Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Implementing SpokenMedia for the Indian Institute for Human Settlements. Presented at Technology for Education 2010: Mumbai, India, July 1, 2010.

SpokenMedia

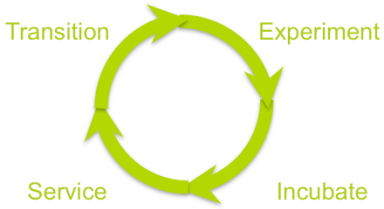
Case Study of Using SpokenMedia for IIHS

- **Goals**
  - Demonstrate value of transcripts + translations for IIHS
  - Test a process to produce transcripts and translations
- **Describe the process and our experiences**
  - Transcribe -> Edit -> Translate -> Present

SpokenMedia

MIT Office for Educational Innovation and Technology

- Dean for Undergraduate Education
- Support Innovation Cycle
- Novel uses of technology to support teaching and learning
- Academic computing without the LMS/VLE



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

SpokenMedia

• Technology transfer—get code running outside of original environment

• 4 components: automatic lecture transcription, search, player, transcript editor

• Suuported by iCampus

SpokenMedia as a Service?

• Still investigating

The Indian Institute for Human Settlements (IIHS) will… "create India's first independent National Innovation University focused on the challenges and opportunities of its urbanisation."

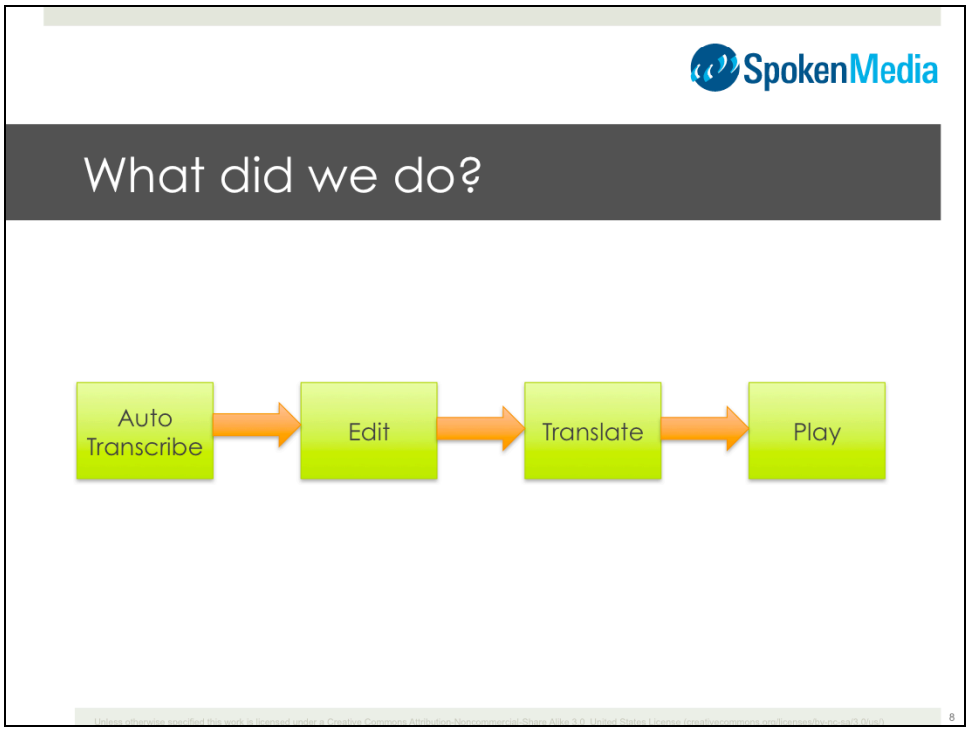
– Indian Institute for Human Settlements: Curriculum Framework Version 3.0
January 2010

