SpokenMedia:
Automatic Lecture Transcription and Rich Media Notebooks

Brandon Muramatsu   mura@mit.edu
Andrew McKinney   mckinney@mit.edu
Peter Wilkins   pwilkins@mit.edu

MIT, Office of Educational Innovation and Technology

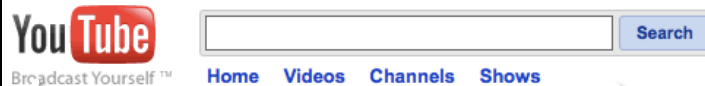SpokenMedia:
Automatic Lecture Transcription
and Rich Media Notebooks

…we now re… …esentation…

YouTube
Broadcast Yourself ™

Search

Home   Videos   Channels   Shows

Broadcasting Ourselves ;)
The Official YouTube Blog

THURSDAY, MARCH 4, 2010

The Future Will Be Captioned: Improving Accessibility on YouTube

Tens of millions of people in the U.S. experience some kind of hearing impairment and recent studies have predicted that over 700 million people worldwide will suffer from hearing impairment by 2015. To address a clear need, the broadcast industry began running caption of regular video programming in the early 1970s. Today, closed captions on television are more typical than not. But closed captions for video on the web have never been broadly available...

Making video easily accessible is something we're working hard to address at YouTube. One of the first steps we took was the development of a caption feature in 2008. In November of last year we released auto-captioning for a small, select group of partners. Auto-captioning combines some of the speech-to-text algorithms found in Google's Voice Search to automatically generate video captions when requested by a viewer. The video owner can also download the auto-generated captions, improve them, and upload the new version. Viewers can even choose an option to translate those captions into any one of 50 different languages -- all in just a couple of clicks.

Today, we are opening up auto-captions to all YouTube users. There will even be a "request processing" button for un-captioned videos that any video owner can click on if they want to speed up the availability of auto-captions. It will take some time to process all the available video, so here are some things to keep in mind:

- While we plan to broaden the feature to include more languages in the months to come, currently, auto-captioning is only for videos where English is spoken.
- Just like any speech recognition application, auto-captions require a clearly spoken audio track. Videos with background noise or a muffled voice can't be auto-captioned. President Obama's speech on the recent Chilean Earthquake is a good example of the kind of audio that works for auto-captions.
- Auto-captions aren't perfect and just like any other transcription, the owner of the video needs to check to make sure they're accurate. In other cases, the audio file may not be good enough to generate auto-captions. But please be patient -- our speech recognition technology gets better every day.
- Auto-captions should be available to everyone who's interested in using them. We're also working to provide auto-captions for all past user uploads that fit the above mentioned requirements. If you're having trouble enabling them for your video, please visit our Help Center: this article is for uploaders and this article is for viewers.

For content owners, the power of auto-captioning is significant. With just a few quick clicks your videos can be accessed by a whole new global audience. And captions can make is easier for users to discover content on YouTube.

MIT, …ology

YouTube. (2010, March 4). The Future Will be Captioned: Improving Accessibility on YouTube. Retrieved on March 8, 2010 from YouTube Website:
http://youtube-global.blogspot.com/2010/03/future-will-be-captioned-improving.html

MIT OEIT

SpokenMedia: What to do if your videos aren't in YouTube

Brandon Muramatsu   mura@mit.edu
Andrew McKinney   mckinney@mit.edu
Peter Wilkins   pwilkins@mit.edu

MIT, Office of Educational Innovation and Technology

MIT OEIT Office of Educational Innovation and Technology

Why are we doing this?

- **More & more videos on the Web**
  - Universities recording course lectures
  - Students relying upon Web video for courses
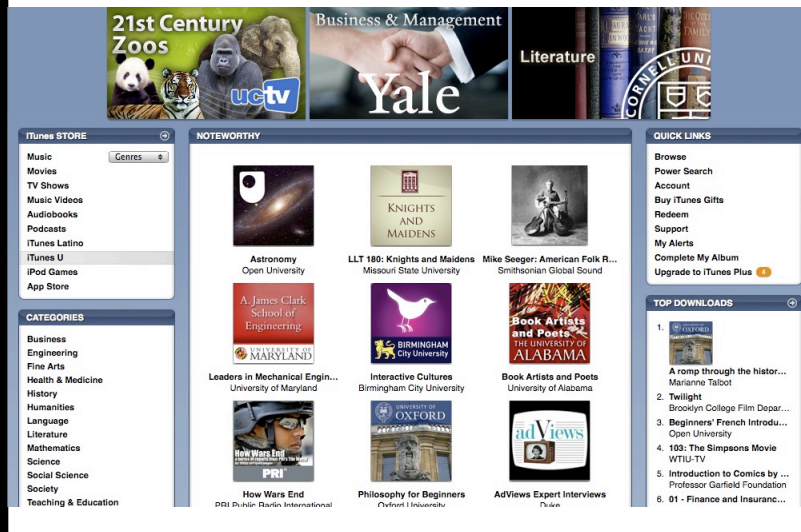
MiT OEiT Office of Educational Innovation and Technology

