Citation: Muramatsu, B., McKinney, A., Long, P. D., & Zornig, J. (2009). The SpokenMedia Project: Toward Rich Media Notebooks for Teaching and Learning.

Presented to the IEEE Computer Society: Bangalore, India, August 6, 2009.

Why are we doing this?

MIT OCW 8.01: Professor Lewin puts his life on the line in Lecture 11 by demonstrating his faith in the Conservation of Mechanical Energy.

- More & more videos on the Web
  - Universities recording course **lectures**
  - **Students** (and universities) **relying** upon Web video for **learning**

MIT OEIT Office of Educational Innovation and Technology

2

Why are we doing this?

• In the last few years, we've seen an explosion of videos on the web.

• Self publishing by millions on YouTube.

• Universities recording course lectures and putting them on the web.

• Students are relying upon web video for learning. Common statistic mentioned by folks like UC Berkeley (which has been doing course webcasts since 1999) is that usage spikes as students prepare for tests, and that they tend to focus on small segments of the video

• Also, cultural organizations (museums, foundations, non-profit organizations) sharing their interviews on the web. Other similar single speaker web video, cost of technology has come down.

<u>What are the challenges?</u>

Large volume of material to search through!

Search results—approximately 3 Million in Google (April 2009):

- Wikipedia, Angular and Conservation of Angular Momentum links might be useful
- Quantum mechanics link is probably too advanced
- Angular Momentum (company) probably not useful
- But no videos

Oh, there's a way of just doing a video search at Google, search is segmented by media type

Google Video Search results—only 400 (April 2009), that's better:

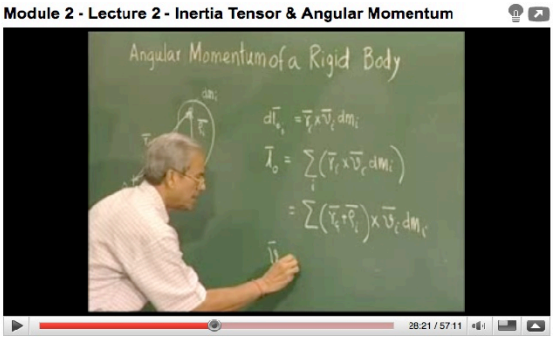
- All appear to be relevant
- Two are lecture length (i.e. 20+ minutes or longer): Mechanical Universe, and Lecture 21 from MIT OCW
- Four are probably demos relating angular momentum to physical examples (tennis, ice skating)

Search results are based on:

- Metadata
- Title of video/link
- Text description of video (typically short), or the text surrounding an embedded video

What are the challenges?

- Interaction & Use
  - Full video vs. Segments
  - Does transcript or captioning exist?



Ghosh, A. (2008). Module 2–Lecture 2–Inertia Tensor & Angular Momentum. Retrieved August 1, 2009 from YouTube Website: http://www.youtube.com/watch?v=a9n2Ztp1Oic

MIT OEIT Office of Educational Innovation and Technology

4

<u>Additional Challenges</u>

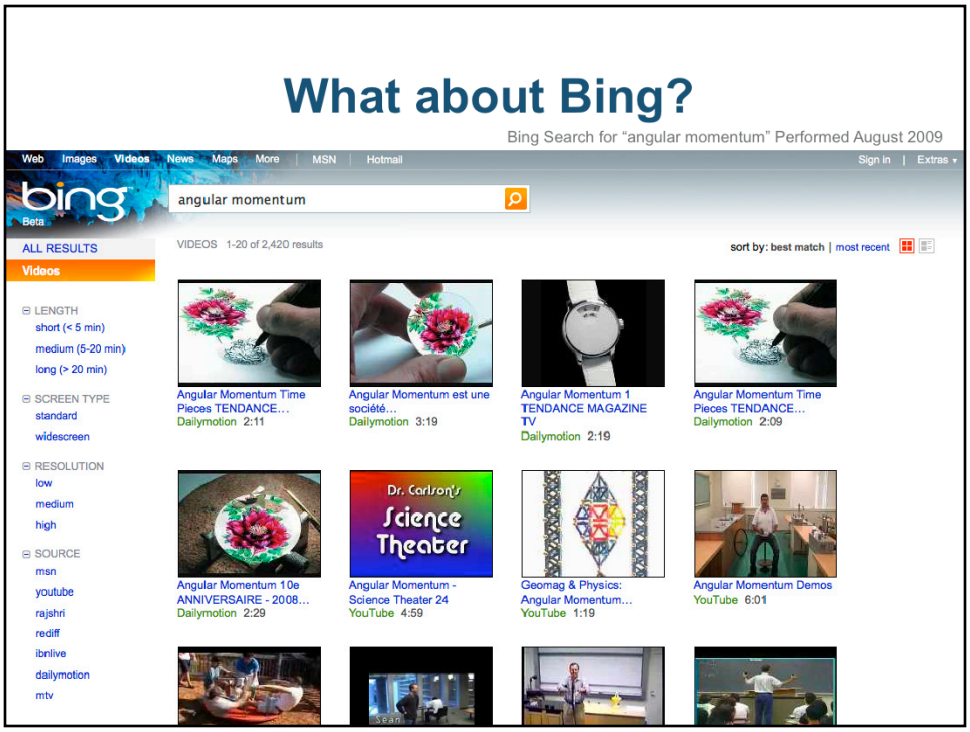
Interaction and Use

• Get the full length video, over 50 minutes

• There may or may not be a transcript, which may or may not be displayed as captioning for accessibility

Policy Implications

• Technology allows for bookmarking and comments, they aren't enabled



What about Bing?