"The IIHS 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

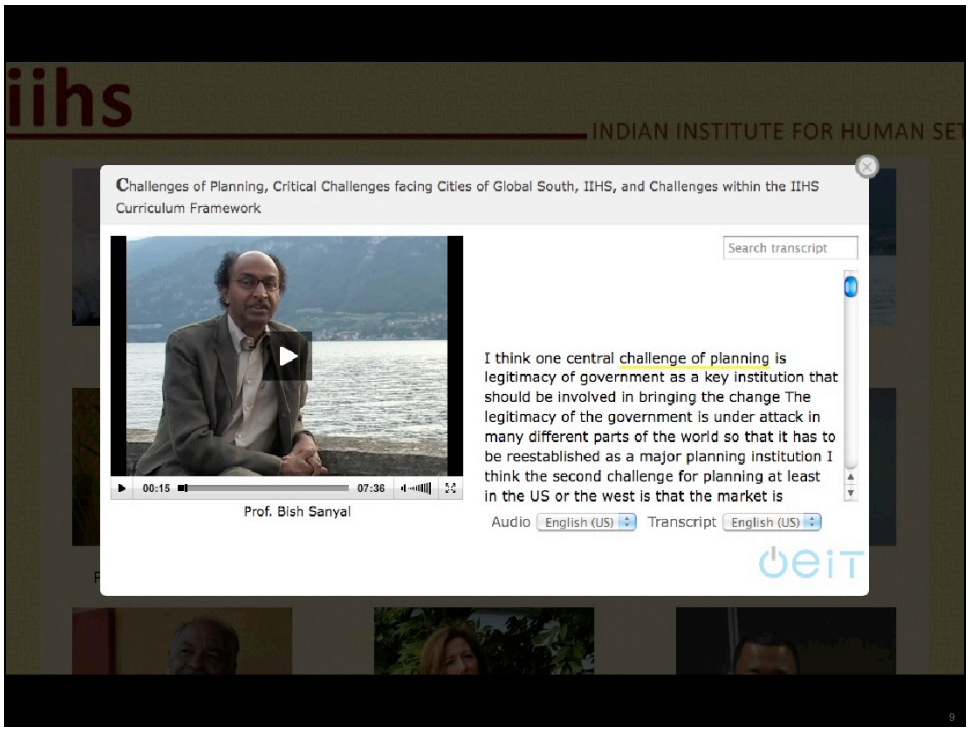


Four step process (simple)

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…



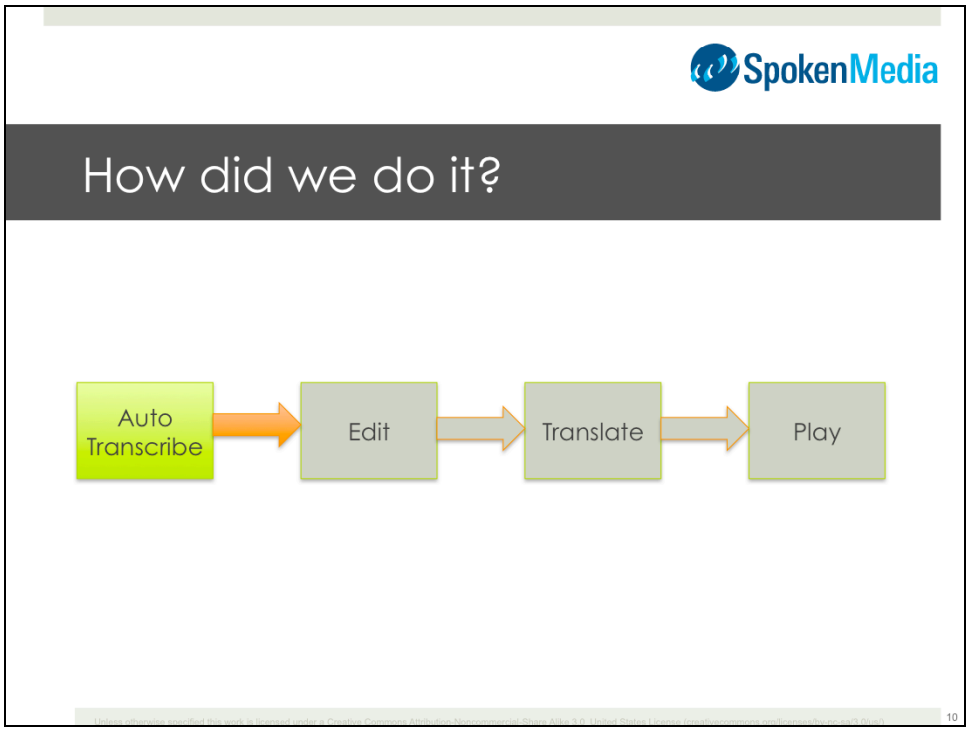
Demo

In this demo, we see a video interview with Prof. Bish Sanyal from MIT. We'll see three things in this demo:

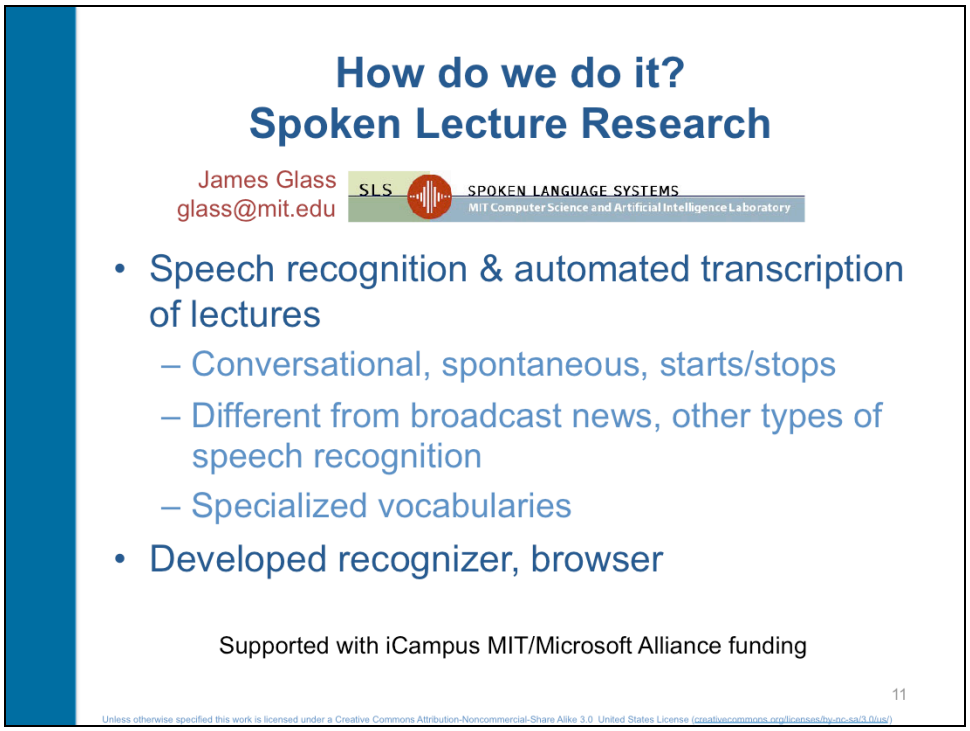
• As Prof. Sanyal speaks, we will see the text in the transcript highlighted and the highlighting will follow along ("bouncing ball")

• Switching the transcript from English to a hand translation into Hindi that is synchronized with the audio, as the switch occurs the playback is seamless

• Searching within the transcript, after entering the search term and selecting the result, the video and transcript seek to that point in the video and playback continues



First we used the SpokenMedia to do automatic transcription.



Lecture Transcription

• Jim Glass and his group have years of research experience for spoken languages

• Lectures are a different type of spoken language

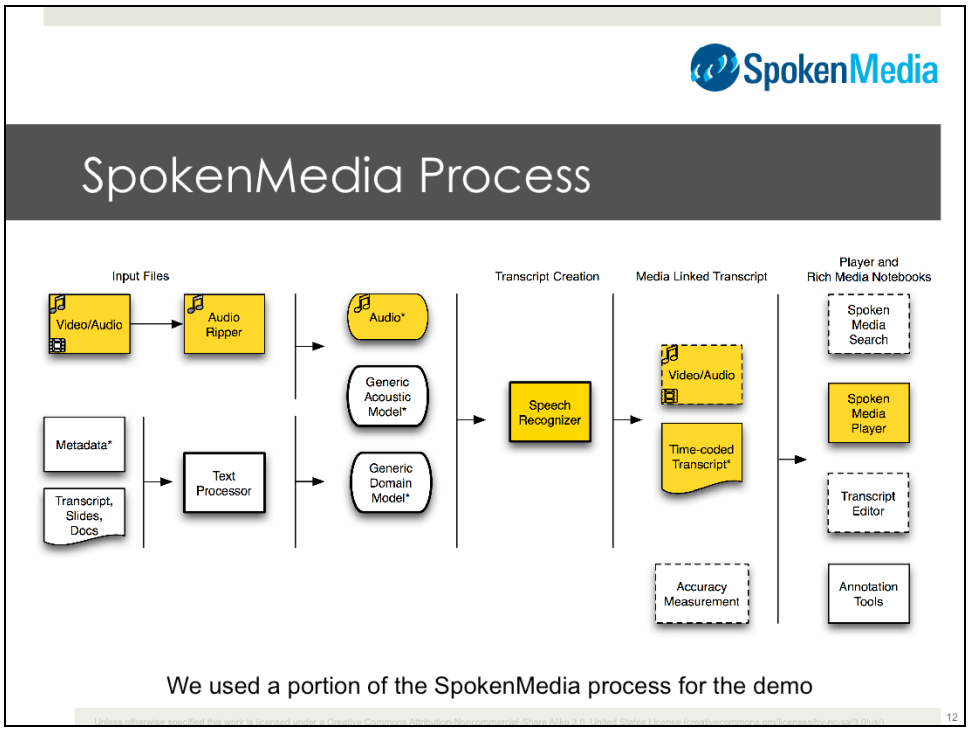
  • Much of the speech recognition research has focused on real time transcription of news broadcasts, or interactive voice response systems (telephone)

  • 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.

