

Improving the OER Experience: Enabling Rich Media Notebooks of OER Video and Audio

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

MIT, Office of Educational Innovation and Technology

Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Improving the OER Experience: Enabling Rich Media Notebooks of OER Video and Audio. Presented at OER10: Cambridge, UK, March 23, 2010.
Unless otherwise specified, this work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License



Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Improving the OER Experience: Enabling Rich Media Notebooks of OER Video and Audio. Presented at OER10: Cambridge, UK, March 23, 2010.

Unless otherwise specified, this work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License

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 relying upon Web video for courses

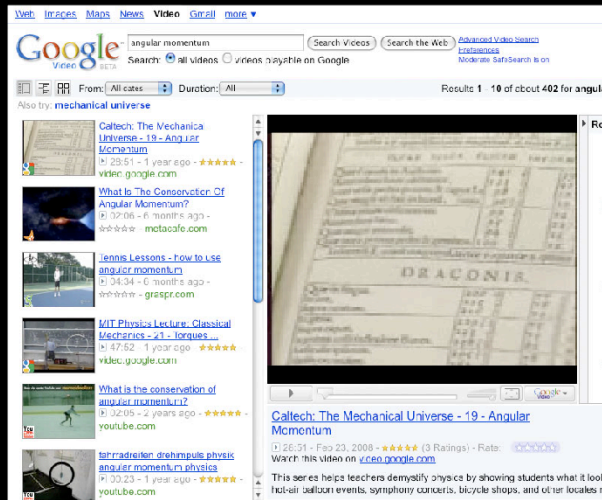
Why are we doing this?

- 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.
 - A couple different models:
 - UC Berkeley (and most of the world) recording courses for matriculated/enrolled students...and then everyone else
 - MIT OpenCourseWare publishing snapshots of courses
- 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
 - Time shifting (ucb)
 - Study tool (ucb, students mark in their personal notes when they don't understand something during the class to go back and review later)
 - Learning from other instructors (ucb)
 - Disabilities (ucb, learning, audio)
 - Course Selection (ucb)
- 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.

What are the challenges?

- Search
 - Volume
 - Segmented by Web, Video

Google Search for
"angular momentum"
Performed April 2009



What are the challenges?

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)

What are the challenges?

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

YouTube, MIT OCW Physics 8.01 - Lecture 20
Retrieved August 2009

webcast.berkeley, Physics 8A, 002, Spring 2009
Retrieved August 2009

webcast.berkeley

Physics 8A, 002 - Introductory Physics
TTh 12:30-2 | 19 SEMESTER
Instructor: Joe PAUMIS
Introductory Physics

This work is licensed under a Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License

Lecture Archive		
Tue, Jan 20	Lecture 1	
Thu, Jan 22	Lecture 2	
Tue, Jan 27	Lecture 3	
Thu, Jan 29	Lecture 4	
Tue, Feb 03	Lecture 5	
Thu, Feb 05	Lecture 6	
Tue, Feb 10	Lecture 7	
Thu, Feb 12	Lecture 8	
Tue, Feb 17	Lecture 9	
Thu, Feb 19	Lecture 10	

Text Comments (26) Options
methur63ahala01haah (6 days ago)
he is so far the best physics instructor i have encountered.

Sign in to post a Comment

Reply

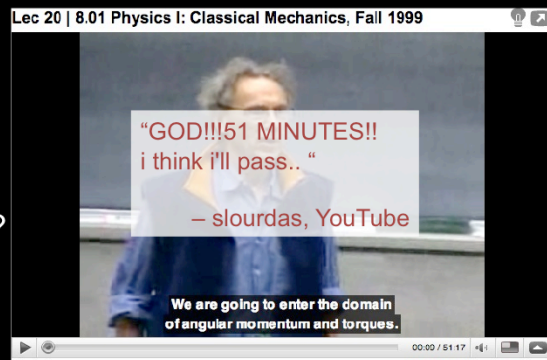
Classical Mech
Lec 23 | 8.01 Ph
Classical Mech
Lec 24 | 8.01 Ph
Classical Mech

What are the Challenges? Description

- Title of video/link
 - Text description of video (typically short), or the text surrounding an embedded video
 - Very infrequently Metadata
-
- Or Videos are described relative to their users, in the case of webcast.berkeley, they're listed by lecture (so are MIT OCW's), but in this example that's all we have, it'll make more sense to the students in the classes.

What are the challenges?

- 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>

What are the additional challenges?

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



We're living in a video world...but only have text to use for search...

