A Crowdsourcing Caption Editor for Educational Videos

Rucha Deshpande, Tayfun Tuna, Jaspal Subhlok
University of Houston
Computer Science Department
Houston, TX
rborgaonkar@uh.edu, ttuna@uh.edu, jaspal@uh.edu

Lecia Barker
University of Texas
School of Information
Austin,TX
lecia@ischool.utexas.edu

***Abstract*—Video of a classroom lecture has been shown to be a versatile learning resource comparable to a textbook. Captions in videos are highly valued by students, especially those with hearing disability and those whose first language is not English. Captioning by automatic speech recognition (ASR) tools is of limited use because of low and variable accuracy. Manual captioning with existing tools is a slow, tedious and expensive task. In this work, we present a web-based crowdsourcing editor to add or correct captions for video lectures. The editor allows a group, e.g., students in a class, to correct the captions for different parts of a video lecture simultaneously. Users can review and correct each other's work. The caption editor has been successfully employed to caption STEM coursework videos. Our findings based on survey results and interviews indicate that this innovative crowdsourcing tool is effective and efficient for captioning lecture videos and has considerable value in educational practice. The caption editor is integrated with Indexed Captioned Searchable (ICS) Videos framework at University of Houston that has been used by dozens of courses and 1000s of students. The ICS Videos framework including the captioning tool is open source software available to educational institutions.**

I. Introduction

Video is gaining popularity as a learning resource. Video recordings of classroom lectures are often made available as additional material for a conventional course, as the core of a distance/hybrid learning course, or posted publicly for community learning. Lecture videos are posted on a large scale on portals such as MIT OpenCourseware and Apple's iTunes University. In recent years MOOCs (Massive open online courses) driven by video and other features have emerged as a potential disruptive technology for delivery of education. There is a substantial body of research that has established that video is a versatile learning resource that is considered valuable by students and instructors [2], [6], [13], [16], [18].

This research is in the context of the *ICS Videos* project that has developed videos enhanced with Indexing, Captioning, and Search capability that are designed for quick access to video content. Videos were typically recorded on Tablet PCs that allow free mixing of prepared (PowerPoint) viewgraphs with hand annotations and illustrations. A snapshot of the ICS video player highlighting the key features is shown in Figure 1. The framework was deployed at University of Houston and partners, and used by 1000s of students for coursework across Biology, Computer Science, Chemistry, Geology, Physics, and Mathematics. Typically, videos were provided as additional learning material and did not replace classroom lectures. Several thousand students were surveyed and hundreds of students participated in focus groups during the project. Conclusions from this project relevant to the research presented in this paper are i) Videos are a very valuable learning resource and ii) Captions enhance the value of videos significantly.

Captioning greatly improves the accessibility of lecture videos for the deaf or hard of hearing, as well as students whose native language is not the primary language in the classroom. Another common scenario where captions are particularly helpful is when the primary language of the instructor is not the classroom language. Not only do captions allow students improved access to the material in a lecture, they are also learning the language. Captions are especially valuable for building vocabulary. Having captions allows students to stop and look up definitions of words, or recognize and remember the words as their professors are speaking them. Many technical lectures have specialized vocabulary. In such cases, seeing the spoken words can be very helpful. If the transcript is available in one language, the captions can be made available in any language desired to cater to the target audience. Finally captioning can make the video searchable which again improves access to the video content.

The major challenge in making captions widely available is the cost/effort involved. Currently available speech recognition technology is too error prone for technical videos for the results to be directly useful. Manual generation of captions is a time consuming process. Generating captions through professional transcription services is typically too expensive for ordinary classroom use. The central goal of the research presented in this paper is generation of captions as automatically and efficiently as possible.

This paper presents a crowdsourced caption generator/editor, named ICS Caption Editor. The basic idea is that students collaboratively caption educational videos for common use. The captioning tool allows groups of students to work collaboratively, cooperatively and asynchronously to caption a classroom lecture. While captioning an hour long video is a long, intimidating task, with crowdsourced captioning each student typically needs to spend only minutes to caption a lecture. The workflow for captioning includes automatically generating initial captions with automatic speech recognition tools followed by corrections by student groups with the ICS Caption Editor.

978-1-4799-3922-0/14/$31.00 ©2014 IEEE

This paper is organized as follows. Section II discusses prior work related to assessment of speech recognition and available caption editors. Section III discusses the ICS Video player and our experience with employing automatic speech recognition for captioning. Section IV presents the design and implementation of the ICS Caption Editor. Section V presents our experience with the caption editor along with survey results on value of captions and usability of the caption editor. Section VI contains conclusions.

II. Related Work

This work is necessary in large part because the current state of the art in speech recognition is inadequate [14], [24], although the subject is an active area of research, e.g., at Google [9]. A particular focus of research is noisy backgrounds [11] that is often the case for lecture videos. Broughton [4] assessed two commercial transcription tools, Dragon Naturally Speaking 5.0 and IBM ViaVoice 8.0, and concluded that there is a significant degradation in the accuracy of commercial ASR tools when conversational or spontaneous speech is used. Assessment of speech recognition systems and impact of errors on usefulness of speech recognition have been addressed in literature [5], [10]. In the background section of this paper, we also present our experience in adopting ASR systems.

