speaker
  - Better use of limited resources
- Search accuracy is closer to 90%, searches tend to be for unique words which the processor is better at recognizing

## How did we do it?



Then we created a player for the presentation of the video and transcripts...

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



Prof. Bish Sanyal

Search transcript

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 market is

Audio English (US) Transcript English (US)



## Check it out for yourself

- Demo site:

<http://oki-dev.mit.edu/spokenmedia>

all the videos from IIHS website...it's not just Bish!

## Thank You!

Brandon Muramatsu, [mura@mit.edu](mailto:mura@mit.edu)

Andrew McKinney, [mckinney@mit.edu](mailto:mckinney@mit.edu)