Collaborative Editing for Improved Usefulness and Usability of Transcript-Enhanced Webcasts

Cosmin Munteanu[1] mcosmin@cs.toronto.edu
Ron Baecker[1,2] rmb@kmdi.toronto.edu
Gerald Penn[1,2] 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
One challenge in facilitating skimming or browsing through archives of on-line recordings of webcast lectures is the lack of text transcripts of the recorded lecture. Ideally, transcripts would be obtainable through Automatic Speech Recognition (ASR). However, current ASR systems can only deliver, in realistic lecture conditions, a Word Error Rate of around 45% – above the accepted threshold of 25%. In this paper, we present the iterative design of a webcast extension that engages users to collaborate in a wiki-like manner on editing the ASR-produced imperfect transcripts, and show that this is a feasible solution for improving the quality of lecture transcripts. We also present the findings of a field study carried out in a real lecture environment investigating how students use and edit the transcripts.

Author Keywords
Webcasting, automatic speech recognition, text transcripts, navigational tools, wiki, field study

ACM Classification Keywords
H5.1 Multimedia Information Systems, H5.2 User interfaces

INTRODUCTION
Humans have long relied on written text to share knowledge. Nowadays more lectures, presentations, and talks are made available on-line. As webcasts are becoming a common mean of broadcasting such events live over the Internet, more of these media are being archived and accessed by users through interactive systems such as the one illustrated in Figure 1.

In the absence of transcripts, humans are faced with increased difficulty in performing tasks that are easily achieved with text documents. For example, a user must listen to or watch a long recording in order to locate a specific passage, instead of quickly skimming through the content of a text document looking for visual landmarks and textual cues. This represents an important hurdle in making webcast archives the digital equivalent (from a user's perspective) of libraries.

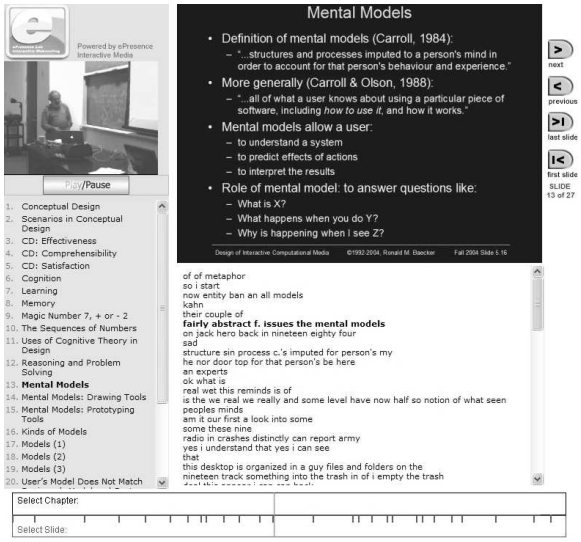

Figure 1. Our transcript-enhanced ePresence system.

Various methods propose improved access to speech recordings by manipulating the audio playback [2, 21] or to webcast archives through a table of contents [3, 23] or timeline [3], although such methods have certain limitations. However, user studies [5] suggest that transcripts are a much-needed tool for carrying out complex tasks that require information-seeking from webcast archives. Moreover, providing access to users with hearing impairments [25] only makes for an even stronger case in favour of offering text transcripts along audio/video media in archives of webcasts. Obtaining transcriptions for online media would also improve the way human users search for, organize, and retrieve specific information from large collections [11].

As manually transcribing the increasing amount of available on-line recordings is an expensive solution, such task would ideally be accomplished by an automatic speech recognition (ASR) system. However, due to adverse acoustic and linguistic characteristics (large vocabulary, speaker independent, continuous speech, imperfect recording conditions), currently-available ASR systems do not perform satisfactorily in domains such as lectures or conference presentations.

CHI 2008, April 5 - 10, 2008, Florence, Italy.