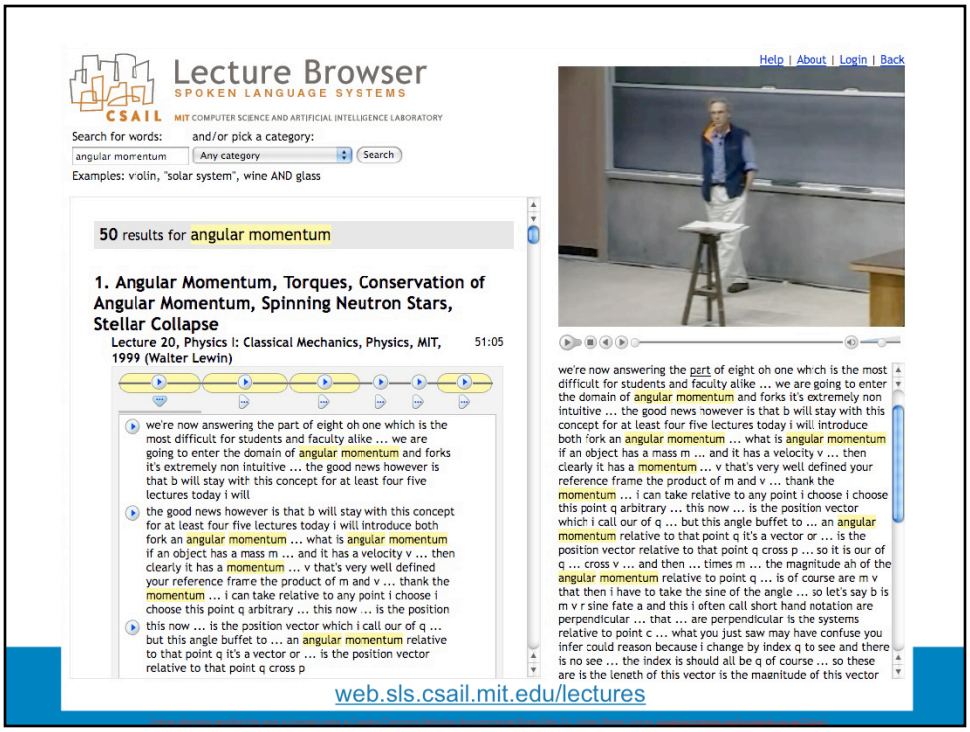
- Fewer Web search results, only 1 Million (August 2009)
    - Three of top six are for companies (two for watchmaker, one for other)
- Still segmented searching (web, video)
- Much less Video search results, only 2,400 (August 2009)
- Video search results much less relevant,
    - First five are for watches,
    - Next three are educational,
    - Does not include Mechanical Universe or MIT OCW videos in first 20 results,
    - NPTEL video is result 19

Why do we want these tools?

- MIT as the customer
- Lots of materials, 1900+ OCW courses, some with video/audio
- Opportunities for positive change: improving presentation and user experience, advocate for new methods of interaction

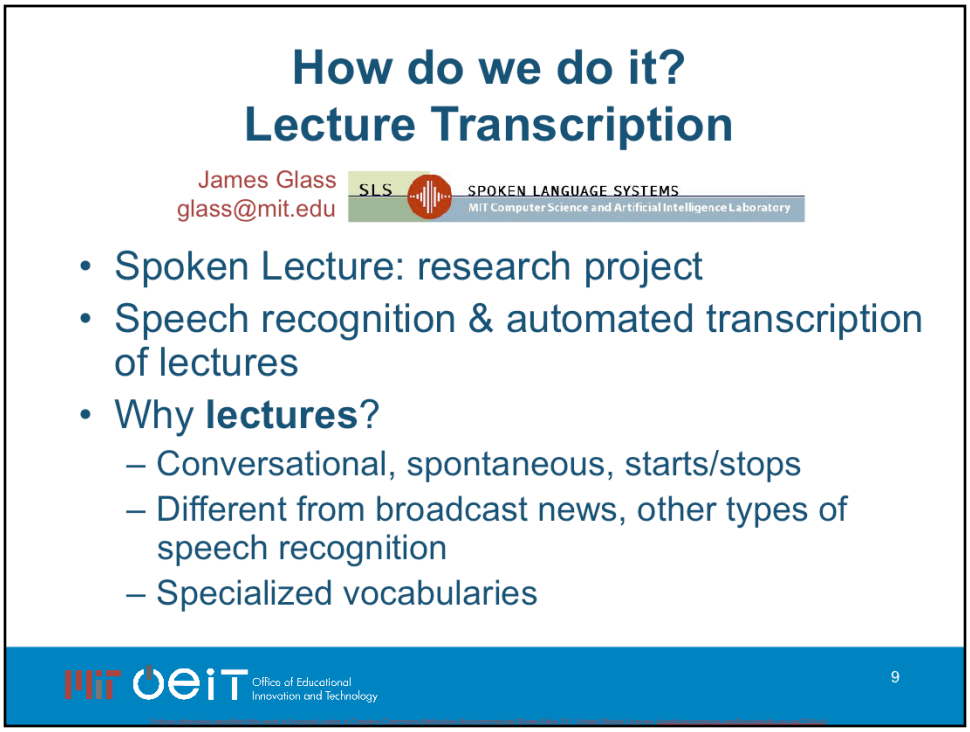
What can we do today?