What video? Where?

What are the challenges?

- **Search**
  - Volume
  - Segmented by Web, Video
  - Text title and Description

Google Search for "angular momentum" Performed April 2009

What about Bing?

Bing Search for "angular momentum" Performed August 2009

What are the Challenges?

- ## Description
  - ### Course and Lecture Title
  - ### Summary
  - ### Metadata?

What are the challenges? Use

- Interaction & Use
  - Transcripts / captions
    - Do they exist?
    - Cost?
  - Full video vs. segments
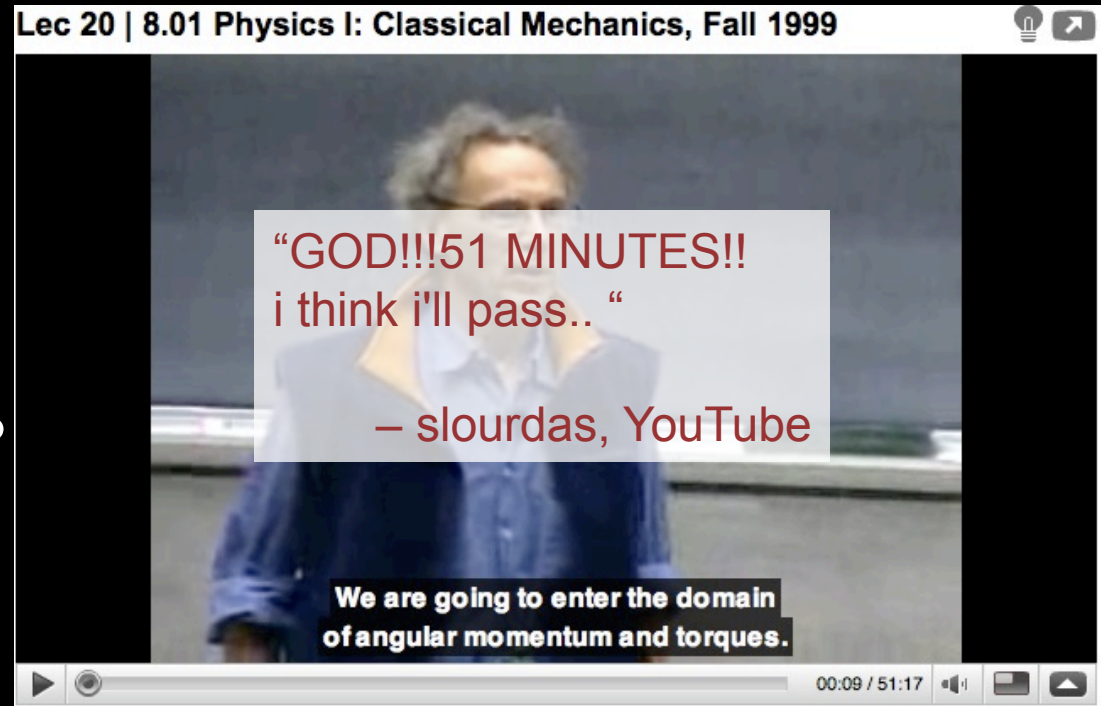
Lewin, W. (1999). Lec 20 | 8.01 Physics I: Classical Mechanics, Fall 1999. Retrieved August 1, 2009 from YouTube Website: http://www.youtube.com/watch?v=ibePFvo22x4