Our recent study [15] 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 Word Error Rates (WERs) equal to or less than 25%. This was determined by assessing users performance in a question-answering task, their perception of transcripts' quality, as well as users' confidence in their performance and their perceived level of task difficulty. Our study also suggested that for most browsing scenarios, users prefer having transcripts even if their quality is less than optimal.

Unfortunately, most recognition systems achieve WERs of about 40-45% in the acoustically and linguistically challenging context of lecture recordings [13, 16] (some reports suggest a 20-30% WER for lectures given in more artificial and better controlled conditions [20]). Moreover, it is expected that such systems will not reach perfect or near-perfect accuracy in the near future [29].

In order to achieve useful and usable transcript-enhanced webcast archives of lectures and presentations, we are proposing and evaluating an alternative tool to reduce current Word Error Rate (WER) levels of 40-45% to the desired 25% or better. For this, we have developed a collaborative tool that extends our webcast system's 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 our webcast system, allowing users to make corrections "on-the-fly" while viewing an archived webcast. In this paper, we present this tool and evaluate it in a field study, showing that it provides a feasible solution for improving the quality of webcast lecture transcripts.

RELATED RESEARCH

The most commonly used measure for the quality of a speech recognition system is the word error rate (WER) – the edit distance (number of substitutions, deletions, and insertions) between an automatically transcribed sentence and its manual version. Current state-of-the-art ASR systems can deliver WERs of less than 10% in ideal conditions (e.g. anechoic room, read text, with proper intonation) or under significant restrictions (limited vocabulary, ASR system trained on same speaker and under same conditions, etc.) However, the conditions for webcast recordings of lectures and presentations are in stark contrast with the ideal conditions. The archives consist of diverse speakers (with particular speech styles and various accents, including non-native), various acoustical conditions (regular lecture or meeting rooms), and the vocabulary is extremely large (determined by the large pool of topics). As shown in [27], the recognition accuracy can be degraded by a factor as large as 1.5 for each external condition that becomes non-optimal for speech recognition.

Due to the adverse conditions characterizing lecture speech, typical WER for lecture speech can reach rates as high as 50% when general-purpose ASR systems are used [16, 17]. Although significant research efforts have been recently spent on improving the performance of ASR systems for the domain of lecture transcription [13, 16], WERs are still greater than 40% under unsupervised (un-controlled) training conditions. Further reductions of WER are only achieved by controlling the ASR training conditions [7–9], such as using large quantities of manual transcriptions of the same speaker or of lectures on the same topic.

For certain automated applications (e.g. the ATIS travel reservation system [26]), a lower WER might not affect the system's performance, as long as keywords are recognized accurately. However, when transcripts are to be used directly by humans, the overall quality of the text is more crucial. Unfortunately, the research that investigates how humans deal with error-ridden transcriptions is scarce. Among the few studies, one assessing human ability to use transcripts [22] for news recording retrieval and summarization revealed that users performed better on several measures when transcripts' accuracy was higher. A follow-up study in the context of skimming through voicemail messages [28] showed that users performed tasks faster when browsing speech and text simultaneously, although performances were lower for improperly transcribed phone numbers and names. However, users' performance can be improved by providing additional information-mining tools [29].

As it is generally accepted that ASR systems are not likely to improve significantly in the near future [29], alternative solutions are needed to reduce the gap between currently-achievable WERs and acceptable WERs for webcasts. A readily available, although expensive, solution is human intervention. Unfortunately, to our knowledge, no research exists that address the cost of this approach for reducing the WER of transcripts. However, in various other scientific areas, computer-supported collaboration has emerged as an alternative. For example, it was shown in two separate studies [1, 24] that the task of indexing and labelling a large collection of images for query-based retrieval can be carried out using web-based collaboration. Collaboration has also been successfully applied to various other tasks, from controlling a mechanical robot over the Internet [10] to open source software development [4] and to geographic information mapping [14]. Recent years have also witnessed an increase in online collaborative writing, mainly in the form of wikis, from large-scale encyclopedias (`http://wikipedia.org`) to classroom projects [6].