Demo of Spoken Lecture Browser



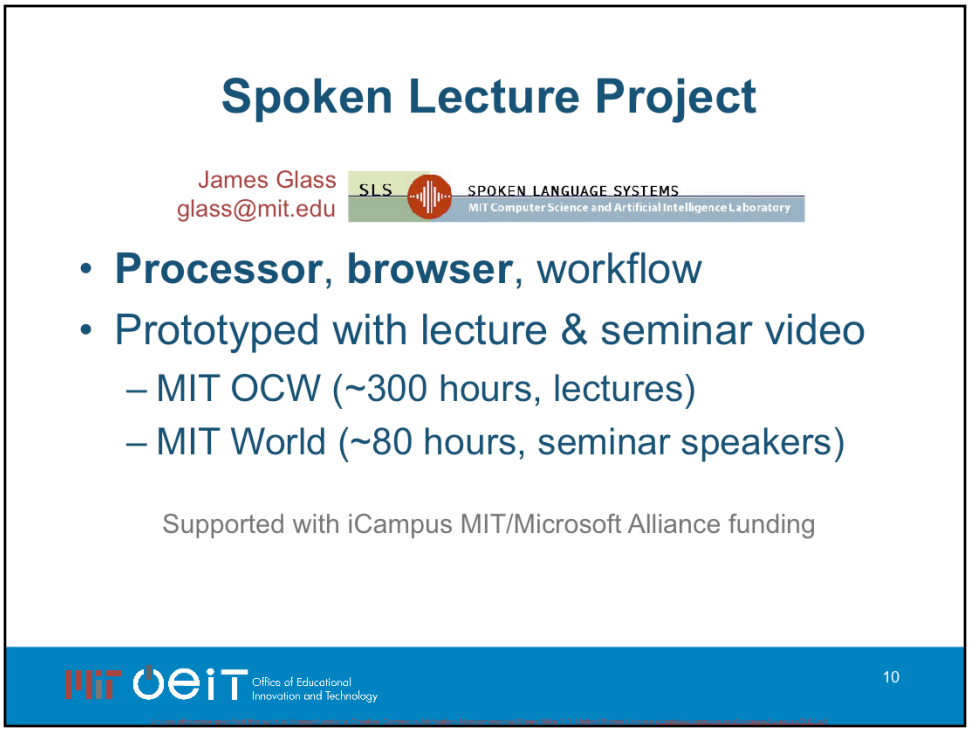
Spoken Lecture Browser Demo

- Search of video library
- Search results show search query in context, query highlighted in yellow
- Results show “conceptual chunks”
- Player controls video and bouncing ball in the transcript
- Uses 2006 technology, Real Player – needs to get updated for current technlogy
    - Much of the work for the Spoken Lecture browser and player completed in 2006 (though recognizer research continues)
    - In 2006, Real could allow within movie bookmarks and playback
    - In 2006, Real was the preferred high-quality video playback mechanism on the Web (lots has changed since then)