A central contribution of this paper is a new caption editor. Several commercial caption editor products are currently available, although they are primarily designed for movies and home videos, not technical lectures. Examples include Subtitle Workshop [22], Express Scribe [12], Caption Maker [19], as well as the caption editor included in YouTube. These editors were not sufficient for our purposes for two kinds of reasons. They lack some of the convenience features to support easy usage for novices, in particular, the editing involved manual pausing and rewinding the audio while typing and correcting the captions. Also, there is no support for crowdsourcing.

Some companies such as 3PlayMedia [1], provide transcription services to a number of academic institutions for a fee. This is not an economical or scalable solution for academic institutions where the number of hours of audio to be transcribed is large.

IBM CCES [15] decomposes audio data of a video into segments and distributes (for example 1 minute audio) to its registered editors. This tool is web-based, supports looping audio, and allows experts to verify the content. Synote [23] is also a web-based caption editing tool. Synote stores edits of all users and uses a matching algorithm to check whether the users are in agreement, with a minimum number of editors required to finalize a correction.

The caption editor implemented in this work builds on these products with the goal of crowdsourced editing of the captions of technical lecture videos. Several innovative features were added including a visible video frame while correcting the captions, variable speed audio looping for the currently chosen sentence, a status interface indicating the correction/verification status of various captions, ability to verify and correct captions of other editors and ability to indicate if a caption appears correct or needs further review.

III. Background

A. ICS Videos

The research presented in this paper is in the context of the ICS (Indexed, Captioned, Searchable) Videos project at the University of Houston[18], [20]. The goal of the project is to ease navigation of lecture videos, making them a companion resource for learning, similar to a textbook. A video lecture is automatically partitioned into topical segments based on text and image analysis [21]. Video is searchable for keywords and concepts. Captions are developed for videos with speech recognition and crowdsourcing by students. All videos for an entire course (or department) are treated as a single videobook stream with global indexing and search capability. The ICS

Fig. 1. ICS Video Player with Index points, Search box, Captions and Transcript

Video Player that encapsulates indexing, search, and captioning is illustrated in Figure 1. An index panel is situated on the bottom of the player; each index point has a screenshot of the video at that point of time. A search box is located above the index points panel to start keyword search within a video. Keyword search across videos is also supported. Captions are displayed as an overlay at the bottom of the main video player and in a separate caption window on the right side. Users can scroll back and forth to read captions anywhere in the video.

The ICS video player has been deployed for dozens of courses and used by thousands of students for STEM coursework at University of Houston. Students surveys show that students consider videos to be a versatile learning resource and consider each one of the indexing, search, and captioning features to be extremely valuable [3], [20].

B. Automatic Speech Recognition tools

Speech recognition technology is constantly evolving. A series of experiments were conducted to understand the accuracy of the state of the art speech recognition tools. The tools deployed were Dragon Naturally Speaking Preferred 10, Windows Speech Recognition (WSR) and YouTube. This study was conducted with 3 professors (2 professors from Computer Science Department, one professor from Mathematics Department) in 3 different scenarios: *Lecture Transcription, Parroting* and *Dictation*.

Lecture Transcription: Recorded audio of live classroom lectures was transcribed to text by deploying ASR tools.

Dictation: The participants were asked to read a prescribed paragraph and the audio was recorded. Along with normal text, this paragraph has numbers, acronyms, proper nouns and some technical words. The resulting audio was transcribed to text with ASR tools.

Parroting: Parroting is the technique of repeating/imitating the words of another speaker. Parroting can be used to enunciate the words to the speech recognition engine and use a speaker for whom a voice profile exists with the ASR tool. In this experiment, a designated speaker (main author of this work) parroted the lectures from their professors. The resulting audio was transcribed to text.

These methods were employed with different tools. The results from the tool that provided *maximum accuracy* in each scenario are listed in Table I. Details of the experiments are available in [7].

TABLE I. ACCURACY WITH TRANSCRIPTION, DICTATION AND PARROTING

Participant	Transcription(YouTube)	Dictation(DNS)	Parroting(DNS)
Professor 1	71.40%	89.93%	94.76%
Professor 2	62.14%	83.13%	96.63%
Professor 3	70.80%	83.03%	96.1%
Average	68.11 %	85.36%	95.83%

It is clear from the results that transcription of live technical lectures has significant errors, making the transcripts/captions almost useless. Dictation has lower but still a substantial error rate, while parroting has good accuracy in the range of 94-97%. Further discussion is beyond the scope of this paper, but we point out that there are significant overheads in employing dictation or parroting for routine lectures.

The errors in sample transcripts were analyzed manually to understand the underlying causes. Results are shown in Table II. Roughly half the errors were attributed to speaker and tool weakness. This indicates that there is significant room for improvement for ASR tools, but bringing down error rates below a threshold will be a major challenge. More details of the experiments, including the precise definitions of terms used in Table II are available in [7].