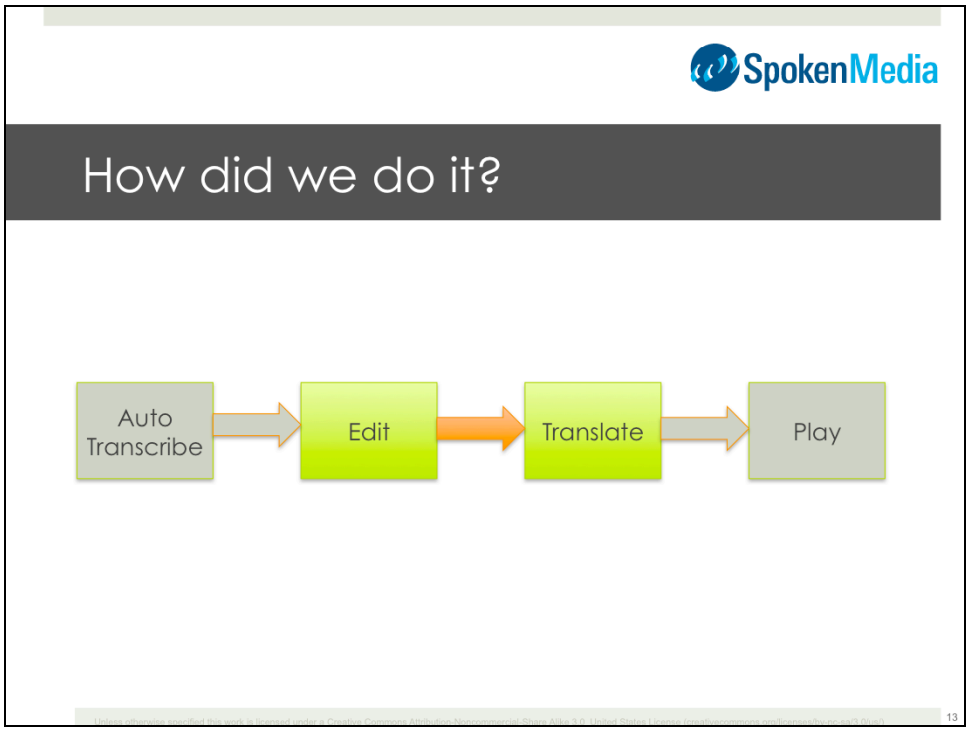
  • Lectures are conversational and spontaneous

  • Can use highly specialized vocabularies, engineering, physical sciences, mathematics

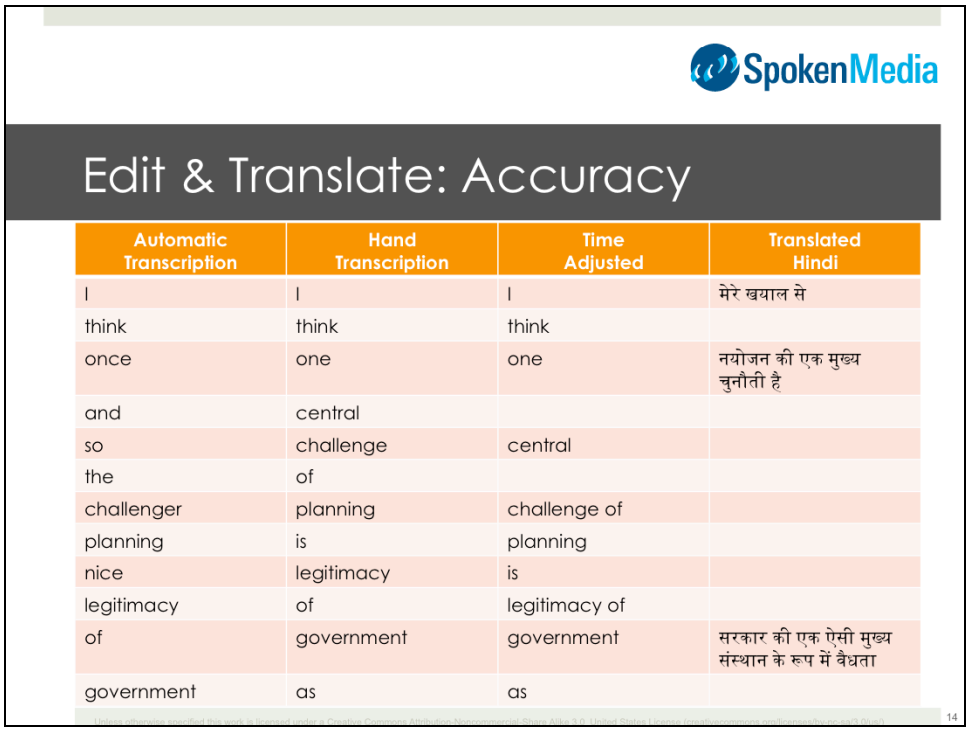


We only used part of the process due to time constraints.