MIT OEIT Office of Educational Innovation and Technology

10
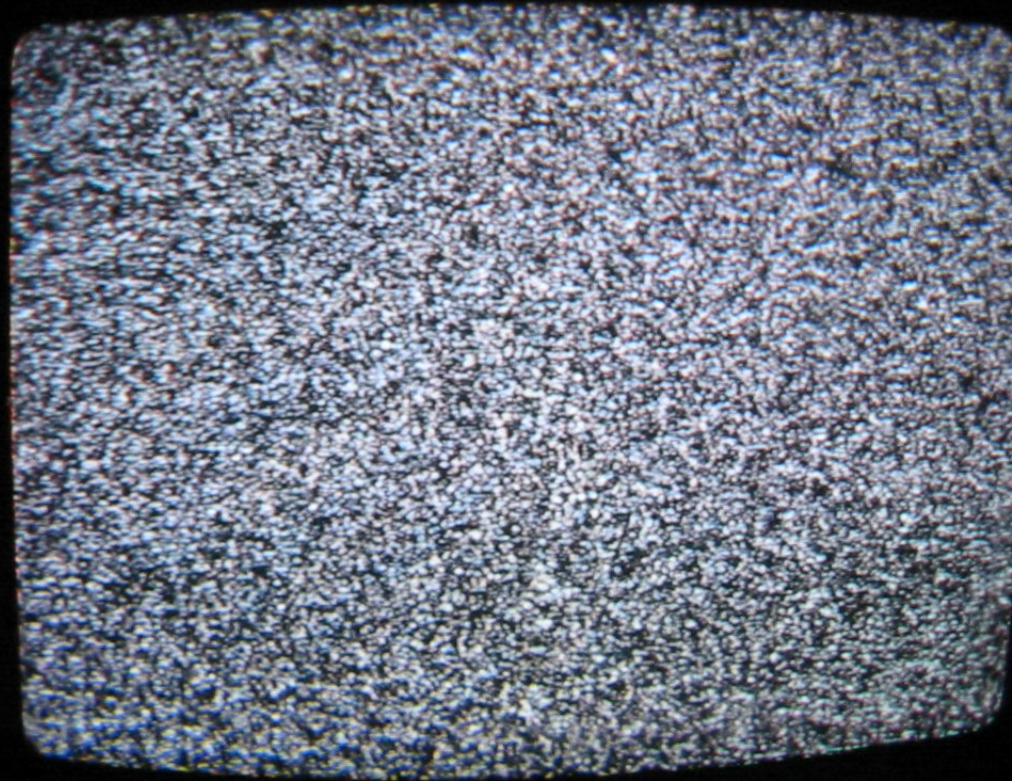
Why do we need these tools?

- Improve search and retrieval
- Improve user experience
- Captioning for accessibility?
- Facilitate translation?
- Other uses?

Broadcasting Ourselves ;)
The Official YouTube Blog

THURSDAY, MARCH 4, 2010

The Future Will Be Captioned: Improving Accessibility on YouTube

Tens of millions of people in the U.S. experience some kind of hearing impairment and recent studies have predicted that over 700 million people worldwide will suffer from hearing impairment by 2015. To address a clear need, the broadcast industry began running captions on regular video programming in the early 1970s. Today, closed captions on video are more prevalent than ever. But generating captions today can be a time-consuming and complicated process.

Making video easily accessible is something we're working hard to address at YouTube. One of the first steps we took was the development of a caption feature in 2008. In November of last year we released auto-captioning for a small, select group of partners. Auto-captioning combines some of the speech-to-text algorithms found in Google's Voice Search to automatically generate video captions when requested by a viewer. The video owner can also download the auto-generated captions, improve them, and upload the new version. Viewers can even choose an option to translate those captions into any one of 50 different languages -- all in just a couple of clicks.

Today, we are opening up auto-captions to all YouTube users. There will even be a "request processing" button for un-captioned videos that any video owner can click on if they want to speed up the availability of auto-captions. It will take some time to process all the available video, so here are some things to keep in mind:

- While we plan to broaden the feature to include more languages in the months to come, currently, auto-captioning is only for videos where English is spoken.
- Just like any speech recognition application, auto-captions require a clearly spoken audio track. Videos with background noise or a muffled voice can't be auto-captioned. President Obama's speech on the recent Chilean Earthquake is a good example of the kind of audio that works for auto-captions.
- Auto-captions aren't perfect and just like any other transcription, the owner of the video needs to check to make sure they're accurate. In other cases, the audio file may not be good enough to generate auto-captions. But please be patient -- our speech recognition technology gets better every day.
- Auto-captions should be available to everyone who's interested in using them. We're also working to provide auto-captions for all past user uploads that fit the above mentioned requirements. If you're having trouble enabling them for your video, please visit our Help Center: this article is for uploaders and this article is for viewers.

