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


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?

• More & more videos on the Web
  – Universities recording course lectures
  – Students relying upon Web video for courses

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.
What are the challenges?

- Search
  - Volume
  - Segmented by Web, Video
What are the challenges?

- Description
  - Course and Lecture Title
  - Summary
  - Metadata?
What are the challenges?

• Interaction & Use
  – Transcripts / captions
    • Do they exist?
    • Cost?
  – Full video vs. segments

We’re living in a video world…but only have text to use for search…
Why do we need these tools?

• Improve search and retrieval
• Improve user experience

• Captioning for accessibility? With correction?
• Facilitate translation?
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 it easier for users to discover content on YouTube.

Twenty hours of video is uploaded to YouTube every minute. Making some of these videos more accessible to people who have hearing disabilities or who speak different languages, not only represents a significant advancement in the democratization of information, it can also help foster greater collaboration and understanding.
## Comparing SpokenMedia and YouTube Auto-Caption?

<table>
<thead>
<tr>
<th>YouTube</th>
<th>SpokenMedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Scale ✔</td>
<td>• Limited</td>
</tr>
<tr>
<td>• Research-basis ✔</td>
<td>• Research-basis ✔</td>
</tr>
<tr>
<td>• For all videos ✔ (soon)</td>
<td>• Service by request</td>
</tr>
<tr>
<td>• No transcript/caption export (?)</td>
<td>• Transcript/caption export available ✔</td>
</tr>
<tr>
<td>• YouTube hosted</td>
<td>• Hosted anywhere ✔</td>
</tr>
<tr>
<td>• Accuracy based on general patterns (?)</td>
<td>• Accuracy based on custom models ✔ (soon)</td>
</tr>
<tr>
<td>• No transcript editing (?)</td>
<td>• Transcript editing ✔ (soon)</td>
</tr>
</tbody>
</table>
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
Spoken Lecture Project

- 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
SpokenMedia Player and Demo

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...
What works today?

Lecture Transcription Workflow
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
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 h will stay with this concept for at least your 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 another way in 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
- Search across multiple video transcripts
- Automate and improve processing
  → Starting a lecture transcription service
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
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
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
Player with Annotation Mockup

Video Title: Physics I: Introduction to Angular Momentum
Speaker Name: Professor Pat Smith, MIT

Snapshots and Transcripts
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


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