In our research, we propose to enhance the quality of webcast lecture transcripts by facilitating, in a wiki manner, the collaboration between users (mainly students attending such lectures) in correcting the ASR-produced errors in the lecture transcripts. For this, we show how transcripts can be integrated into webcast archives and present the wiki editing tool. We then describe the evaluation of the editing tool through a field study, followed by an analysis of the collected data, and the re-design and re-evaluation of the tool based on the findings of the field study. We conclude the paper with a discussion of our proposed solution and with recommendations for future work.

ENHANCING WEBCASTS WITH TRANSCRIPTS

The ePresence (`http://ePresence.tv`) webcast system 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 (TOC – containing "chapter" headings and the title of the slides), and through a timeline (a clickable fine-grained time-progress indicator), as illustrated in Figure 1.

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

MANAGING IMPERFECT TRANSCRIPTS

Current ASR systems deliver transcripts of webcast lectures and presentations of 40-45% WER, while the necessary WER threshold is 25% (as shown in [15]). The collaborative editing tool that we developed for our webcast interface 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.

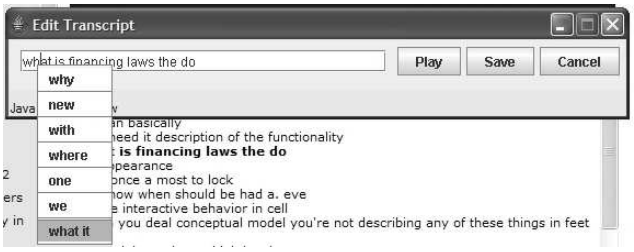

Figure 2. Wiki-like editing of imperfect transcripts

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).

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.

A FIELD STUDY

A field study was carried out to assess the feasibility of the proposed tool and to obtain further insight into how users accept and manage lectures with imperfect transcripts. In this section we describe the setup of this in-situ evaluation, while the following section will present the results and recommendations arising from the field study.

Research Questions

As we propose a collaborative tool that engages users to edit and correct imperfect transcripts of webcast lectures, several specific research questions need to be answered. For this, we devised and conducted a field study whose main objectives were to assess:

The feasibility of the interface for wiki-editing of webcast transcripts as a solution for *completing the task* of improving the quality of computer-generated transcripts.

User experience when using the transcript-enhanced webcast system and the editing tool, encompassing several components: *users' acceptance* of such interface, *transcript quality*'s influence on user experience, the *attitude* towards using it in a real lecture setting, and other *indirect benefits* gained by users (such as better lecture comprehension). Moreover, users' *confidence in using the system* was also measured[1].

[1]Since our previous laboratory-based experiment [15] also investigated users' acceptance of machine-generated transcripts as an enhancement of webcast systems, a similar assessment was conducted for the present field study in a real-life setting.

Appropriate motivational schemes for increasing users (students') *involvement* and *effectiveness* in wiki-editing of webcast lecture transcripts.

General user feedback on the interface design elements that can be improved in order to maximize the benefits of transcript editing (such as to improve the amount of edits that can be done in a short period of time).

METHODS

The field study was carried out in the context of a real classroom, over a 13-week semester consisting of 21 lectures (each approximately one hour long). The lectures are part of the same course (third year Computer Science course). Although the recordings took place in-classroom, all measures were taken such that the recording did not interfere with the regular lecture proceeding. 26 students were enrolled in the course, while typical classroom attendance was approximately 15 students every week. The lecture recordings were made available online (using our webcast system) within a day of the lecture date.

System

Our webcast interface is entirely web-based, and the recordings can be accessed through most browsers (Internet Explorer and Mozilla Firefox) on several platforms (Windows, Linux, Mac OS) without installing additional software (beside the Real Player plugin, which is freely available for a variety of browsers and platforms).

For every recorded lecture, the webcast system gives users full control of the archive, mainly through the display of the slides used in lectures and the video recording, through interaction with the TOC (at the left of the screen, which contains "chapter" headings and the title of the slides), and through the timeline (an interactive, clickable, fine-grained time-progress indicator).

