Interoffice Memorandum

Refer to OCE DJO 14-024
April 21, 2014

To: Charles White (172A)
From: David Oberhettinger
Subject: Experiment in Low Cost Video Capture of Technical Interchange Meetings

JPLTube is the result of a Microsoft Research partnership with JPL and the NASA IT Labs that allows the user to perform textual searches of transcribed audio-video content within the 1155 currently active videos that have been uploaded to *JPLTube*. The site is restricted to U.S. Persons within the JPL firewall, but anyone can upload a video and, if desired, restrict access to a limited set of JPL individuals. An interactive (i.e., synchronized scrolling) transcript is automatically generated for each upload using MAVIS technology. To date, there have been 1.193 million page views of *JPLTube*, or an average of 11,400 page views per day.

Video capture and sharing of technical knowledge has been of limited utility to date due to the difficulty of searching for specific information among hundreds of hours of video. Hence, *JPLTube* may represent significant disruptive technology because a search brings up one or more video snippets specific to a search. For example, the search string "alcohol wipes" brings up a video snippet from an 8-hour seminar taught by a Planetary Protection expert; in the short snippet, the experts lectures on the effectiveness of alcohol wipes in sterilizing spacecraft surfaces.

However, a remaining constraint to widespread use of video capture is the cost of videography. Using JPL media services professionals to record a one-hour seminar or technical interchange meeting may cost over $1000. Hence, the Office of the Chief Knowledge Officer (OCKO) conducted an experiment in low cost video capture of technical knowledge. We uploaded three versions of the seminar by Roger Diehl ("Earth Fly-By Risk Considerations") to *JPLTube;* they were recorded using a:

1. Consumer-quality Sony HandyCam zoom video camera connected to a lapel microphone clipped to Mr. Diehl,
2. Consumer-quality Panasonic Lumix, Model DMC-FZ70, still camera with video capability, and a
3. iPhone 5s.

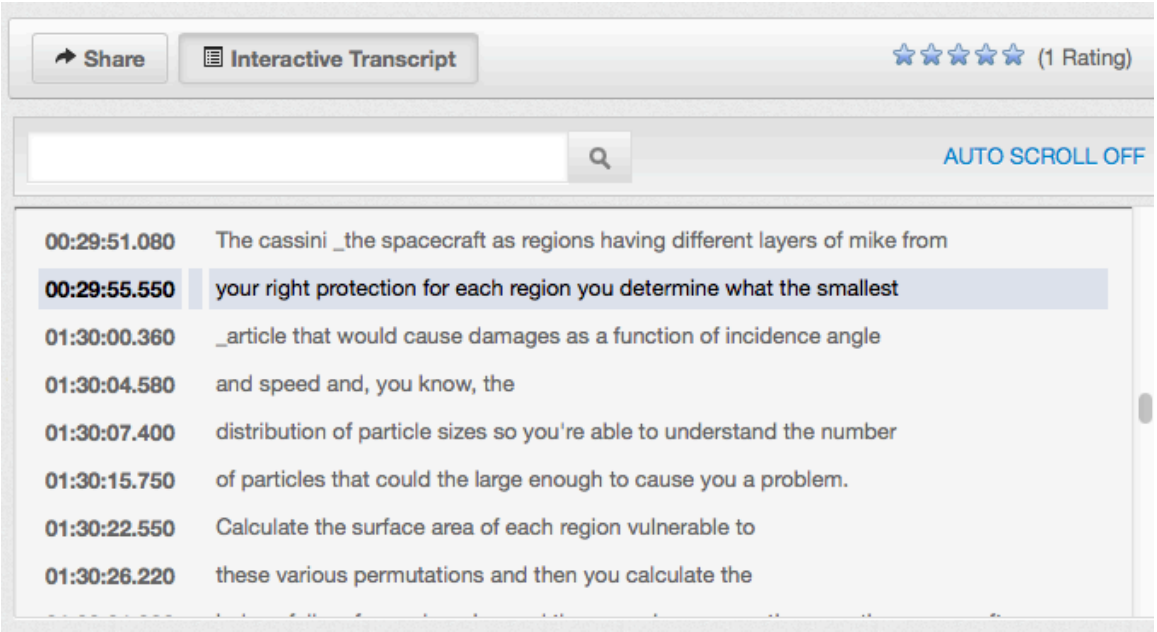
The primary concern was providing an audio track of sufficient quality to enable high fidelity audio transcription to text. The venue employed was the 301-367 conference room, an average size (35-seat) conference room with average fluorescent lighting. The video camera wielded by Charles White of the JPL OCIO was free to move about the conference room, the tripod-mounted Lumix was positioned against the left wall of the conference room, and the iPhone was propped up in the middle of the conference table. All three cameras captured both the speaker and the presentation slides with good video quality, except the color balance produced by the iPhone under fluorescent lighting needed post-filming adjustment. All three audio tracks provided good clarity; listening to them, the audio quality appeared to match that of the live presentation that was recorded. Although the