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

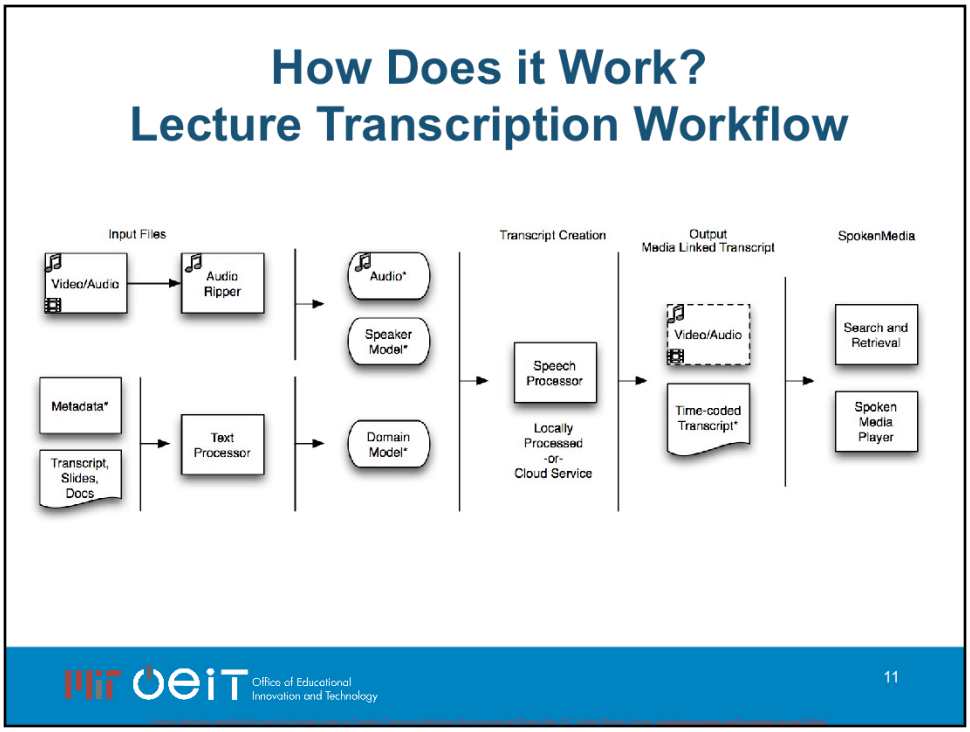


Spoken Lecture Project

• 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

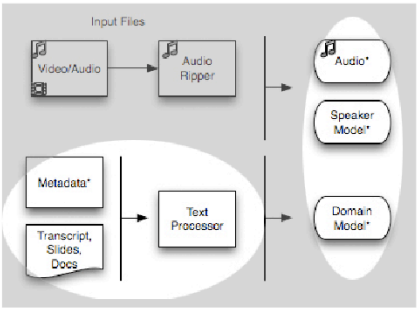


How does it work?

- 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

Recognizer Accuracy ~85%

- Accuracy
  - Domain Model and Speaker Model
  - Internal validity measure
  - Transcripts



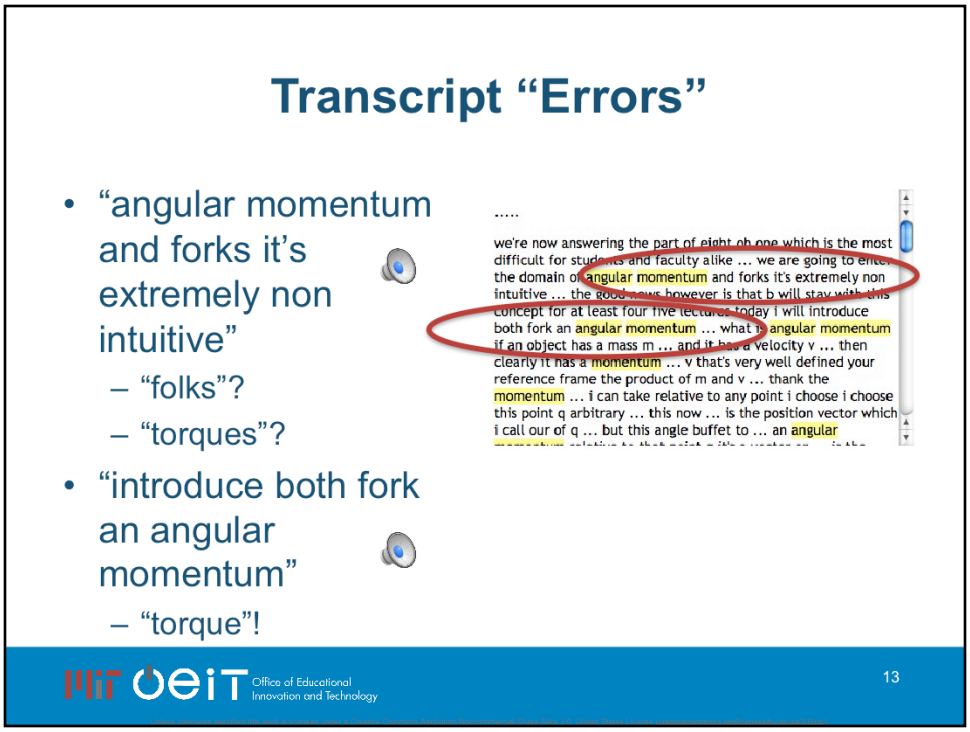
- Ongoing research by Jim Glass and his team

MIT OEIT Office of Educational Innovation and Technology

12

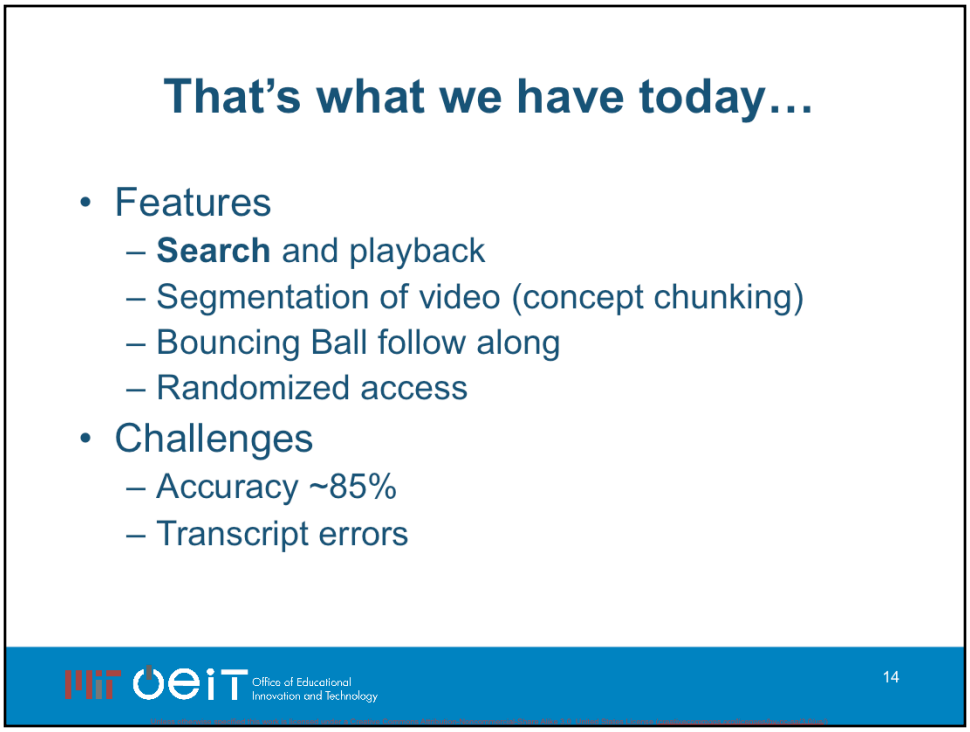
Recognizer Accuracy

- 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



Transcript "Errors"

