Wiki-like Editing of Imperfect Computer-Generated Webcast Transcripts

Cosmin Munteanu[1] Yuecheng Zhang[1] Ron Baecker[1,2] Gerald Penn[1,2]
mcosmin@cs.toronto.edu jacey.zhang@utoronto.ca rmb@kmdi.toronto.edu gpenn@cs.toronto.edu
[1]) Department of Computer Science University of Toronto Toronto, M5S 3G4, Canada
[2]) Knowledge Media Design Institute University of Toronto Toronto, M5S 2E4, Canada

ABSTRACT

As the use of Internet broadcasting (webcasting) increases, more webcasts will be archived and accessed numerous times retrospectively. One challenge in skimming and browsing through such archives is the lack of textual transcripts of the archived medias' audio channel. Ideally, transcripts would be obtainable through Automatic Speech Recognition (ASR). However, current ASR systems can only deliver, in realistic conditions, Word Error Rates (WERs) of around 45% – unsatisfactory, as shown in our recent study [1], which revealed that transcripts are useful and usable in webcast archives for WERs equal to or less than 25%. We therefore propose an extension to the ePresence webcast system that engages users to collaborate in a wiki manner on editing the imperfect transcripts obtained through ASR.

1. INTRODUCTION

Webcasts are increasingly used to deliver live information-rich media over the Internet [2] (such as on-line lectures). Most webcast media are also archived, and can be accessed by users through interactive systems such as ePresence (`http://epresence.tv/`), illustrated in Figure 1, which serves as the framework for our research. Without transcripts, humans are faced with far greater difficulty in performing tasks that are easily achieved with archives of text documents, such as retrieval, browsing, or skimming. Research evidence indicates that transcripts are the most suitable tool for performing tasks that require information seeking from webcast archives [3].

Our recent study [1] has shown that, when using a fully-featured webcast browsing tool, users' task performance and perception of difficulty was better than using no transcripts at all only for transcripts with WERs equal to or less than 25%. Unfortunately, current ASR systems cannot perform this well in domains such as open-domain lecture or conference presentation transcription. Manual transcription, on the other hand, is costly. Most lecture ASR systems achieve WERs of about 40-45% [4] (at most 20-30% in more artificial and better controlled conditions [5, 6]). Also, it is expected that such systems will not reach perfect or near-perfect accuracy in the near future [7].

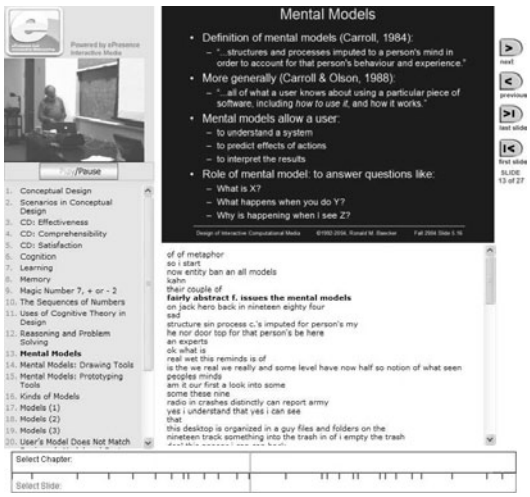

Figure 1: An ePresence webcast with transcripts

In order to achieve useful and usable transcript-enhanced webcast archives of lectures and presentations, we are proposing alternative tools to reduce current WER levels of 40-45% to the desired 25% or better. For this, we have developed a collaborative tool that extends ePresence functionality by allowing users to edit and correct, in a wiki-like manner, the webcast transcripts. The editing tool is seamlessly integrated into the regular archive viewing mode of ePresence, allowing users to make corrections "on-the-fly" while viewing an archived webcast.

2. SYSTEM DESCRIPTION

2.1 Webcast Archives with Transcripts

ePresence gives users full control of the archive, mainly through the display of the slides used in lectures and a video recording of the lectures themselves, through interaction with a table of contents (containing "chapter" headings and the title of the slides), and through a timeline (a clickable fine-grained time-progress indicator). To this interface, we have added transcripts of the webcast, obtained through ASR. The lines were time-synchronized with the video, by boldfacing the current line of the transcript, thus emulating a closed-captioned system, while fully displaying the transcript of the segment of lecture for the current slide. Transcript lines correspond to pauses longer than 200ms. Users can re-synchronize the playback of the video

CSCW '06, November 4–8, 2006, Banff, Alberta, Canada

by clicking on a line in the transcript. Figure 1 shows a screen capture of the system, with transcripts of 45% WER.

2.2 Wiki-like Editing of Transcripts

Current ASR systems deliver transcripts of webcast lectures and presentations of 40-45% WER, while the necessary WER threshold is 25%. The collaborative editing tool that we developed for ePresence allows users to correct and edit the transcripts. It extends the basic functionality of the system without burdening the user at the same time.