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

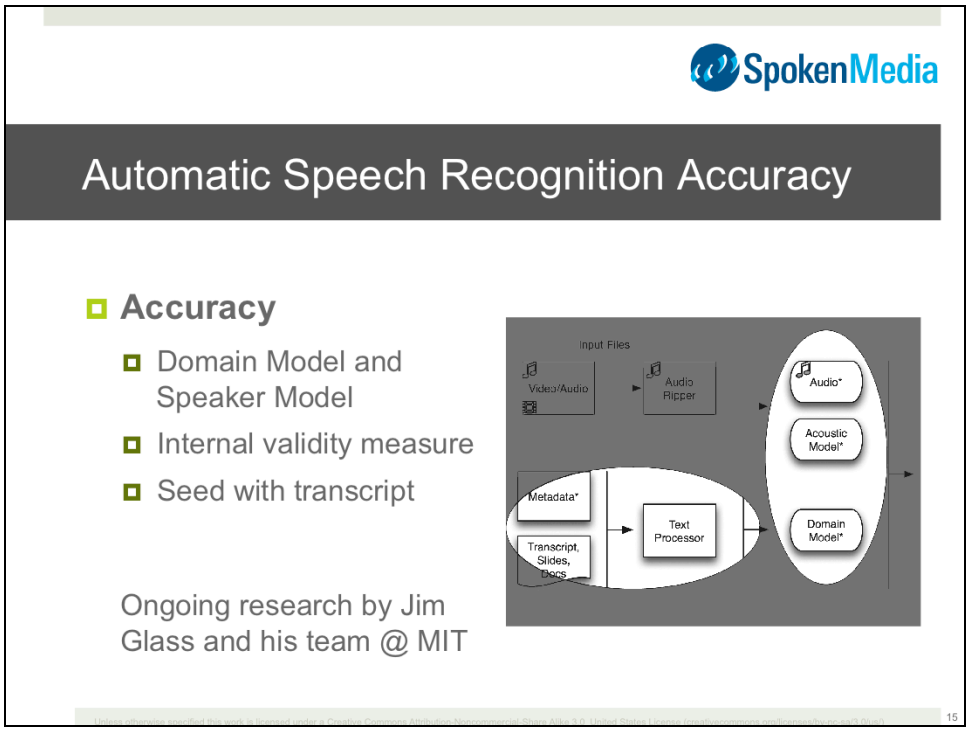


Next we did hand edit and translation steps.



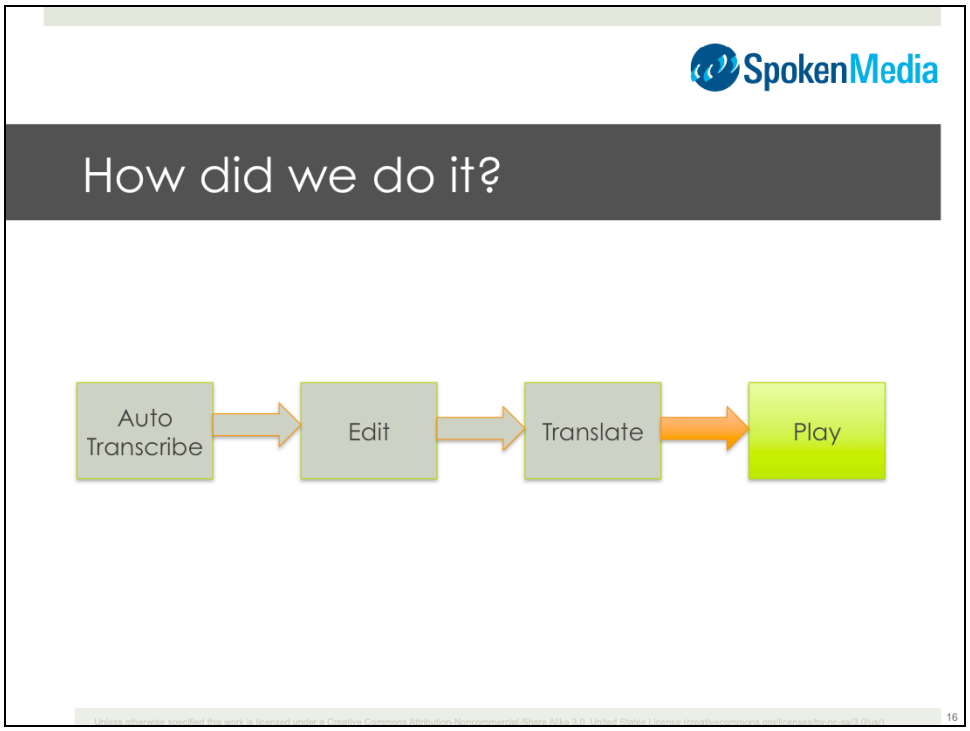
For this demo, we did computer-based automatic transcription, sent a file to IIHS for “editing” that consisted of performing a hand transcription (due to the format we sent, and the low accuracy of the automatic transcription in this case), a time alignment (though I actually feel that it’s “off” or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

Automatic transcription is in the ~50-55% accuracy range (by way of comparison YouTube Auto Caption for this same video is ~68% accuracy).