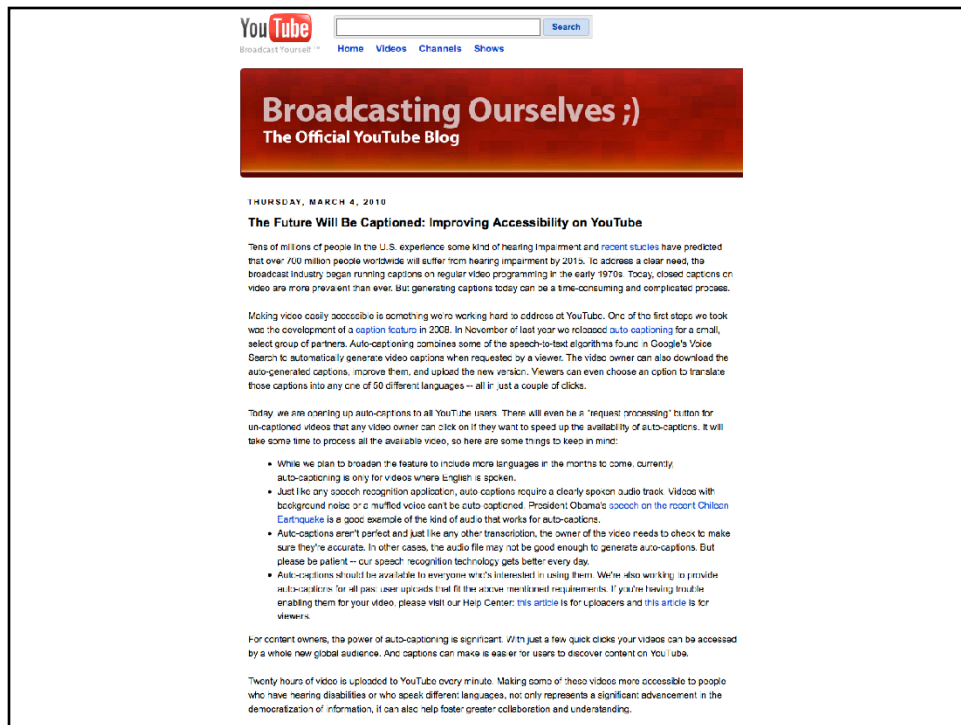
Disconnects—static—between finding/describing, segments/full-video, etc.

Why do we need these tools?

- **Improve search and retrieval**
- **Improve user experience**
- Captioning for accessibility? With correction?
- Facilitate translation?

Why do we need 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
- We want it for search and retrieval
- If a collaborative process, perhaps wiki-like correction and editing can improve the transcripts for captioning
- Or perhaps we can use it for translation



We should have something by April 2010

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)

We're not trying to compete with Google. But since you're probably wondering, how what we're doing compares...

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...

Enabling Research

James Glass
glass@mit.edu



- Speech recognition research
 - 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



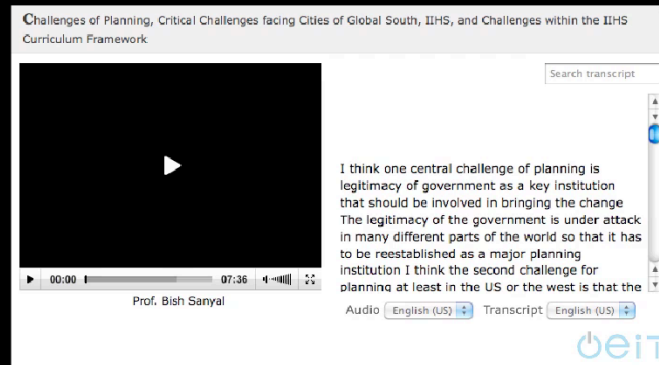
- Processor, browser, workflow
- Prototyped with lecture & seminar video
 - MIT OCW (~300 hours, lectures)
 - MIT World (~80 hours, seminar speakers)