For content owners, the power of auto-captioning is significant. With just a few quick clicks your videos can be accessed by a whole new global audience. And captions can make is easier for users to discover content on YouTube.

Comparing SpokenMedia and YouTube Auto-Caption?

YouTube

- Scale ✔
- Research-basis ✔
- For all videos ✔ (soon)
- No transcript/caption export (?)
- YouTube hosted
- Accuracy based on general patterns (?)
- No transcript editing (?)

SpokenMedia

- Limited
- Research-basis ✔
- Service by request
- Transcript/caption export available ✔
- Hosted anywhere ✔
- Accuracy based on custom models ✔ (soon)
- Transcript editing ✔ (soon)

MIT OEIT Office of Educational Innovation and Technology

Developing SpokenMedia...

- What do we have at MIT?
  - Existing videos & audio, new video
  - Lecture notes, slides, etc. (descriptive text)
  - Multiple videos/audio by same lecturer
  - Diverse topics/disciplines

- Research from Spoken Language Systems Group !!!

MIT OEIT Office of Educational Innovation and Technology

14

Enabling Research

James Glass
glass@mit.edu

SLS   SPOKEN LANGUAGE SYSTEMS
MIT Computer Science and Artificial Intelligence Laboratory

- 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

Spoken Lecture Project

James Glass
glass@mit.edu

SLS — SPOKEN LANGUAGE SYSTEMS
MIT Computer Science and Artificial Intelligence Laboratory

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

MIT OEIT Office of Educational Innovation and Technology

Tech Transfer Timeline:
Research -> Service

SpokenLecture Browser

Tech Transfer Begins
SpokenMedia Player !

1990   2000   2006   2009   2010

Spoken Language Systems Group
Research

17
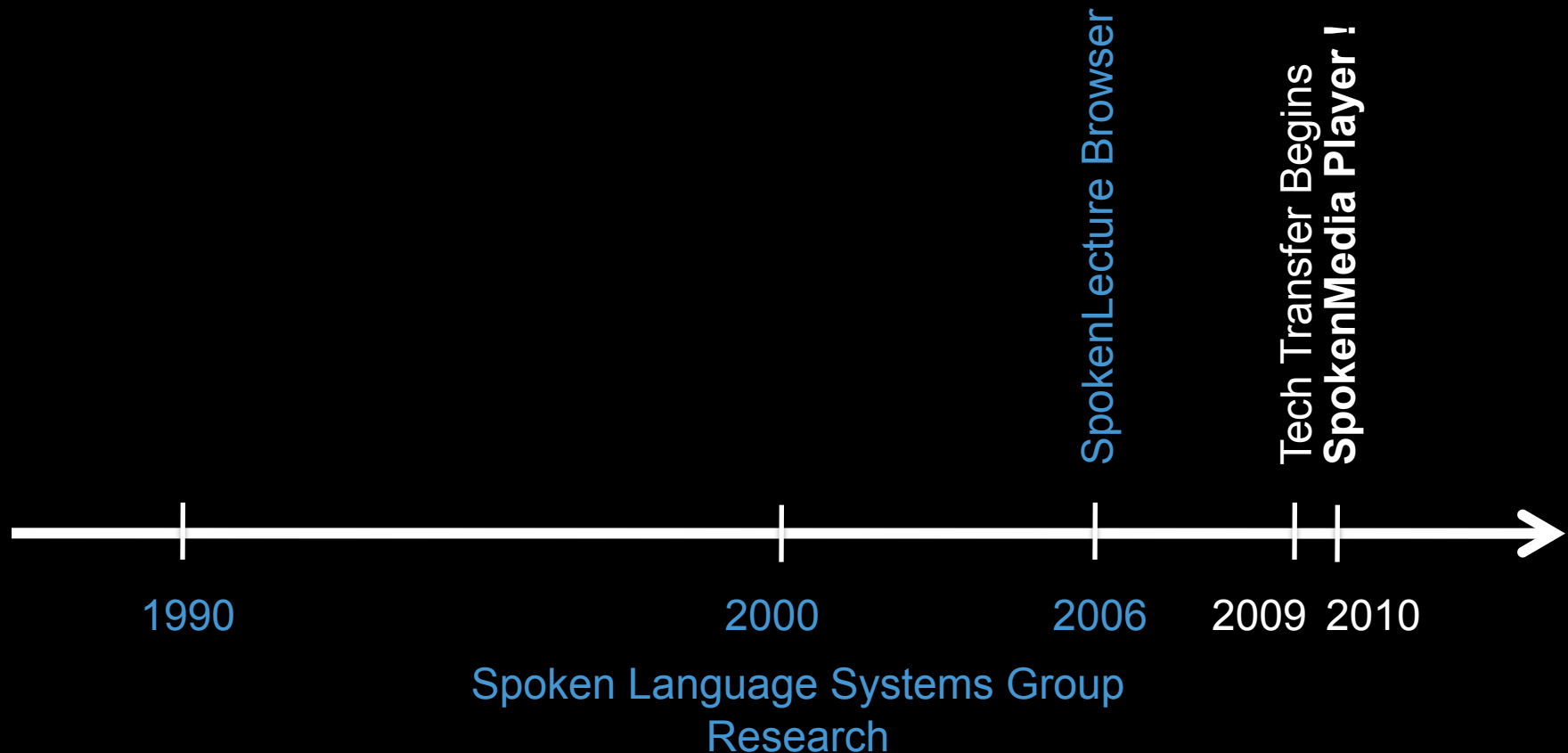
Let's see a demo!