Textual transcripts of the recording are displayed below the slide. The lines of text are time-synchronized with the video, by boldfacing the current line of the transcript, thus emulating a closed captioning system, while fully displaying the transcript of the segment of lecture for the current slide. The line breaks do not represent ends of sentences, but rather correspond to pauses longer than 200ms. To further enhance the user's control over the lecture, users can re-synchronize the playback of the video by clicking on a line in the transcript. The transcripts were editable, as described in the *Managing Imperfect Transcripts* Section.

Transcripts were obtained using the SONIC ASR toolkit [18]. The lecturer is male, late 30s, native (but accented) speaker of English. Due to the high speaking rate, accent, and speaking style, the WER was between 50 and 60%.

Task and Procedures

Participants were provided access to the web-based interface for accessing webcast archives. No restrictions were imposed on how the participants watched the recorded lectures. They were encouraged to make use of these as an additional course material (and as thus, it was linked from the course website). With respect to editing the lecture transcripts users were also left the choice of which lectures and which parts of lecture to correct. Since the purpose of the study was to investigate the use of the wiki editing tool in a real situation, no further requirements (beside watching lectures and editing transcripts) were formulated during the course of the in-situ evaluation. All participants were required to complete a questionnaire and brief interview after the course was completed.

The first 9 of the total of 21 lectures were freely accessible to all students in the class, while the rest were available only to the participants of the field study. Transcript editing was restricted to users registered as participants in the study. The remaining 12 lectures could be accessed ("unlocked") by participants through a credit-gaining scheme: for each user the number of words edited was recorded as credits that can be exchanged for access to the restricted lectures.

In a previously-run pilot study, we have determined that students correct an average of 300 erroneous words per hour. In agreement with the course lecturer[2], the required amount of participants' involvement was set to 4 hours during the entire semester (beside normally watching the lecture recording). As such, the amount of credits needed to "unlock" access to one lecture was set to 80 words (thus allowing full access to all lectures in exchange of 4 hours of transcript editing). As an additional incentive for editing the transcripts, students received a small course grade bonus and a modest financial compensation according to the amount of transcript corrections (significant only for those dedicating more than 4 hours).

Participants

The study was conducted with 15 participants, all third-year Computer Science undergraduate students enrolled in the course that was recorded (participation in the study was not compulsory). Two participants had previous experience with the webcast system (as it is used, without transcripts, for other courses). However, due to the intuitiveness of the webcast interface controls, no training was required for the other participants (beside a brief explanation of the system's web-based controls). Moreover, all participants indicated they are familiar with various forms of Internet-based media.

Instruments and Measures

In order to answer the research questions that motivated this field study, four types of data were collected: *task completion data*, *user experience data*, *involvement and motivation data*, and *general user feedback*.

Task completion

One of the objectives of this field study is to assess the feasibility of the wiki-editing tool as a solution for improving the quality of the lecture transcripts. As such, we have collected data indicating what percentage of the lecture transcripts were corrected by users. As it will be shown in

[2]The authors were not affiliated with the course in which the study took place.

the *Results: Task completion* Section, the task completion is assessed through the percentage of edited transcript lines (sentences), as well as through the relative WER reductions (a commonly used measure of ASR accuracy).

User experience

A post-study questionnaire was used as the instrument for collecting user experience data. The questionnaire consisted of multiple-choice questions and indicated agreement/disagreement with various statements. The user experience was assessed through a series of indicators:

User acceptance. This indicator measured students' willingness to use the transcript-enhanced webcast system through statements such as "Being able to access lectures through the webcast system helps me better review the course material", "I would like to see the system used for more classes", "I didn't need the lecture archives, the slides and examples on the prof's page are enough", and "I only need to review parts of lectures occasionally, I don't need transcripts for that".

Transcript quality 's influence on user experience was assessed through the answers to several statements: "I would rather use this system without transcripts", "Having transcripts for every lecture means I don't have to attend classes anymore", "I think the quality of the transcripts was good enough for what I needed".