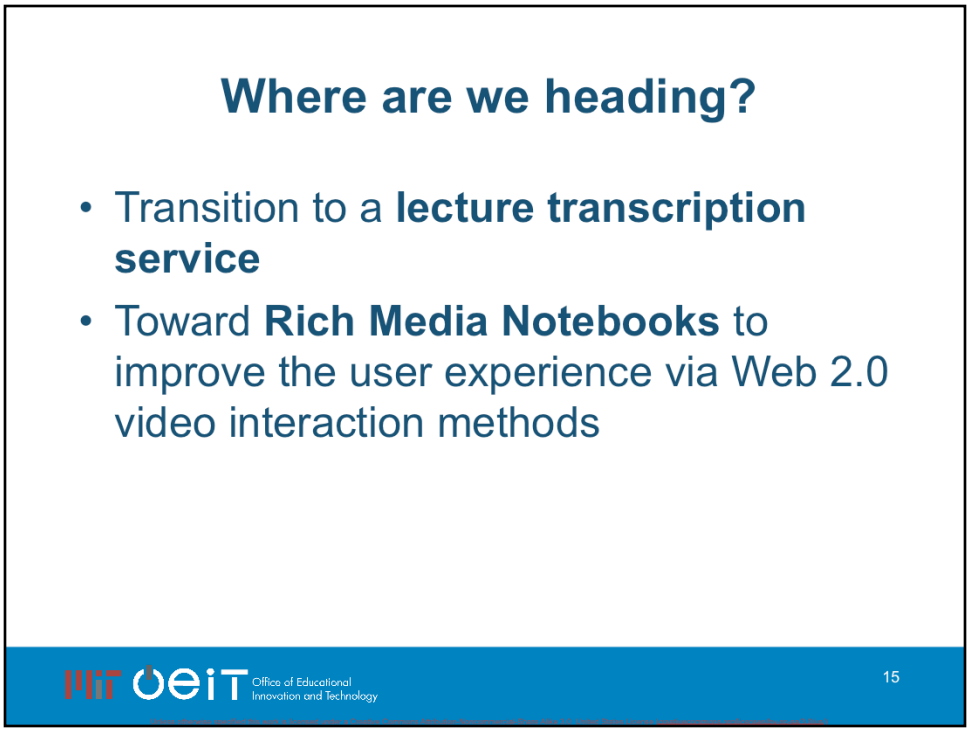
- Recall, processor has 85% accuracy
- Here are two examples of recognizer errors…
  - In the first case, looking at the transcript, it's hard to say what the speaker (Lewin) might have said
  - Continuing … it's unlikely that he used the word "fork" twice
  - Let's listen…ok. It's torque not fork
- Recognizer can recognize when it's guessing—that's not exposed in a public interface, but could be



What we have today

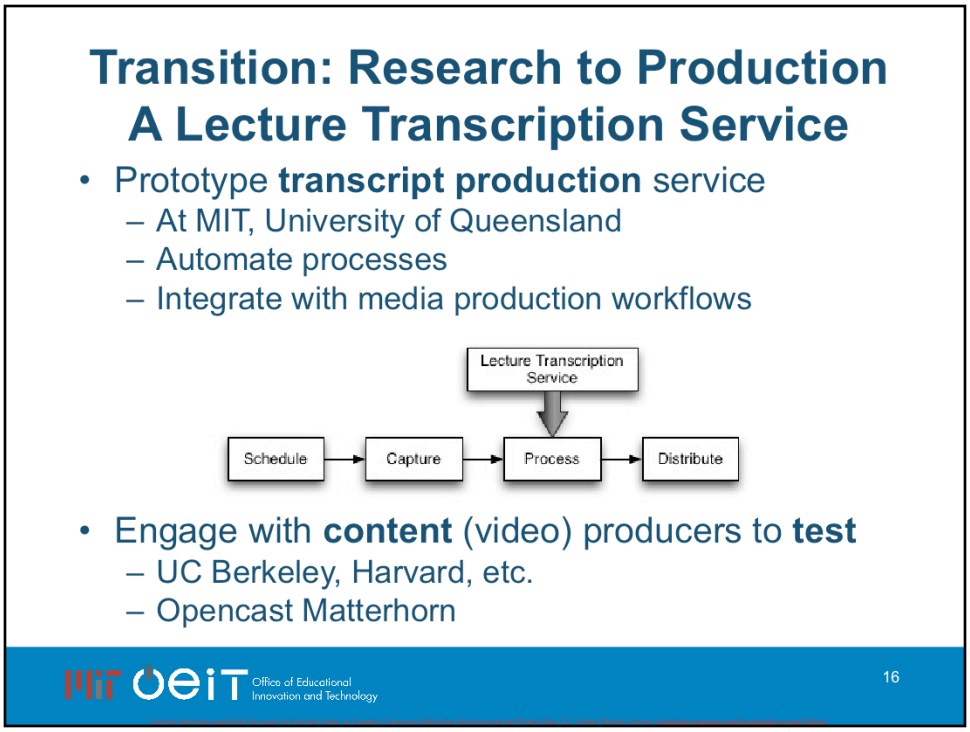
• It's not perfect, but a pretty good start