Demo

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

Search transcript

Prof. Bish Sanyal

00:00   07:36

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)

How Does it Work?

Lecture Transcription Workflow

Input Files | Transcript Creation | Output Media Linked Transcript | SpokenMedia

- Video/Audio → Audio Ripper
- Audio*
- Acoustic Model*
- Metadata*
- Transcript, Slides, Docs → Text Processor → Domain Model*
- Speech Recognizer — Locally Processed -or- Cloud Service
- Video/Audio
- Time-coded Transcript*
- Search and Retrieval
- Spoken Media Player
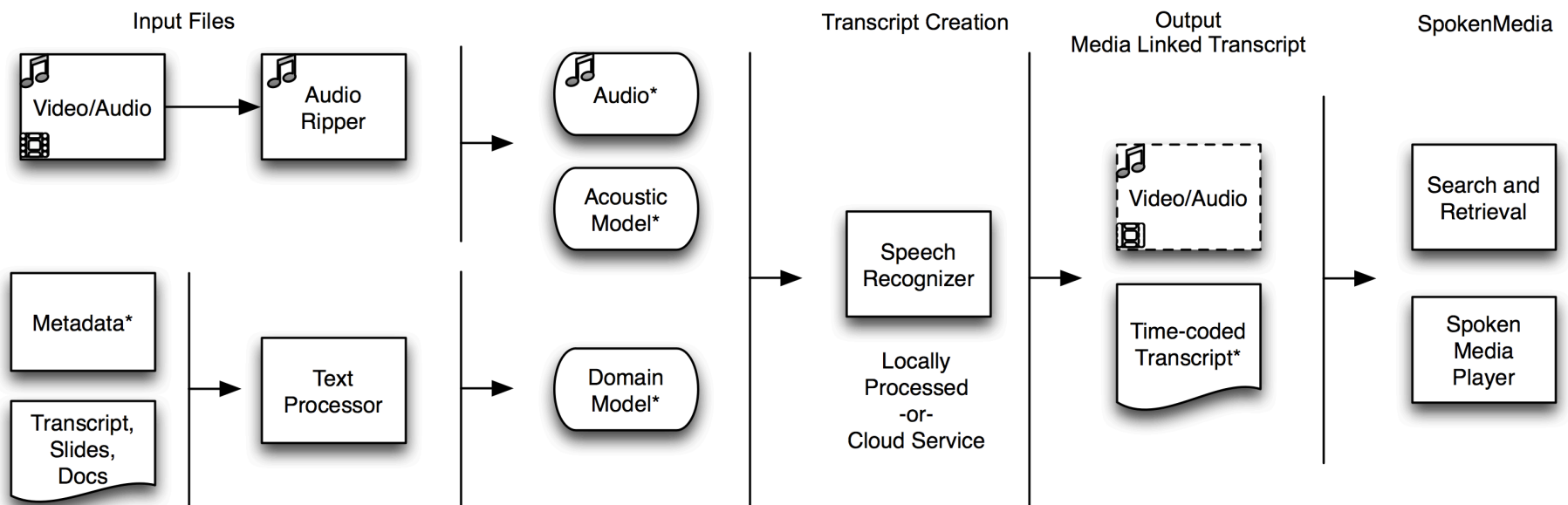
Recognizer Accuracy? ~85%