Attitudes toward wiki editing. Measuring task completion rates as the percentage of corrected words provides an objective assessment of the feasibility of the wiki editing solution. However, a subjective evaluation (from the users' perspective) is also needed. For this, we have collected data related to users' attitudes toward the editing tool, mainly focused on determining if students perceive the editing tool as a useful addition to the webcast system and if they are willing to use it. Several statements on the questionnaire were used for this, such as: "Being able to correct errors in the transcripts really improved access to the course material", "I think I also benefited from other users' editing of the transcripts", "I would gladly help the class by editing transcripts for lectures using the webcast system", "I would have rather payed to access perfect transcripts than do my part of the editing", and " I rather go to class and take notes than edit transcripts".

Perception of indirect benefits. Beside determining the attitudes toward using the wiki editing tool, we queried students about how they perceived the educational benefits of the editing tool (mainly motivated by the hypothesis that more exposure to the lecture material would be beneficial). For this, we have asked questions such as: "I don't think that my editing of the transcripts helped me better prepare for the course", "When editing the transcripts, I payed more attention to the lecture", and "I think editing the transcripts helped me better understand the course material".

Confidence in using the system. Similar to our previous study [15] (and to compare differences between the use of transcripts in an artificial experiment with that of a real lecture setting), participants indicated the context in which they would choose to use the transcript-enhanced webcast system. The contexts ranged from very critical to less critical: "Prepare for an examination instead of going to classes," "Prepare for an examination in addition to going to classes," "Prepare for an assignment," and "Make up for a missed class." For each context, participants could choose "Yes," "No,", or "Only if transcripts have no errors.". In addition to these, a new choice was added to the possible answers: "Only if everyone is helping correct the transcripts".

Involvement and motivation

In order to determine if the wiki editing tool is an appropriate solution for correcting the automatic transcription errors, it is important to investigate how much time students are willing to dedicate for improving the lecture transcripts, and how to better motivate them. Multiple-choice questions on the study completion questionnaire were used to collect data about the number of hours spent weekly for transcript editing, the amount of time willing to spent, and students' estimate of the amount of time others would be willing to spend.

Ideally, students would voluntarily edit the transcripts for the benefit of the entire class. However, this might not be a realistic expectation for smaller-size classrooms. Therefore, we have also investigated possible motivational schemes. Beside the edit-for-access scheme employed during the field study, we have also asked students to indicate their preference for others, such as cost increases, making transcript editing part of course requirements, and edit-for-access combined with course bonus marks for editing more than the required minimum. Preference for each of the scheme was indicated by choosing one of "Fair", "Maybe", or "Not a fair deal" options.

General user feedback

We have invited users to also provide free-form feedback (as answers to an interview-like questionnaire), suggesting as possible dimensions features of the webcast system, features of the editing tool, positive/negative impressions of the entire system, and general comments.

RESULTS

Task completion

The improvements in transcript quality through collaborative editing are measured at sentence level (through the percentage of corrected transcript lines), but also in terms of relative WER reductions. As Figure 3 shows, most lectures had a significant number of sentences corrected (for example, 16 of the 21 lectures had more than 75% of transcript lines corrected). On average, 84% of all transcript lines were edited. This resulted in an average relative WER reduction of 53%.

In order to facilitate users' editing of transcripts, no restrictions were imposed with respect to what users were allowed to type in the edit box, resulting in inconsistencies between the 12 participants (such as abbreviations, formulas, proper names, even spelling errors that do not influence the text readability – we allowed for such variations when computing the relative error reductions). Thus, determining