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

SpokenMedia Player and Demo

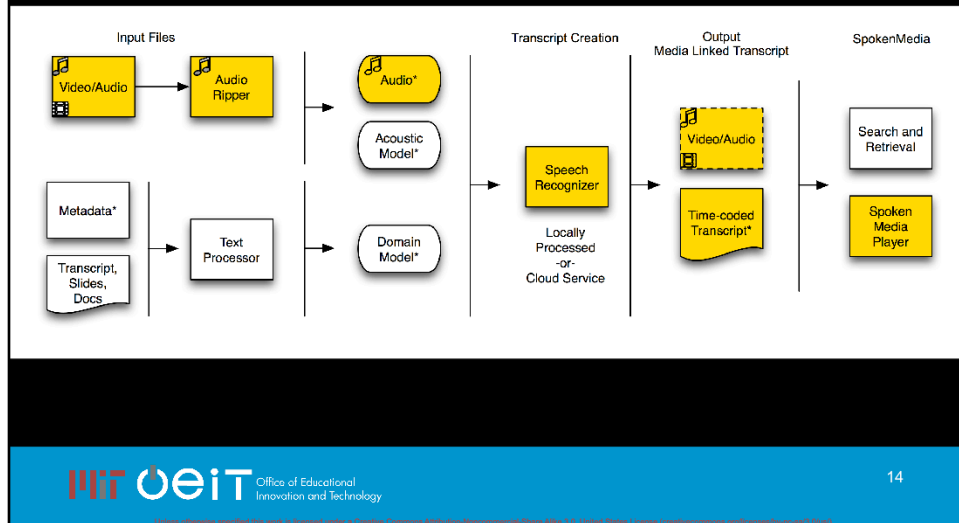


This demo is from the Indian Institute for Human Settlements

- There are a wide variety of speakers with different dialects of English
- Try out Bish Sanyal for a 100% accurate hand transcript in our player, along with a Hindi translation. Search in either English or Hindi.
- Or try Geetam Tiwari, for another 100% accurate hand transcript (to demonstrate what's possible)
- All the other speakers have transcripts from 40-60% accuracy using the SpokenMedia processing.

What works today?

Lecture Transcription Workflow

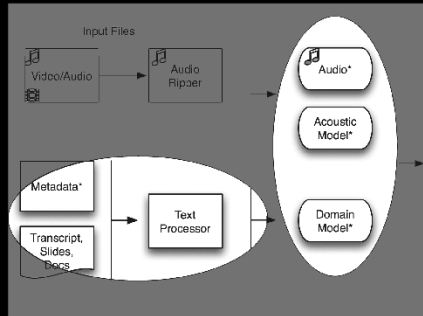


What works as of March 2010?

- Audio
 - System only needs audio (waveform), extracts from video
- Domain Model (base is generic domain model)
 - Using a Generic Domain model
- Acoustic model (base is generic speaker model)
 - Using the American-English-male-voice generic speaker model
- Process—With audio, domain and speaker models
- Output
 - Time coded transcript (standard formats)
 - Links media and transcript
- Applications
 - Player

Recognizer Accuracy? Up to 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

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”

- “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

Transcript “Errors”

- Recall, recognizer can have up to 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

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)

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?

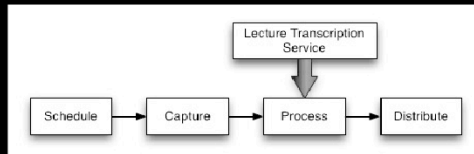
- Improved accuracy
- Search across multiple video transcripts
- Automate and improve processing
 - > Starting a lecture transcription service

Where are we heading?

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

Goals of the Lecture Transcription Service

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



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

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

- 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

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
 - Most of worldwide research has been in English, there is research in other languages – ones we've been talking with Jim Glass about include Chinese, Spanish
 - Need speech researchers in the language, coupled with research Jim Glass has done
- Current plan to host a web service from MIT
 - Contribution to research and a hosted collection will be important aspect of participation

Toward Rich Media Notebooks Improving the User Experience

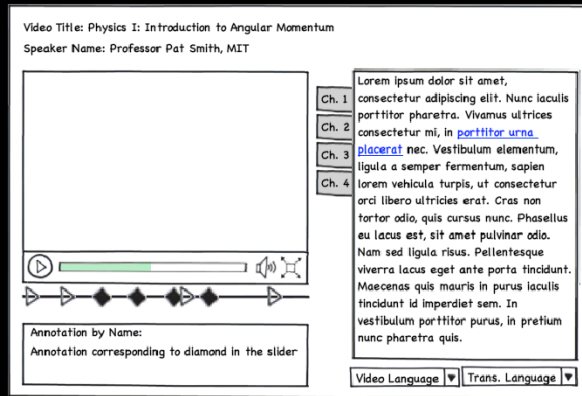
- Innovative player interfaces (prototypes)
 - Bookmarking and annotation
 - Clip creation and authoring
- Transcript editing (prototypes)
- Searching across collections of videos (soon-ish)

In Collaboration with the Univesite de Lyon 1

Toward Rich Media Notebooks

- Implement innovative player interfaces including 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
 - Searching across collections of videos