- ## Accuracy
  - Domain Model and Acoustic Model
  - Internal validity measure
  - Single 100% accurate transcript for a full course
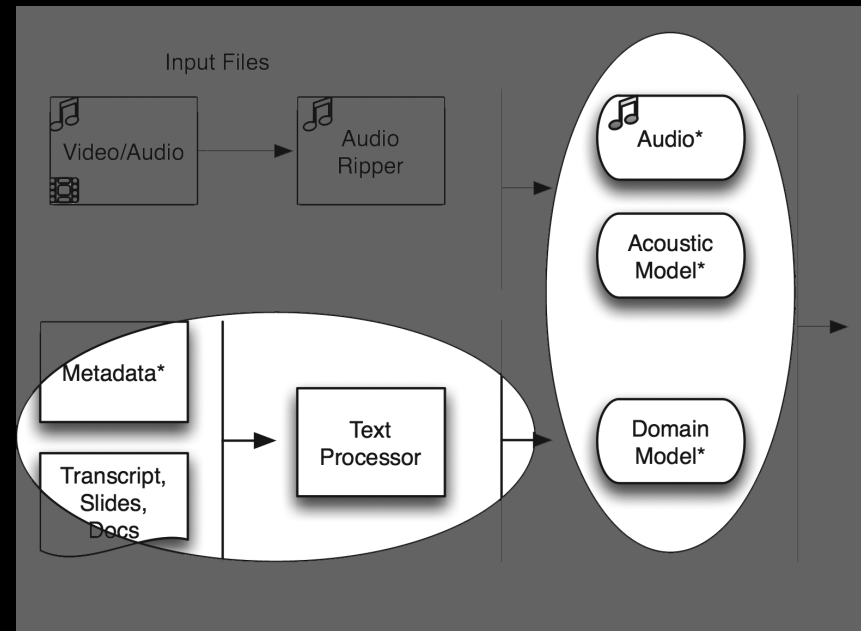
Ongoing research by Jim Glass and his team

What works today?

Lecture Transcription Workflow

Input Files

Transcript Creation

Output
Media Linked Transcript

SpokenMedia

- Video/Audio → Audio Ripper
- Audio*
- Acoustic Model*
- Metadata*
- Transcript, Slides, Docs → Text Processor
- Domain Model*
- Speech Recognizer

  Locally Processed
  -or-
  Cloud Service
- Video/Audio
- Time-coded Transcript*
- Search and Retrieval
- Spoken Media Player
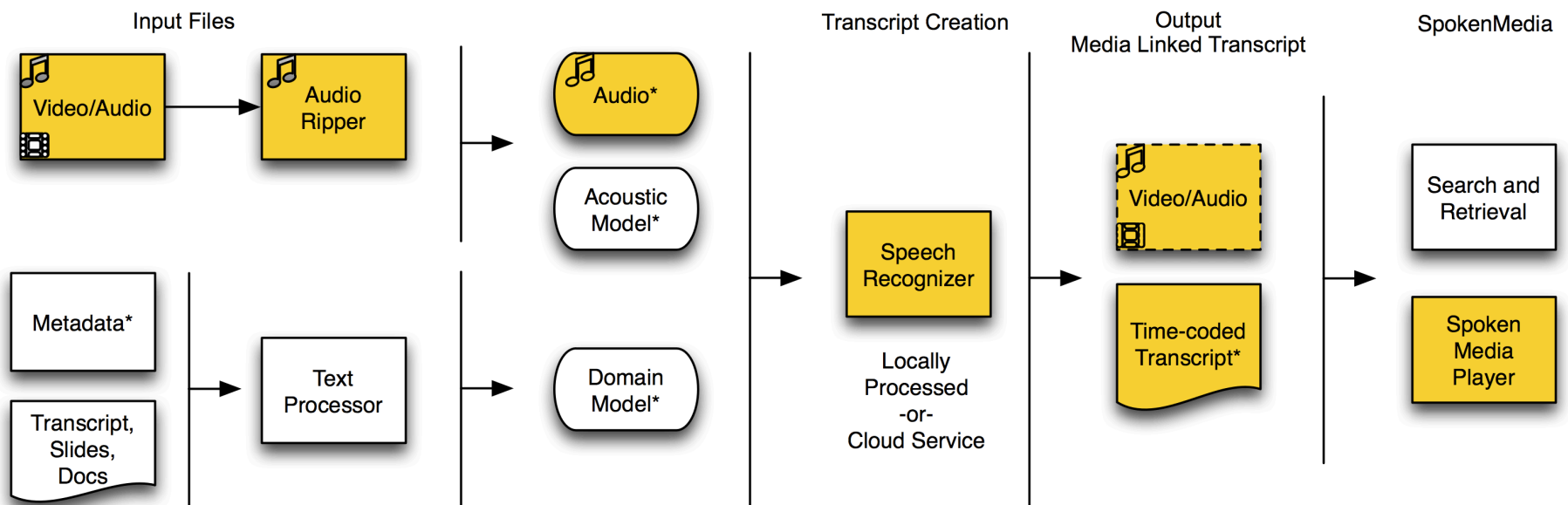
Transcript "Errors"