During regular playback of a webcast archive, users can right-click on any transcript line (not necessarily the one currently being played back), and an edit box (Figure 2) is displayed, allowing users to make corrections to the selected line. This line becomes highlighted in red, which potentially differentiates it from the current line, which is bold-faced. Besides colour-highlighting, the edit box is popped up on the screen about two transcript lines above the selected line, to maintain a visual connection with the transcript context.

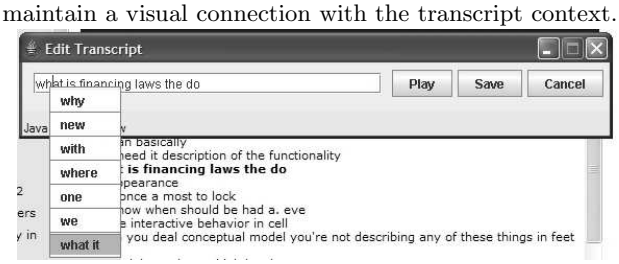

Figure 2: Wiki-like editing of imperfect transcripts

To avoid editing conflicts, a server-side locking mechanism prevents users from simultaneously editing the same line. When trying to edit a locked line, users are informed that the line is being edited by a different user, and that a browser refresh might be needed to update the transcript (webcasts need accurate time synchronization between all components, so regularly checking for transcript updates is not possible).

This on-the-fly editing mode has the advantage of being light-weight on the users – the tool is "invisible" unless explicitly invoked — while at the same time allowing users to carry out corrections to the transcripts without explicitly loading a different interface (the webcast playback is resumed automatically after the edit pop-up is closed).

2.3 Features of the Transcript Edit Tool

Edit area: users can freely make corrections to the transcript line displayed in the edit box.

Suggestion drop-down: when right-clicking on words in the edit box, a list of possible replacement words is displayed. These are choices under consideration by the ASR system during the recognition process, and extracted from the word lattices produced by the ASR system – only words that overlap by more than 70% in time alignment with the original word in the lattice are considered as alternatives.

Play button: plays the audio recording corresponding to the selected transcript line, extracted off-line from the original recording (before processing and compression of the streaming video) to ensure optimum quality.

Save: both the transcripts in the webcast window and the originals stored on the webcast server are instantly updated.

Other collaborative features: users can verify the amount of editing work they carried out, quantified as the number of word-level edit actions, viz. deletions, insertions, and substitutions. Also, editing access can be restricted to certain users up to the level of transcripts corresponding to certain slides, which is useful for defining a collaboration model of students lecture transcript editing.

3. DISCUSSIONS AND FUTURE WORK

We are currently conducting a (pilot) user study aimed at quantifying the WER reductions brought about directly by such corrections, as well as using these corrections as a source of ASR re-training and fine-tuning that will further improve the quality of the transcripts. Our study is also investigating how this approach can be applied to the domain of lecture transcript correction.

During the Fall 2006 Term, we will extend the scope of the study to address research questions related to wiki-like collaboration, such as how best to motivate students to correct transcripts. Based on the positive feedback received from students watching lecture archives during the Summer 2006 Term, we will evaluate a reward system where users gain access to the archives by editing transcripts.

4. CONCLUSIONS

The usefulness and usability of webcast archives can be significantly improved by the integration of text transcripts. Unfortunately, manual transcription is expensive, while ASR systems yield error rates of 40-45%, below the 25% threshold of usability and usefulness determined in [1]. As a solution to bridging the WER gap, we have developed a collaborative tool that extends the basic functionality of a transcript-enhanced webcast system by engaging users to collaborate in transcript editing and correction for webcast lectures and presentations. This tool seamlessly integrates with the webcast interface and allows for on-the-fly corrections during normal viewing of the archived webcast.

5. ACKNOWLEDGEMENTS

This research was funded by the NSERC Canada Network for Effective Collaboration Technologies through Advanced Research (NECTAR).

6. REFERENCES

[1] C. Munteanu, R. Baecker, G. Penn, E. Toms, and D. James, "The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives," in *Proc. of CHI*, 2006.

[2] P. Ritter, "The business case for on-demand rich media," Wainhouse Research Whitepapers, 2004.

[3] C. Dufour, E. G. Toms, J. Lewis, and R. M. Baecker, "User strategies for handling information tasks in webcasts," in *Proc. of CHI*, 2005.

[4] E. Leeuwis, M. Federico, and M. Cettolo, "Language modeling and transcription of the TED Corpus lectures," in *Proc. of the IEEE ICASSP*, 2003.

[5] I. Rogina and T. Schaaf, "Lecture and presentation tracking in an intelligent meeting room," in *Proc. of IEEE ICMI*, 2000.

[6] K. Kato, H. Nanjo, and T. Kawahara, "Automatic transcription of lecture speech using topic-independent language modeling," in *Proc. of ICSLP*, 2000.

[7] S. Whittaker and J. Hirschberg, "Look or listen: Discovering effective techniques for accessing speech data," in *Proc. of the Human-Computer Interaction Conference*. 2003, Springer-Verlag.