TABLE II. ANALYSIS OF ERRORS BY ASR TOOLS

Category	Nature of error in DNS	Prof. 1	Prof. 2	Total	%	Total % per category
Tool's weakness incorrect	Incorrect hypothesis by the tool	113	39	152	50.16	50.16
Speaker's weakness	Disfluent speech	21	10	31	10.23	40.26
	Heavily accented speech	7	11	18	5.94	
	Conversational speech	18	12	30	9.90	
	Mixed words/ not enunciated well	25	17	42	13.86	
	Very low volume or moving away from microphone	0	1	1	0.33	
Independent	Out of Vocabulary words	1	2	3	0.99	9.57
	Homonyms	1	1	2	0.66	
	Ungrammatical construct (because of technical words)	19	0	19	6.27	
	Inaudible (student interaction)	5	0	5	1.65	
	Total	**210**	**93**	**303**	**100**	**100**

IV. ICS CAPTION EDITOR

A. Objectives

The reason for developing a custom ICS caption editor was that the state-of-the-art speech recognition tools often provided captions that were too erroneous to be of practical use. Also, the available commercial caption editors were not up to the task for STEM classroom lectures. Manual correction of the captions is required to be able to deliver reasonably accurate captions. Following were the key objectives in designing the caption editor:

Teamwork: The caption editor must allow a group of students to work collaboratively and asynchronously from different geographical locations to caption a video lecture.

Technical content: The audio content can be complex and technical in nature.

Audio quality: The quality of audio can be poor for a number of reasons such as poor recording, instructor accents, and conversational nature of a classroom environment.

Simple navigation: The tool is designed for ordinary students not professionals, so it must be easy to use and navigate.

B. Implementation

The ICS Caption Editor was designed and implemented based on these goals. It is a custom built web-based, crowdsourcing tool that allows addition or editing of captions efficiently. Typically the editing process starts after initial captions are obtained from a speech recognition tool. Feedback from early versions of the tool influenced the final design. A snapshot of the user interface of the caption editor is shown in Figure 2. The letters *A, B....K* refer to various components/icons of the tool. We discuss the main features of the tool and how the design goals were achieved.

Partitioning for crowdsourcing: Caption text is divided into sections, each with 5 sentence long blocks of audio/text. An individual user visually identifies a section that is available to work on, locks it, and starts the editing process. Different users can work on different sections at the same time. The time span and current text is displayed for each section. These aspects are indicated with labels *B, C, F* in Figure 2.

Collaboration and cooperation: After entering/editing and saving a caption, the user decides if the caption is satisfactory or needs further review. Because of the technical nature of content and the variable quality of the audio, a user is often not certain if their captions are accurate. The user can accordingly mark a caption as *Complete* or *Needs further review*. In the latter case, another user will review the caption. A count of the number of users that have reviewed a caption is also maintained, allowing a policy where a minimum number of users must review a caption before it is complete. An intuitive color coding scheme is used to indicate the status of captioning