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



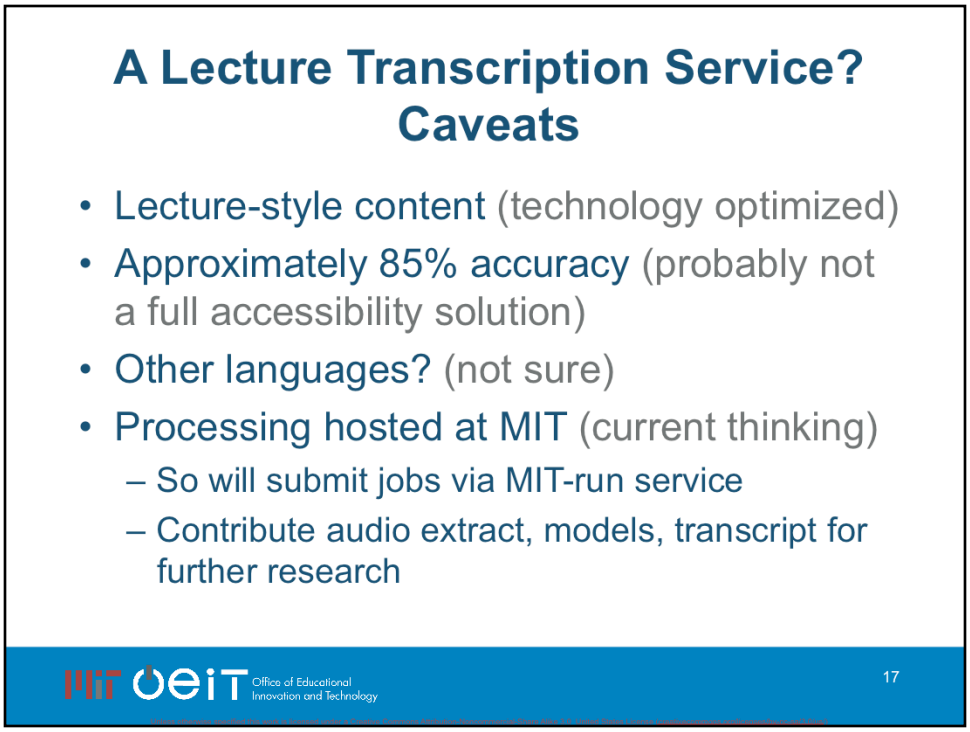
Where are we heading?

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



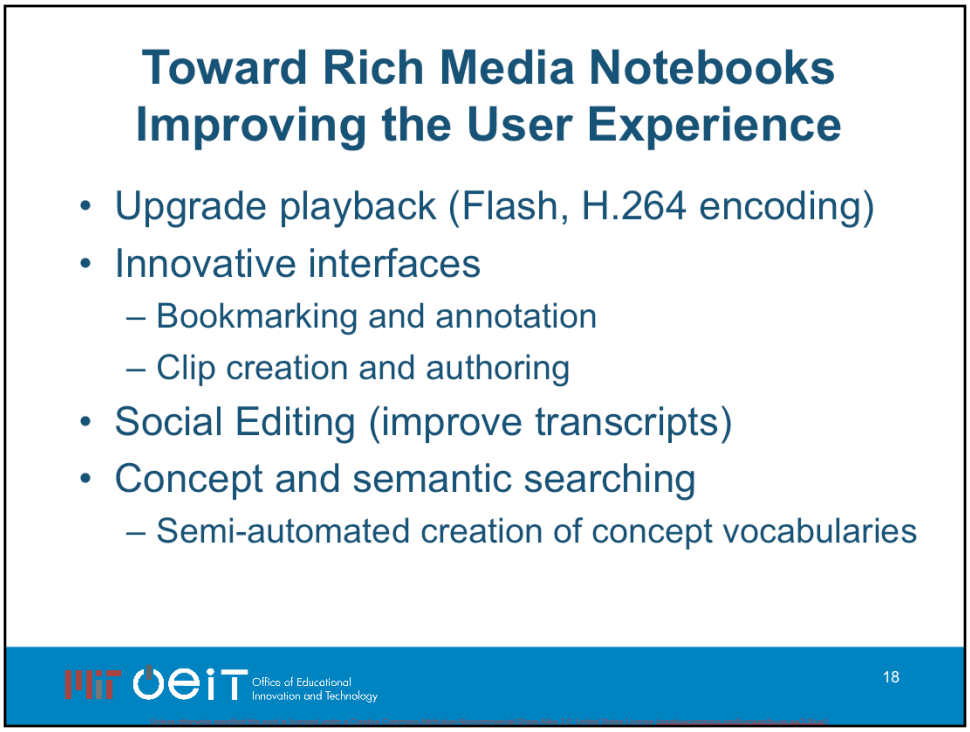
Towards a Lecture Transcription Service

- OEIT at MIT's goal is to transition from research to production
    - First priority to get running on our servers
- Prototype a transcript production service—second priority
    - For MIT
    - Automate a mostly hand process
    - Considering integration with local Podcast Producer workflow engine (Apple)
    - Integrate into media production workflow, as a plugin
- Partner with other content producers to test service—tied for third priority
    - See how it meets needs of other content producers
    - See how it plays with Opencast Matterhorn, distributed service