Lumix still camera and the iPhone used only their built-in microphones, they captured both the speaker and the questions asked by the audience. (Cord and cordless external microphones are available from third-party vendors to attach to desktop tripods that are compatible with the Lumix and iPhone.)

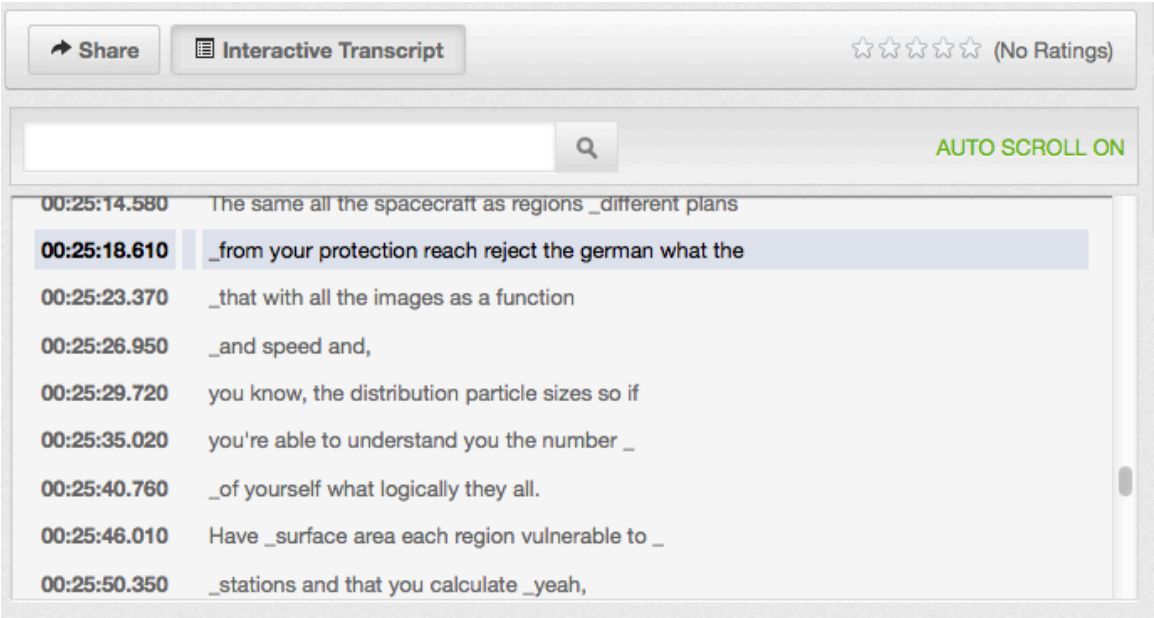
All three video versions of the seminar were uploaded to *JPLTube* so that the accuracy of the three automated transcriptions could be compared. As a baseline for comparison, I manually transcribed the following half-minute segment chosen at random.