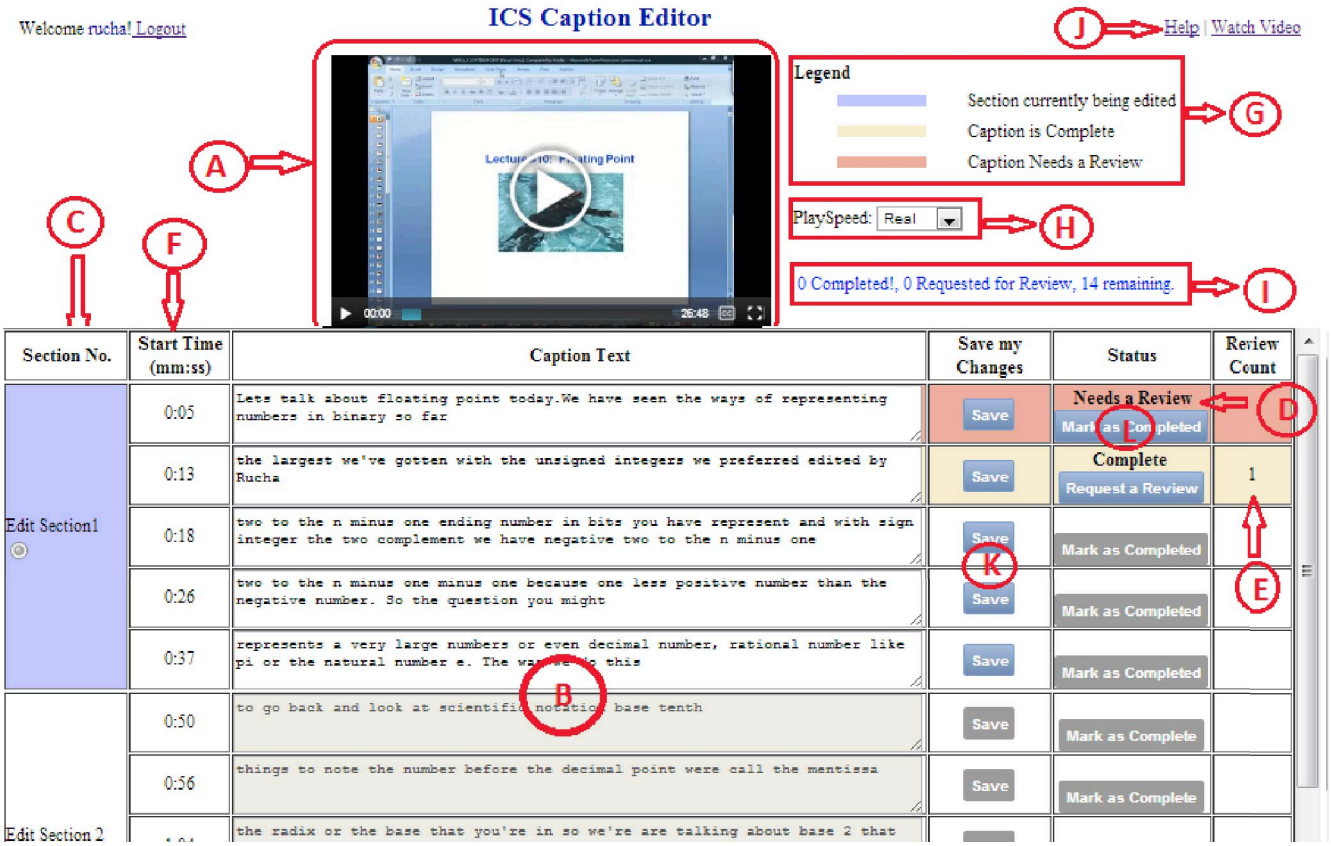

Fig. 2. ICS Caption Editor Interface

of various segments. These aspects are indicated with labels *D, E, G, I, L, K* in Figure 2.

Convenience features: The audio for a segment repeats in a loop until the user has completed editing the corresponding text. There is a *Playspeed* tool that can increase or decrease the speed of audio playback. A visual display of the video corresponding to the caption being created is displayed simultaneously. These features are important for classroom videos as the content can be difficult to understand and is often connected to the video display such as a viewgraph. These aspects are indicated with labels *A, H* in Figure 2.

C. Technology

The Caption Editor is built using PHP Version 5.2.4. The Apache Tomcat 2.2 web server is used. The video is displayed using HTML5 elements and the MediaElement.js. MySQL 5.2 is used in the backend for user management to store editor related data such as status of captions, locked sections, etc. Database updates are done in the background using AJAX. JQuery (version: 1.4.1), a JavaScript library, is used to traverse the HTML document and to make AJAX requests. Initial captions are obtained from YouTube using Google API Java Clients 1.10.3. More implementation details are available in [7]. The framework is entirely built on top of open source technology.

V. Evaluation

The key objectives of evaluation were to determine the value attached to captions by students and the usability and effectiveness of the ICS Videos caption editor. The evaluation was conducted by employing the captioning tool to caption videos by crowdsourcing and making the captioned videos available to the students. The courses selected for the study were *Introduction to Computing* and *Computer Organization and Programming*, with 9 and 12 lectures captioned, respectively. The results are based on an online survey conducted at the end of the semester and completed by 24 students. The fluency of students in English based on the survey, which can influence the value of captions, is shown in Figure 3. An overwhelming majority of students did not have English as their first language.

A. Value of Captions

Captioning has been assessed and found to be effective in specialized scenarios such as foreign language learning [8] and deaf students in mainstream classrooms [17]. In this evaluation we focus on the added value of captions when university students are employing video as a learning resource.

Figures 4, 5, and 6 capture the response of the surveyed students on different metrics for assessing the value of videos. The results are unequivocal that students found the captions on videos to be very useful. Students believed that the captions

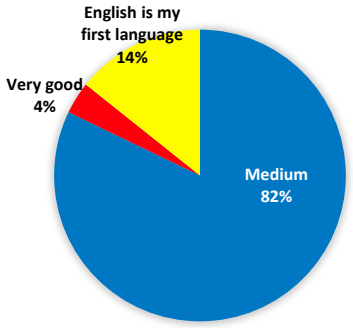

Fig. 3. Question: How would you describe your fluency with the English language?

helped them to understand the video content, found the transcripts useful for quickly browsing through the video content, and that videos with captions were preferable to videos without captions.

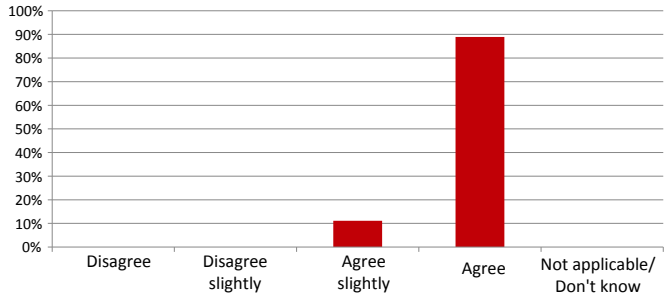

Fig. 4. Question: The captions and transcript helped me understand what the professor was saying. Please express the strength of your agreement.

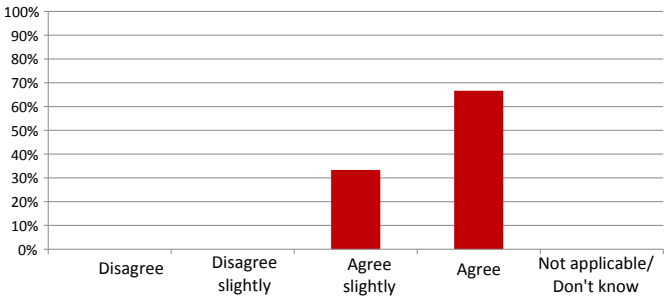

Fig. 5. Question: Do you believe that the transcript is useful to quickly read through what is discussed in the video without having to watch the entire video? Please express the strength of your agreement.

Figure 7 presents student response to questions about how different aspects of the learning experience were influenced by captions. There was strong agreement that *Efficiency*, *Note-taking*, *Attention*, and *Learning* were improved with captions. Opinion was mixed on the impact on *Quiz performance* and *Motivation*; around a third of the students answered that these aspects were also improved. There was strong agreement that captions had no impact on attendance. Finally no student answered that any aspect of learning was negatively affected by captions. Overall, the survey clearly showed a clear and meaningful enhancement of the value of lecture videos with captioning.