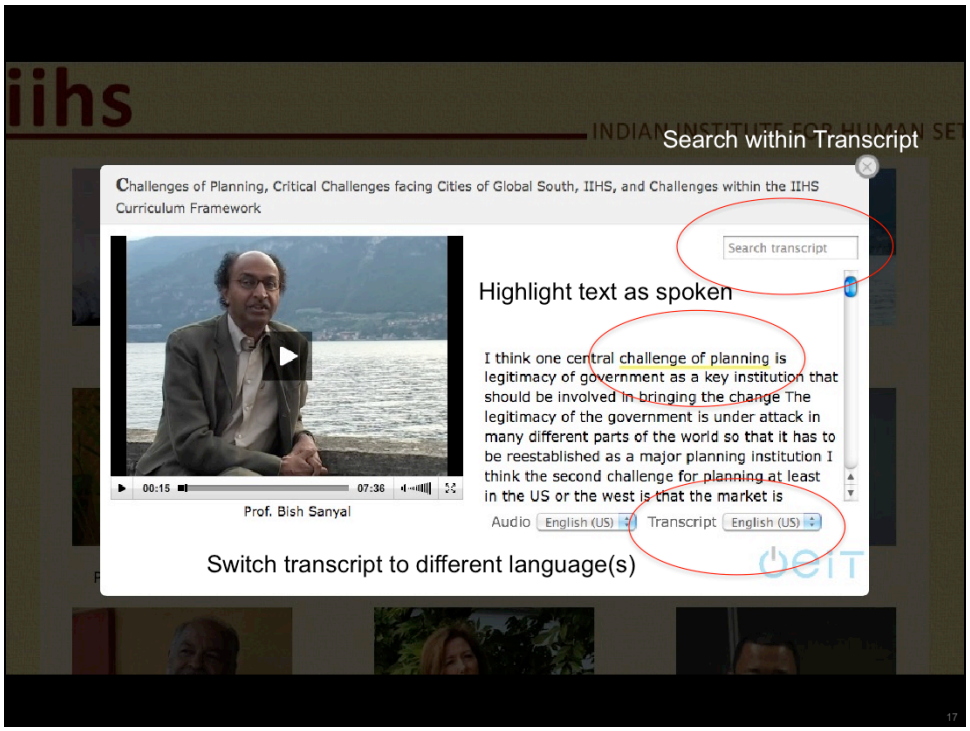


Recognizer Accuracy

- 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

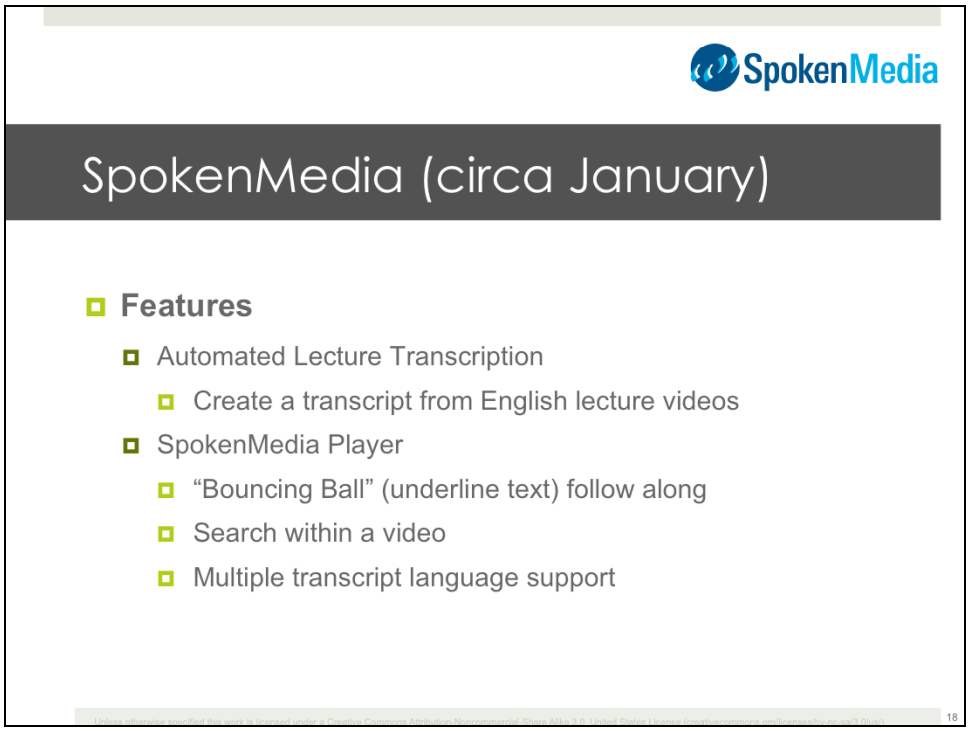


Then we created a player for the presentation of the video and transcripts…



SpokenMedia Player 1.0

- Video-linked transcript
- Highlighted text follows along as the speaker speaks
- Switch transcript to a different transcript track seamlessly during playback
- Search within a transcript



What we have today

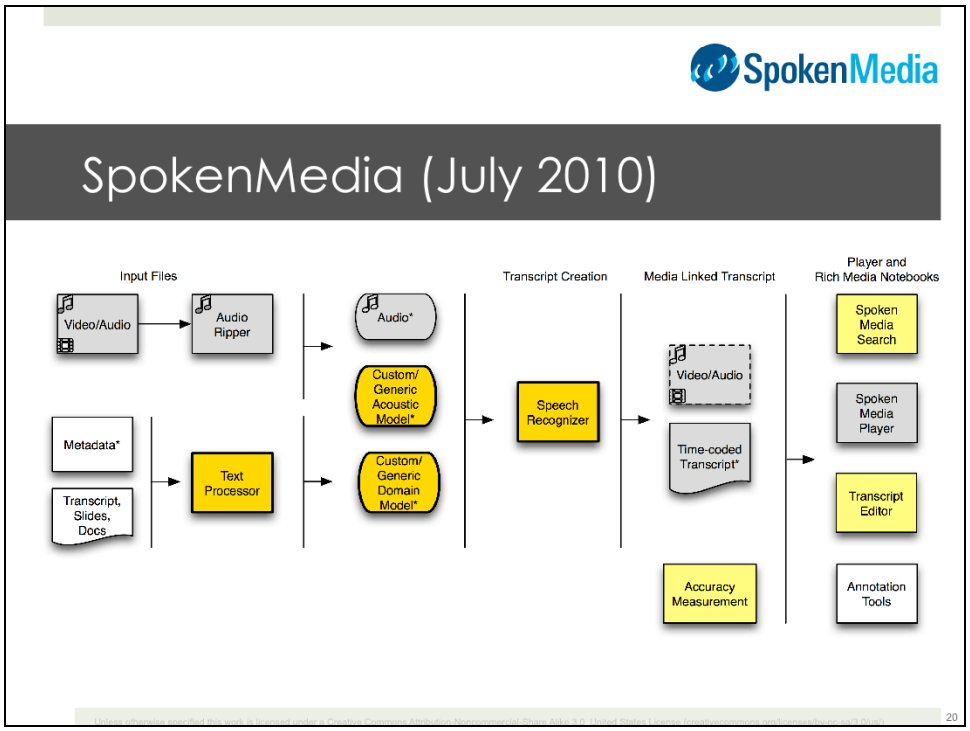
• It’s not perfect, but a pretty good start

• Prototype has a number of useful features that demonstrate search interfaces and interaction interfaces

SpokenMedia

Challenges

- **Accuracy!**
  - We've seen as high as 90% accuracy, but its not going to be 90% for everyone
  - Goal is to do this without special training, and be useful for any lecture video
  - Developing an editing tool to correct transcripts by hand
- **Applicability of speech recognition**
  - Indian context and speech require customization of underlying speaker models and speech recognizer