- "angular momentum and forks it's extremely non intuitive"
  - "folks"?
  - "torques"?
- "introduce both fork an angular momentum"
  - "torque"!

.....

we're now answering the part of eight oh one which is the most difficult for students and faculty alike ... we are going to enter the domain of angular momentum and forks it's extremely non intuitive ... the good news however is that b will stay with this concept for at least four five lectures today i will introduce both fork an angular momentum ... what is angular momentum if an object has a mass m and it has a velocity v ... then clearly it has a momentum ... v that's very well defined your reference frame the product of m and v ... thank the momentum ... i can take relative to any point i choose i choose this point q arbitrary ... this now ... is the position vector which i call our of q ... but this angle buffet to ... an angular momentum relative to that point it's a vector or ... is the

That's what we have today...

- Features
  - Video linked transcripts
  - "Bouncing Ball" follow along
  - Search within a video
  - Multiple transcript language support
- Challenges
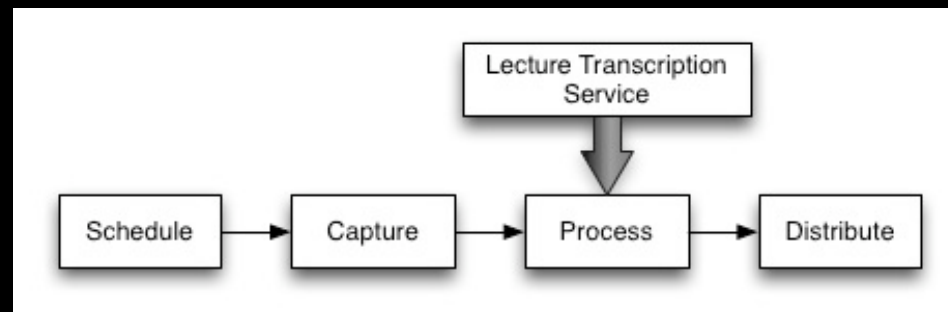 - Accuracy (partial toolset)

Where are we heading?

- Improved accuracy
- Automate and improve processing
- Search across multiple video transcripts

- Starting a lecture transcription service

Lecture Transcription Service

- ## Integrate with media production workflows
  - At MIT, University of Queensland

- ## Stand-alone service
  - Test with external content (video) producers

A Lecture Transcription Service? Caveats

- Lecture-style content (technology optimized)
- Up to 85% accuracy
  - (good for search, not sure about accessibility)
- English-language audio
  - (need much more research for other languages)
- Processing hosted at MIT (current thinking)
  - Submit jobs via MIT-run service
  - Contribute audio, models, transcript for further research

MIT OEIT Office of Educational Innovation and Technology

Test it for yourself!