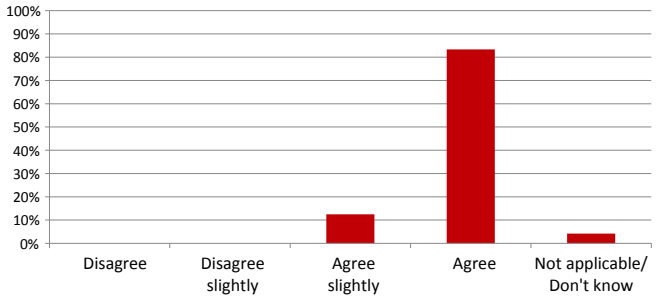

Fig. 6. Question: The videos with captions/transcript (text given for spoken sentences) are preferable than videos without them. Please express the strength of your agreement.

B. *Crowdsourced captioning study*

The ICS Videos caption editor was field tested as follows. Groups of 10-12 students were assigned to caption one lecture which runs typically between an hour and 80 mins. The participants worked independently and were given 5 days to complete the captioning assignment. The captions generated were very accurate with few errors. Some of the data from a representative experiment is presented in Table III and Table IV. Table III shows that all students participated in captioning although there was significant variation in the time they invested. The median time spent on captioning by a student was around 45 minutes. Note that captioning of such a lecture by one person can take between 6 to 10 hours. Table IV shows the rate of progress, indicating that the bulk of the captions were completed on Day 3 and Day 4. However, since captions have to be generated and possibly reviewed multiple times, it is normal that not many would be completed in the first day or two. Hence, the actual work is probably more evenly distributed. Overall, the crowdsourced caption editor enabled the students to caption their lectures without any significant problems.

TABLE III. PERFORMANCE OF INDIVIDUAL PARTICIPANTS

Users	Time taken in minutes	# of captions saved	# of captions completed
User 1	10.06	10	9
User 2	16.38	20	25
User 3	25.35	20	33
User 4	38.50	35	42
User 5	41.41	35	35
User 6	45.21	42	50
User 7	56.12	32	45
User 8	60.66	57	54
User 9	64.13	50	55
User 10	69.16	54	52
User 11	76.11	31	0

TABLE IV. PROGRESS OF WORK

Day	Day of week	# of sentences completed	% work done
1	Wednesday	0	0
2	Thursday	26	7.38
3	Friday	155	44.03
4	Saturday	134	38.06
5	Sunday	37	10.51

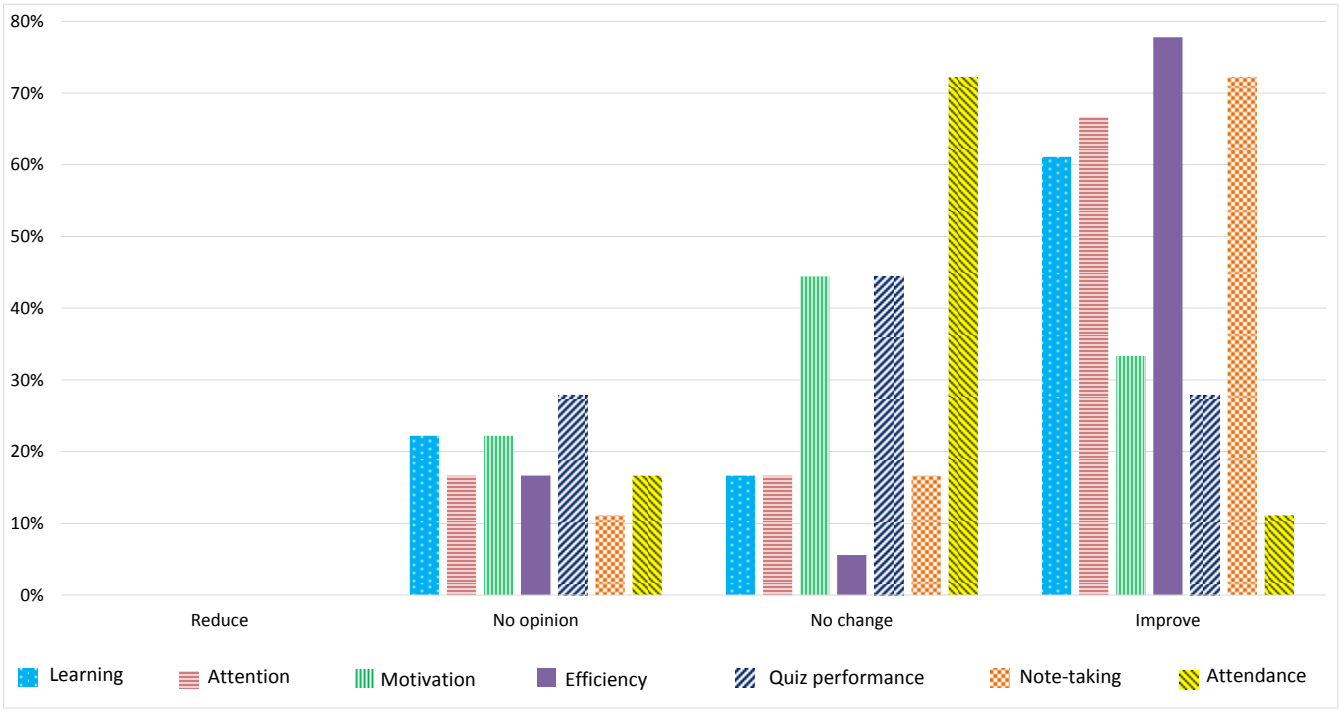

Fig. 7. Question: Please comment on the effect Captions had on your learning experience in the following aspects: Choose the option improve, no change or reduce.