A Lecture Transcription Service? Caveats

- Full disclosure, limitations we know about or think are important
- We've been asked about other languages
    - Should be possible
    - Jim Glass is experimenting with Chinese
    - Would have to create a language model, not sure what's involved with that
- Current plan to host a web service from MIT
    - Contribution will be important aspect of participation



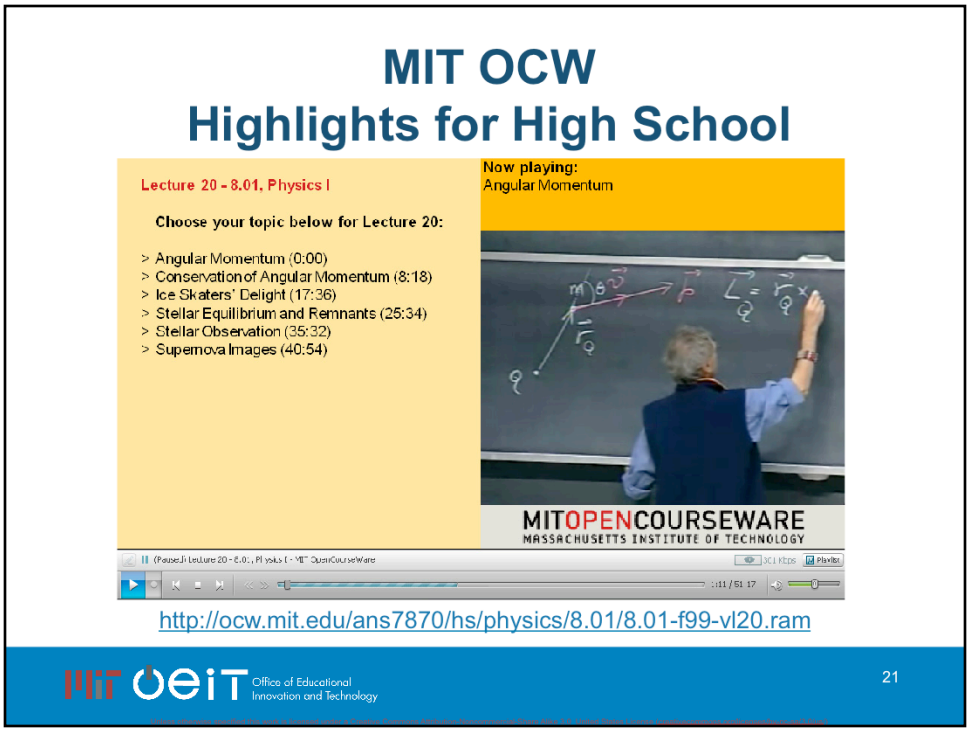
Toward Rich Media Notebooks

- Updating the playback—third priority
    - Flash and H.264 encoded video
- Implement other common video features (e.g., from YouTube and other commercial video sites)
    - Bookmarking, annotations and comments (timestamp, text fields)
    - Clip creation (ala XMAS cross media annotation system)
- Down the road
    - Social editing to improve transcripts, wiki interfaces, trust systems
    - Concept and semantic searching—current system breaks text up into "logical" chunks, and lets users search for a term or phrase, but doesn't really get to concepts.

Alternate Representations

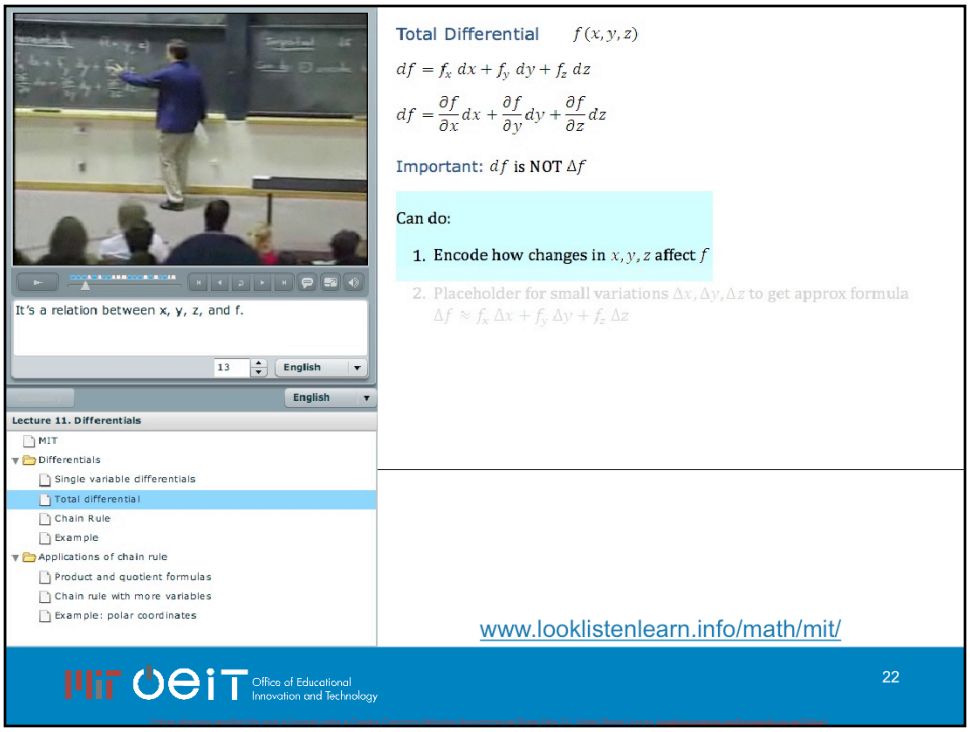
- Microsoft Project Tuva: Enhanced Video Player
  - research.microsoft.com/apps/tools/tuva/
- MIT OCW Highlights for High School
- Look Listen Learn
  - Alternate view of MIT OCW video
  - www.looklistenlearn.info/math/mit/
- Google Audio Indexing
  - labs.google.com/gaudi
  - U.S. political coverage (2008 Elections, CSPAN)