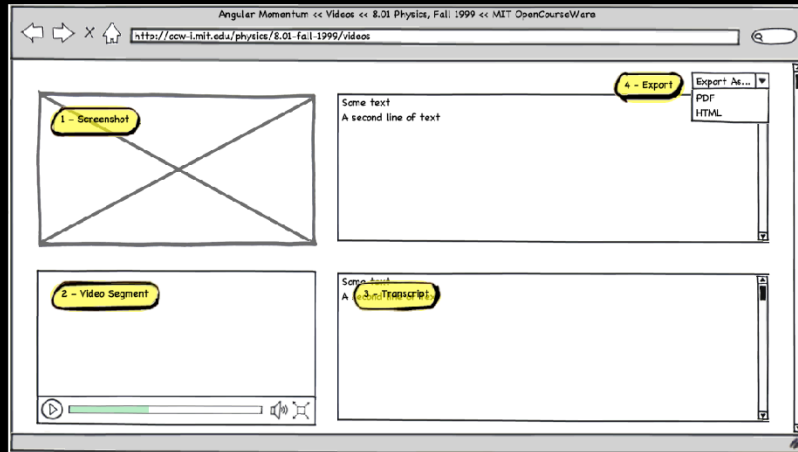
Player with Annotation Mockup



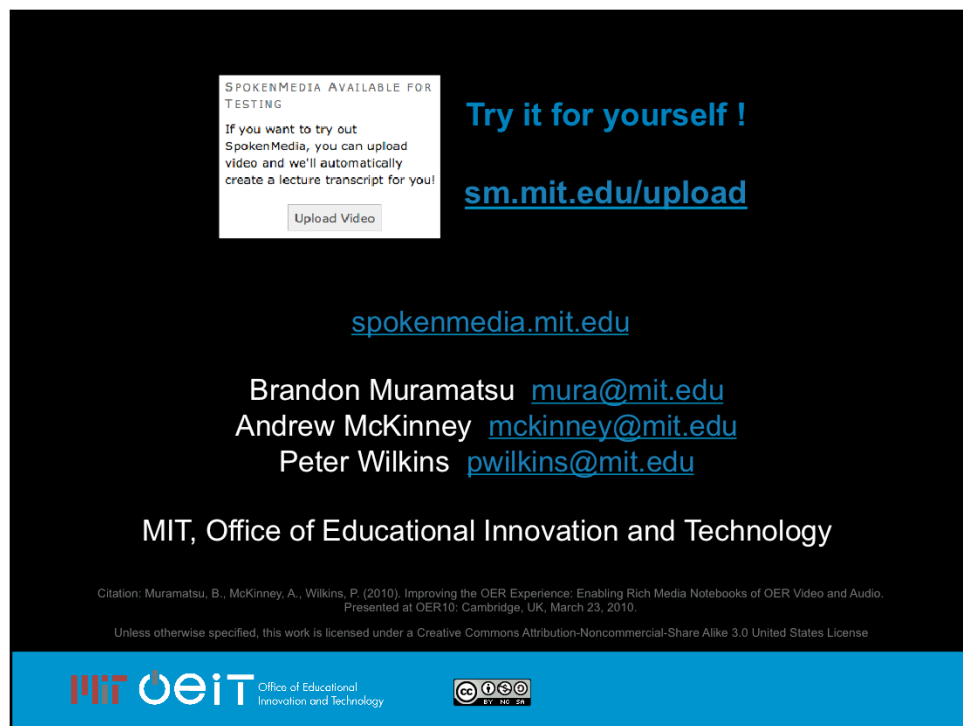
Here's an example of what our next generation player might look like.

- Ability to add “chapters”, “annotations” and “bookmarks”
- Still can change audio/transcript languages
- We did this mockup in late-February 2010
- Universite de Lyon 1 has a functional prototype in mid-March 2010

Snapshots and Transcripts



- New mockup for display of video by chapter and related transcript, mid-March 2010
- In collaboration with Universite de Lyon 1
- Playback of segment from video (e.g., 10 minute segment of 1 hour lecture) along with transcript
- Or, export as stand-alone HTML
- Or, export to PDF with thumbnails (keyframes) from video and transcript for offline “viewing” (if the video images are important)



SPOKENMEDIA AVAILABLE FOR TESTING

If you want to try out SpokenMedia, you can upload video and we'll automatically create a lecture transcript for you!

[Upload Video](#)

Try it for yourself !

sm.mit.edu/upload



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

Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Improving the OER Experience: Enabling Rich Media Notebooks of OER Video and Audio. Presented at OER10: Cambridge, UK, March 23, 2010.

Unless otherwise specified, this work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License

 Office of Educational Innovation and Technology 

Citation: Muramatsu, B., McKinney, A., Wilkins, P. (2010). Improving the OER Experience: Enabling Rich Media Notebooks of OER Video and Audio. Presented at OER10: Cambridge, UK, March 23, 2010.

Unless otherwise specified, this work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License