C. Caption Editor surveys

A series of survey questions were included to assess the usability and effectiveness of the ICS Videos caption editor. We present the results most relevant to evaluating the usefulness of the caption editor.

Figure 8 shows that almost all students considered the captions to be accurate. This was also validated by our manual analysis that showed that the final captions generated by the students in the class were very accurate, typically reaching around 99% accuracy.

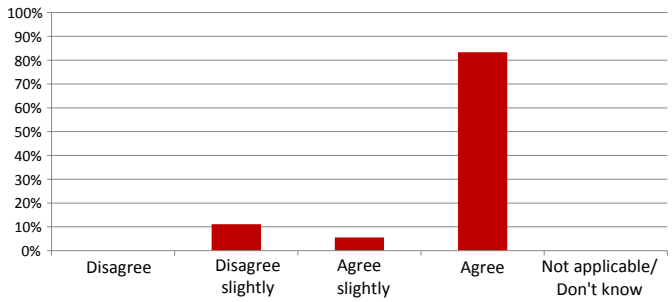

Fig. 8. Question: The captions and the transcript represented accurately what the professor said. Please express the strength of your agreement.

Results presented in Figure 9 show widespread agreement that the captioning tool was easy to use, one of our key design objectives. Figure 10 indicates broad agreement that the ability to adjust the speed of the audio with the PlaySpeed tool was useful. In instances where the audio is hard to decipher, slowing down the playback can be helpful. Figure 11 shows that almost all the students agreed that being able to mark the captions as candidates for review by another editor was useful. Our experience indicates that it is not uncommon for students to have difficulty in deciphering a word or a phrase in a technical lecture. In such scenarios, the students can seek help from others by simply marking the caption appropriately. One student commented that it was nice to be able to go directly to the problem when a caption was marked for review by other students.

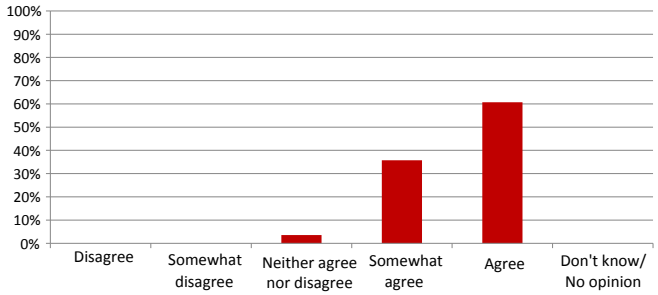

Fig. 9. Question: The ICS Caption Editor is easy to use. Please express the strength of your agreement.

Students were queried on the placement of the key elements and controls of the user interface with the results presented in Figure 12. Students overwhelmingly agreed that placements were appropriate for all features. We treat this as a validation of our user-interface design, which itself was finalized after several rounds of feedback from users.

Finally, the students were queried if they would be interested in participating in such crowdsourced captioning of lectures from their classes in the future. The response shown