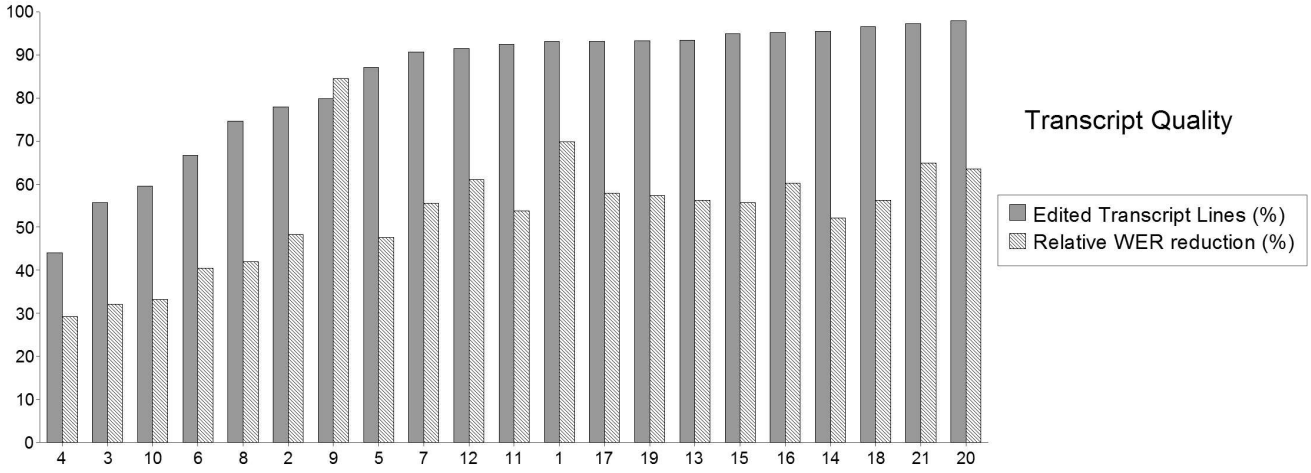

Figure 3. The percentage of edited transcript lines and relative WER reduction for each of the 21 lectures after all transcripts were corrected.

%	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1	25	58.33	16.67	0	0
Q2	63.64	27.27	9.09	0	0
Q3	0	33.33	8.33	50	8.33
Q4	0	8.33	16.67	66.67	8.33

Q1: "Being able to access lectures through the webcast system helps me better review the course material"
Q2: "I would like to see the system used for more classes"
Q3: "I only need to review parts of lectures occasionally, I don't need transcripts for that"
Q4: "I didn't need the lecture archives, the slides and examples on the prof's page are enough"

Table 1. User acceptance of the transcript-enhanced webcast system.

the WER of the corrected transcripts is not relevant. As our previous study showed that lecture transcripts become usable for WER below 25%, we can note that in the case of the present field study, this goal was achieved overall. However, while the 25% threshold was computed over uniformly-distributed imperfect sentences, the wiki editing of lecture transcripts creates a slightly uneven distribution of corrected sentences (as illustrated in Figure 3 by the non-linear variations in the percentage of corrected lines and WER reductions).

User Experience

As previously mentioned, user experience data was collected through a questionnaire administered at the end of the semester. Students indicated their level of agreement with several statements related to the indicators described in the *Instruments and Measures* Section.

User acceptance

Participants indicated their agreement with four questions relevant to their willingness to use the transcript-enhanced webcast system. As Table 1 shows, most users consider it a necessary addition to traditional lecture preparation materials.

%	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1	0	16.67	25	33.33	25
Q2	0	16.67	16.67	41.67	25
Q3	0	33.33	25	33.33	8.33

Q1: "I would rather use this system without transcripts"
Q2: "Having transcripts for every lecture means I don't have to attend classes anymore"
Q3: "I think the quality of the transcripts was good enough for what I needed"

Table 2. Users' attitudes toward imperfect transcripts.

The influence of transcript quality

Table 2 shows users' responses to questions assessing their acceptance of imperfect transcripts. While users indicate they prefer having transcripts (significant disagreement with Q1), these are shown to not provide the same experience as attending the lecture (disagreement with Q2), the quality of the transcripts continuing to pose challenges (Q3). It should be noted however that responses to Q3 could also be the result of users being exposed to transcripts of varying level of correctness, as access to lecture archives (as well as transcript editing) was not uniformly distributed over the entire semester.