http://spokenmedia.mit.edu/

http://sm.mit.edu/upload

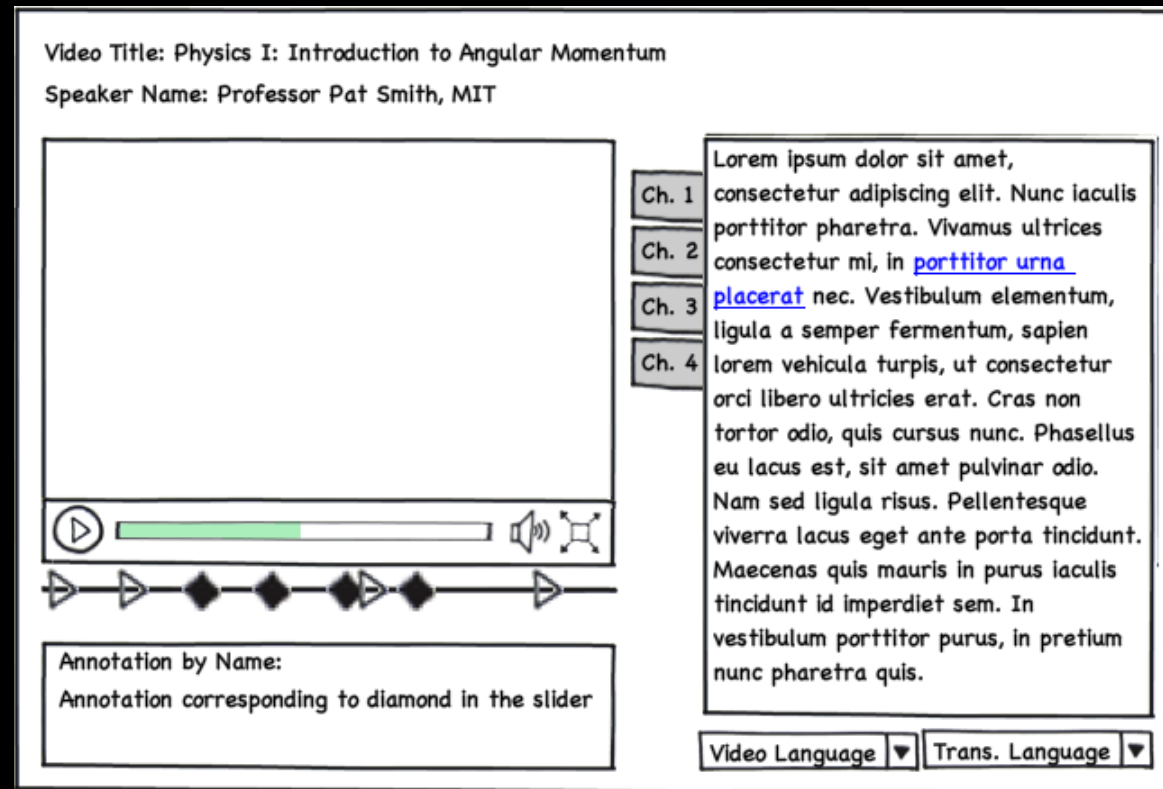**MIT OEIT** Office of Educational Innovation and Technology

Toward Rich Media Notebooks
Improving the User Experience

- Innovative player interfaces (soon)
  - Bookmarking and annotation
  - Clip creation and authoring
- Transcript editing (soon)
- Searching across collections of videos

MIT OEIT Office of Educational Innovation and Technology

Player with Annotation Mockup

Video Title: Physics I: Introduction to Angular Momentum

Speaker Name: Professor Pat Smith, MIT

Ch. 1
Ch. 2
Ch. 3
Ch. 4

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc iaculis porttitor pharetra. Vivamus ultrices consectetur mi, in porttitor urna placerat nec. Vestibulum elementum, ligula a semper fermentum, sapien lorem vehicula turpis, ut consectetur orci libero ultricies erat. Cras non tortor odio, quis cursus nunc. Phasellus eu lacus est, sit amet pulvinar odio. Nam sed ligula risus. Pellentesque viverra lacus eget ante porta tincidunt. Maecenas quis mauris in purus iaculis tincidunt id imperdiet sem. In vestibulum porttitor purus, in pretium nunc pharetra quis.

Annotation by Name:
Annotation corresponding to diamond in the slider

Video Language ▼   Trans. Language ▼

Editing Interfaces

Soon
(we're designing the editing interfaces right now)

Thanks!

spokenmedia.mit.edu

Brandon Muramatsu mura@mit.edu
Andrew McKinney mckinney@mit.edu
Peter Wilkins pwilkins@mit.edu

MIT, Office of Educational Innovation and Technology

MIT OEiT Office of Educational
Innovation and Technology