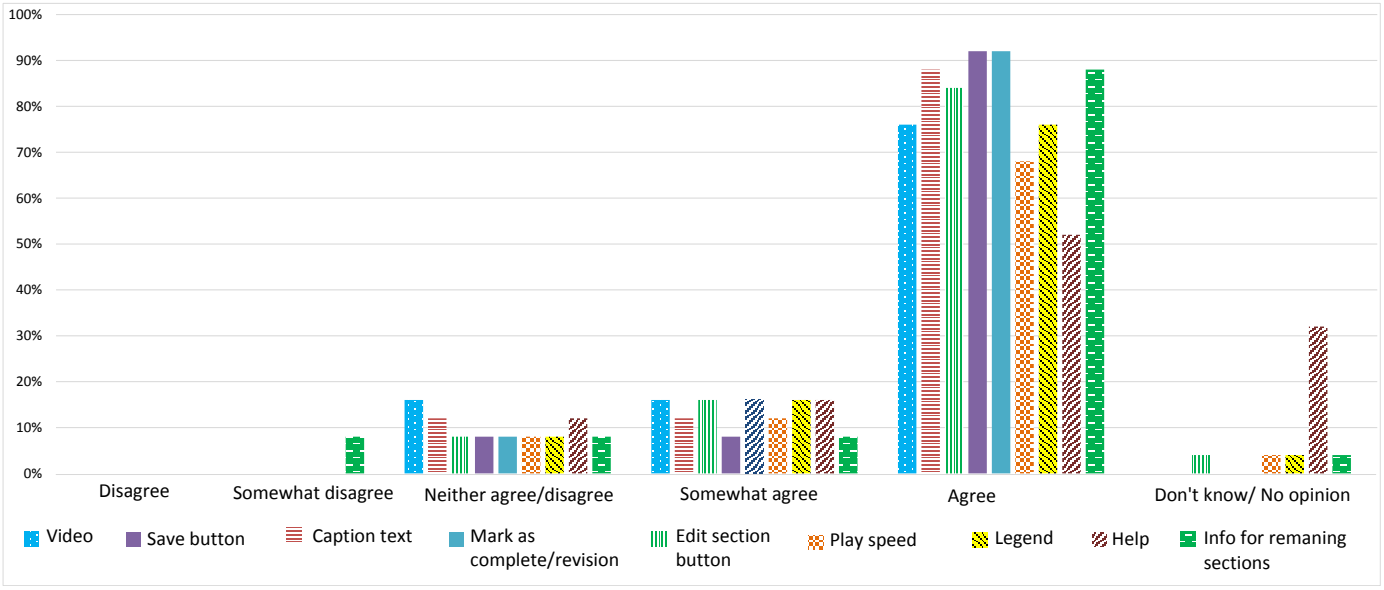

Fig. 12. Question: The placement (position) of the following elements and controls on the Caption Editor interface was appropriate. Please express the strength of your agreement.

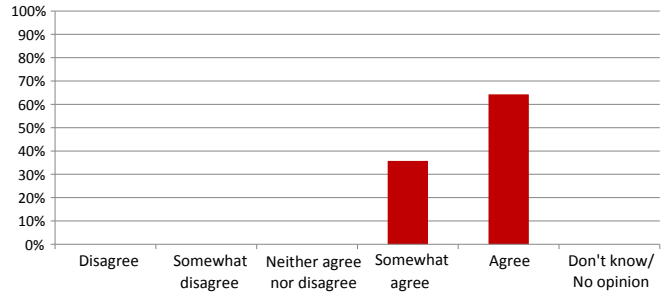

Fig. 10. Question: The PlaySpeed tool is useful. Please express the strength of your agreement. (The PlaySpeed tool can be used to adjust the speed of the audio)

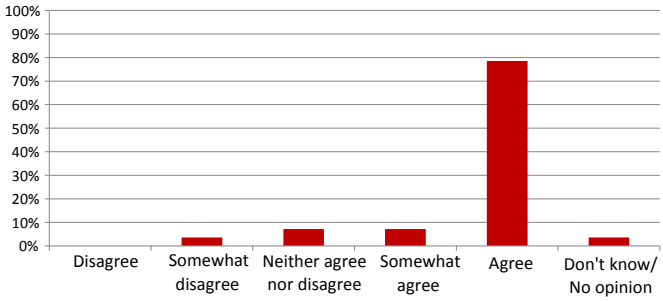

Fig. 11. Question: The feature of having the status of the caption as "Needs a Review" is useful. Please express the strength of your agreement. (If you are unable to hear the audio clearly or are unsure of the accuracy of your correction, there is an option to have the status of the caption as "Needs a Review".)

in Figure 13 was strongly positive. We consider this to be a very important result as it shows the potential of crowdsourced video captions as a practical approach to enhance the value and usability of classroom lecture videos.

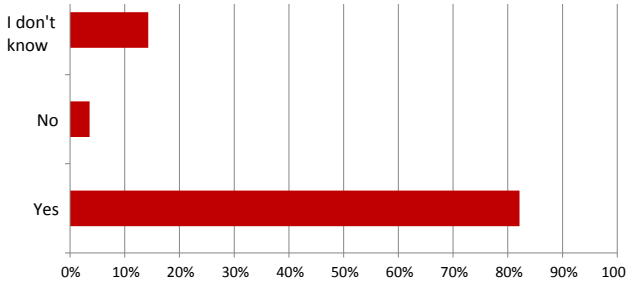

Fig. 13. Question: Would you be interested in working with other students to correct captions for your class lectures using this caption editor if you receive some incentive (for example academic credit)?

VI. Concluding Remarks

The main contribution of this paper is the design and evaluation of a semi-automatic captioning framework centered around ICS Videos crowdsourced caption editor for STEM coursework. The paper also presents results from studies on effectiveness of contemporary automatic speech recognition tools and the value of captions. We point out the key conclusions and their significance.

State-of-the-art speech recognition tools were found to have limited accuracy for technical lectures recorded live in a classroom. We speculate that many of the problems in speech recognition will be resolved with ongoing research. At the same time some challenges are likely to remain, such as those related to technical terminology and the nature of conversational speech. In some cases, speech was very difficult for even humans to decipher.

This paper adds to the growing body of evidence that captions are very valuable. Virtually all students surveyed considered the captions to be useful and valuable. We recommend that instructors consider adding captions to lectures when feasible.

The process of adding captions semi-automatically with the ICS Videos caption editor was effective and efficient. Groups of students were able to caption video lectures from their classes with a relatively low individual effort. Most students liked the design of the caption editor and found it easy to use. Most students were also willing to be part of crowdsourced captioning of their lectures in the future. The conclusion is that the approach of using students groups to caption classroom lectures collectively is promising and deserves to be investigated further and considered for deployment.

VII. Future Work

The paper presented evidence from student surveys that captioning enhances the value of video lectures and the caption editor was well designed and easy to use. However, the survey was answered by only 24 students from 2 classes. Hence the conclusions are preliminary; usage and survey by larger groups is needed to arrive at firm conclusions. Also, only 14% of students had English as their first language; usage and survey with a higher fraction of students with English as their first language is needed to understand the value of captions for native English speakers.