> **Baseline**: Oberhettinger's manual transcription of the audio track
>
> *The Cassini modeled the spacecraft as having different layers of micrometeoroid protection for each region to determine what the smallest mass particle that would cause damage as a function of incidence angle and speed. And you know the distribution of particle sizes and so you are able to understand the number of particles that could be large enough to cause you a problem. Calculate the surface area of each region vulnerable to these various permutations, and then you calculate…*

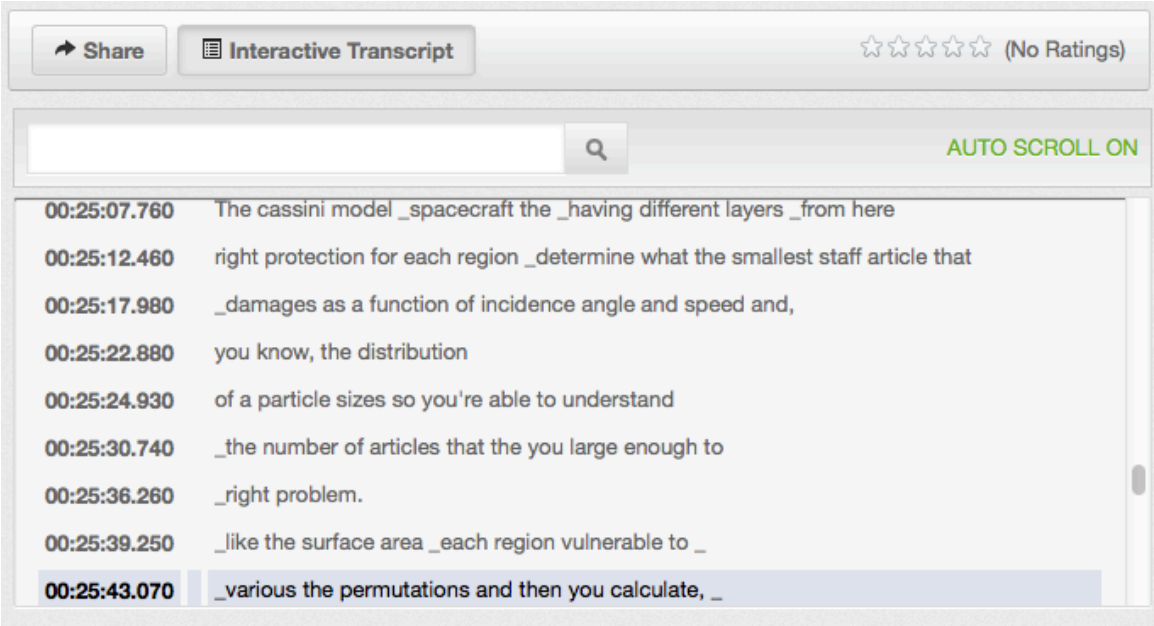
Version #1: The automated transcription generated by *JPLTube* from the Sony HandyCam video camera audio track, following the video upload to *JPLTube*, is shown below. Note that words or phrases that are unintelligible to the system are depicted with an underscore.



Version #2: The automated transcription generated by *JPLTube* from the audio track associated with the video made by the Lumix still camera, uploaded to *JPLTube* as Part 1 and Part 2, is shown below:



Version #3: The automated transcription generated by *JPLTube* from the iPhone video is shown below, uploaded to JPLTube as Part 1 and Part 2:



The transcripts for the video camera with lapel microphone (Version #1) and the iPhone (Version #3) appear to have achieved about 80% accuracy. Technical terms that might be used in searches of *JPLTube* seem to have achieved almost the same level of accuracy. Version #2 using the still camera's microphone and video capability has the lowest fidelity, perhaps because it was the farthest camera/microphone from the speaker.

In conclusion, a cellphone like the iPhone appears quite practical as a low cost and flexible alternative for capturing audio-video with a single speaker giving the presentation in a medium-sized conference room. The video quality may be adequate even without supplementary lighting. The recorded audio may be adequate for *JPLTube* automated

transcription even without an external microphone. A small desktop tripod with a clamp for the cellphone would be preferable to propping up the cellphone. A lapel microphone plugged into the iPhone would likely provide the most consistent transcript quality.