Microsoft Research
PROJECT TUVA
3 results
angular momentum
Richard Feynman: The Messenger Series: The Great Conservation Principles
NOTES | 1
44:31
Here Feynman covers Angular Momentum
Insert Note at: 45:47
0 / 150
The explanation is, that angular momentum appears in two forms:
45:47
56:03
Notes
Credits: Contributors | Stimulant
research.microsoft.com/apps/tuva/
Contact Us | Terms of Use | Trademarks | Privacy Statement
©2009 Microsoft Corporation. All rights reserved.
Microsoft
MIT OEIT Office of Educational Innovation and Technology
20



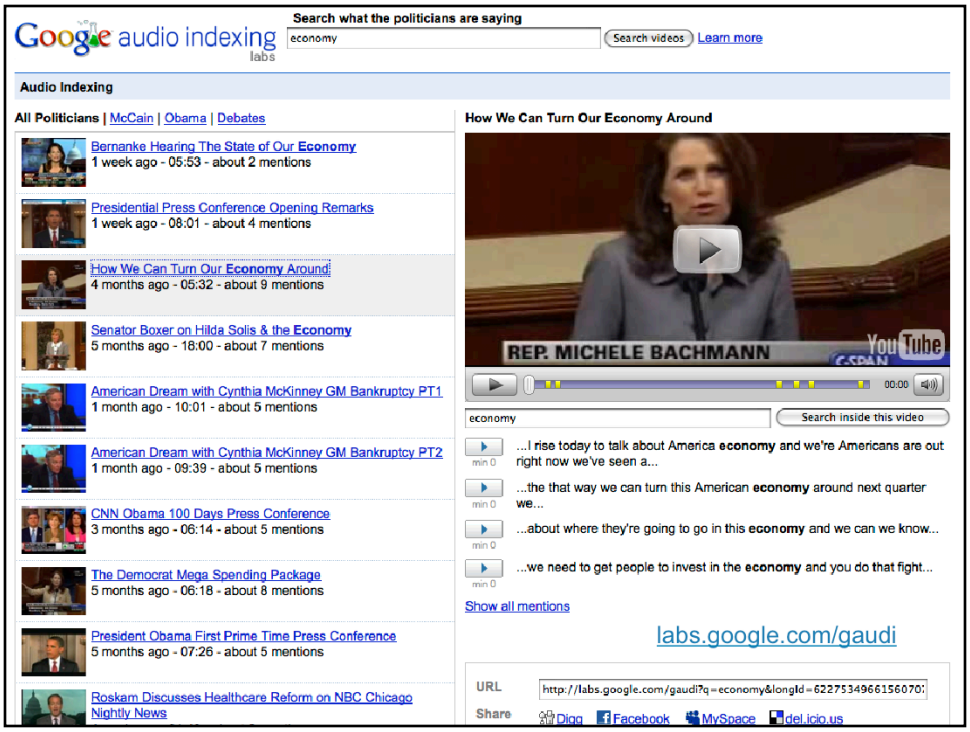
OCW Highlights for High School

- Links MIT OCW course materials to Advanced Placement tests
- Created by undergrads and high school students
- Hand-created link between topics in AP Physics and existing course resources
- Links concepts to specific video segments



Look Listen Learn

- Hand crafted from MIT OCW video
  - Created equations that are written on the board (in LaTeX)
  - Displays equations as they are written
  - Displays transcripts as well
  - Must have taken quite a bit of time to do this
  - Probably a really good interface for the material



Google Audio Indexing

- Google Labs Project
- Indexed CSPAN on economy and 2008 U.S. Presidential elections
- Similar search results and display of segments (though less text for context)
- Takes advantage of playback from within Flash at YouTube (which they haven't made available to the general public as far as I know yet)

Thanks!

oeit.mit.edu/spokenmedia

Brandon Muramatsu
mura@mit.edu
MIT, Office of Educational Innovation and Technology

Andrew McKinney, MIT OEIT
Phillip Long and John Zornig, University of Queensland

Citation: Muramatsu, B., McKinney, A., Long, P. D., & Zornig, J. (2009). The SpokenMedia Project: Toward Rich Media Notebooks for Teaching and Learning. Presented to the IEEE Computer Society: Bangalore, India, August 6, 2009.

MIT OEIT Office of Educational Innovation and Technology