The caption editor is a research prototype that can be enhanced. A spell-check mechanism or auto-prediction mechanism could help reduce errors in captions and make the process of editing captions more efficient. Techniques from natural language processing can be used to partially automate the correction of captions. A mechanism to generate highly accurate captions can be developed by enabling multiple users to edit the same sentence followed by automatic validation.

An important goal of the project is enabling broad use of video lectures with captions. This can be achieved by developing a robust software that inter-operates with other video generation and presentation frameworks.

Acknowledgments

We would like to acknowledge the contributions of the members of the ICS Videos group that developed the infrastructure for the project. We thank Shishir Shah, Leigh Hollyer, Nouhad Rizk and Olin Johnson, the instructors who participated in the evaluation, along with the students in their classes. Holly Hutchins provided help in designing the evaluation questionnaire. Partial support for this work was provided by the National Science Foundation's Division of Undergraduate Education under Course, Curriculum, and Laboratory Improvement (CCLI) program with Award No. DUE-0817558. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] 3playmedia. 3playmedia: A captioning, transcription, and translation solution. [Online]. Available: http://www.3playmedia.com

[2] G. D. Abowd, "Classroom 2000: An experiment with the instrumentation of a living educational environment," *IBM Systems Journal*, vol. 38, pp. 508–530, 2000.

[3] L. Barker, C. L. Hovey, J. Subhlok, and T. Tuna, "Student perceptions of indexed, searchable videos of faculty lectures," in *Proceedings of the 2014 IEEE Frontiers in Education Conference (FIE)*, Madrid, Spain, Oct 2014.

[4] M. Broughton, "Measuring the accuracy of commercial automated speech recognition systems during conversational speech," *Virtual Conversational Characters (VCC) Workshop, Human Factors Conference*, 2002.

[5] S. P. Casali, B. H. Williges, and R. D. Dryden, "Effects of recognition accuracy and vocabulary size of a speech recognition system on task performance and user acceptance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 32, no. 2, pp. 183–196, 1990.

[6] S. Chandra, "Lecture video capture for the masses," *SIGCSE Bull.*, vol. 39, no. 3, pp. 276–280, Jun. 2007.

[7] R. Deshpande, "Captioning for classroom lecture videos," Master's thesis, University of Houston, 2013.

[8] T. J. Garza, "Evaluating the use of captioned video materials in advanced foreign language learning," *Foreign Language Annals*, vol. 24, no. 3, pp. 239–258, 1991.

[9] Google. Speech recognition research at google. [Online]. Available: http://research.google.com/pubs/SpeechProcessing.html

[10] K. S. Hone and R. Graham, "Towards a tool for the subjective assessment of speech system interfaces," *Natural Language Engineering*, vol. 6, no. 3&4, pp. 287–303, 2000.

[11] A. Misra, "Speech/nonspeech segmentation in web videos." in *INTERSPEECH*, 2012.

[12] NCH. Express scribe transcription software. [Online]. Available: http://www.nch.com.au/scribe/index.html

[13] H. Odhabi and L. Nicks-McCaleb, "Video recording lectures: Student and professor perspectives," *British Journal of Educational Technology*, vol. 42, no. 2, pp. 327–336, 2011.

[14] B. Shneiderman, "The limits of speech recognition," *Commun. ACM*, vol. 43, no. 9, pp. 63–65, Sep. 2000.

[15] A. Sobhi, R. Nagatsuma, and T. Saitoh, "Collaborative caption editing system–enhancing the quality of a captioning and editing system," in *Proc. CSUN*, 2012.

[16] S. K. A. Soong, L. K. Chan, C. Cheers, and C. Hu, "Impact of video recorded lectures among students," *The 23rd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, pp. 789–793, 2006.

[17] A. Steinfeld, "The benefit of real-time captioning in a mainstream classroom as measured by working memory," *Volta Review*, vol. 100, no. 1, pp. 29–44, 1998.

[18] J. Subhlok, O. Johnson, V. Subramaniam, R. Vilalta, and C. Yun, "Tablet pc video based hybrid coursework in computer science: Report from a pilot project," in *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education*. New York, NY, USA: ACM, 2007, pp. 74–78.

[19] Telestream. Captionmaker: A caption editing software. [Online]. Available: http://www.cpcweb.com/nle/

[20] T. Tuna, J. Subhlok, L. Barker, V. Varghese, O. Johnson, and S. Shah, "Development and evaluation of indexed captioned searchable videos for stem coursework." in *Proceedings of the 33th SIGCSE Technical Symposium on Computer Science Education*. ACM, 2012, pp. 129–134.

[21] T. Tuna, J. Subhlok, and S. Shah, "Indexing and keyword search to ease navigation in lecture videos," in *AIPR'11*, 2011, pp. 1–8.

[22] Urusoft. Software:subtitle workshop. [Online]. Available: http://subtitle-workshop.en.softonic.com

[23] M. Wald, "Crowdsourcing correction of speech recognition captioning errors," in *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*. New York, NY, USA: ACM, 2011, pp. 22:1–22:2.

[24] X. Yao, P. Bhutada, K. Georgila, K. Sagae, R. Artstein, and D. R. Traum, "Practical evaluation of speech recognizers for virtual human dialogue systems." in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. European Language Resources Association, 2010.