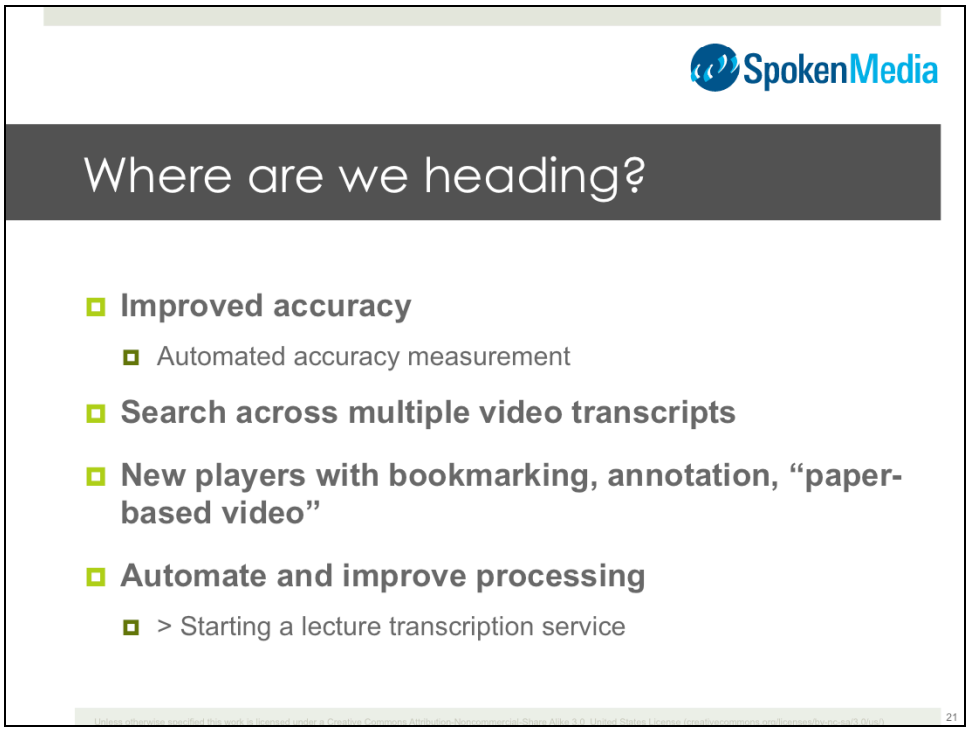
The bright yellow indicates features working in the last two months…

- Text processing to create a custom domain model
- Creation of custom acoustic models in unsupervised mode
- Updated speech recognition software

The gray indicates features we've had working since December 2009 and that were used for IIHS

The light yellow indicates features on which we've just started working.

- Accuracy measurement
- SpokenMedia Search (search across multiple videos)
- Transcript Editor



Where are we heading?

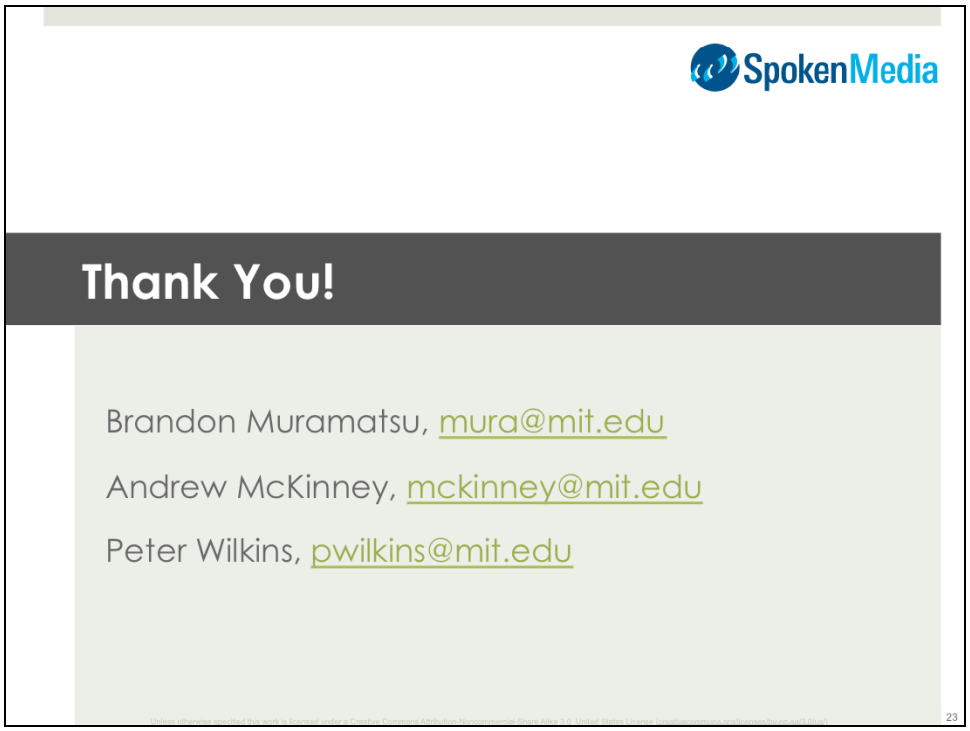
- Transition from research project to service
- Explore new interactions—what we're calling Rich Media Notebooks

SpokenMedia

Check it out for yourself

- **IIHS Demo:**
  http://spokenmedia.mit.edu/demo/iihs/
- **SpokenMedia Website:**

  http://spokenmedia.mit.edu/



Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Implementing SpokenMedia for the Indian Institute for Human Settlements. Presented at Technology for Education 2010: Mumbai, India, July 1, 2010.