Attitudes toward wiki editing

One of the areas that was the main focus of this field study is the users' attitude toward wiki editing as a viable solution for improving the quality of the transcripts. As illustrated in Table 3, students manifested a positive attitude toward the system, not only as a transcript-correction tool, but as an enhancement of the classroom experience as well.

Perception of indirect benefits

One of the hypotheses that motivated the field study was that of user editing of transcripts not only as a solution

%	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1	8.33	58.33	33.33	0	0
Q2	25	50	8.33	16.67	0
Q3	8.33	83.33	8.33	0	0
Q4	0	0	16.67	33.33	50
Q5	8.33	25	16.67	50	0

Q1: "Being able to correct errors in the transcripts really improved access to the course material"
Q2: "I think I also benefited from other users' editing of the transcripts"
Q3: "I would gladly help the class by editing transcripts for lectures using the webcast system"
Q4: "I would have rather payed to access perfect transcripts than do my part of the editing"
Q5: " I rather go to class and take notes than edit transcripts"

Table 3. Users' attitudes toward wiki editing.

%	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1	8.33	16.67	16.67	41.67	16.67
Q2	16.67	41.67	25	0	16.67
Q2	0	16.67	41.67	25	16.67

Q1: "I don't think that my editing of the transcripts helped me better prepare for the course"
Q2: "I think editing the transcripts helped me better understand the course material"
Q3: "When editing the transcripts, I payed more attention to the lecture"

Table 4. Users' perception of the indirect benefits of wiki editing of transcripts.

%	Yes	Only if transcripts have no errors	Only if everyone is helping correct the transcripts	No
Q1	25	16.67	0	58.33
Q2	75	0	8.33	16.67
Q3	41.67	8.33	0	50
Q4	83.33	8.33	0	8.33

Q1: "Would you consider using the ePresence system to prepare for an examination (instead of going to classes)"
Q2: "Would you consider using the ePresence system to prepare for an examination (in addition to going to classes)"
Q3: "Would you consider using the ePresence system to prepare an assignment"
Q4: "Would you consider using the ePresence system to make up for a missed class"

Table 5. Users' confidence in using the system as a relation of transcript quality.

for improving the accuracy of the transcripts, but as having the added benefit of providing students with more exposure to the lecture materials. Indeed, as Table 4 indicates, most users perceive this as a benefit (by disagreeing with Q1 and agreeing with Q2), although such benefits do not seem to stem from increased attention to the lecture (slight preference for disagreement with Q3).

Confidence in using the system

Similar to our study that determined the minimum WER level for usable transcripts [15], we have assessed users' overall confidence in using the system, with respect to the application where the system is to be used. The responses (presented in Table 5) are consistent with the findings of our previous study – the transcript quality is critical in more important applications. Compared to the previous study, a fourth condition was available now as an answer: "Only if everyone is helping correct the transcripts", although this does not seem to significantly influence the relation between transcript quality and application's importance.

Involvement and Motivation

Out of the total fifteen participants, two chose to contribute more than the required 4 hours of transcript editing. The combined contributions of these two participants accounted for approximately 75% of the transcript corrections. Ten more participants contributed the full required amount of editing (thus gaining access to the entire archive of lectures). Three participants contributed less than the required amount and did not respond to the questionnaires.

Users' level of involvement was also assessed through the final questionnaire. Participants indicated that they are willing to spend an average of approximately 50 minutes a week for transcript editing (two users indicating 2 hours, while other two indicating only 15 minutes). However, when asked about other students' willingness to edit transcripts, most of the twelve responses estimated 15 minutes (four responses) or no time (also four responses). Participants were also asked to estimate how much time they spent weekly editing transcript during the field study – the majority (six responses) indicated one hour, although two (likely the most active users) indicated more than two hours a week.

In terms of users' preference for the motivational scheme, 75% of respondents indicate they consider as fair crediting the transcript editing work (as the number of corrected words) and use this credit to gain access to lectures. This percentage increased to 83% for the same scheme that also allow students to gain course marks for additional editing work. As expected, increasing university costs to provide perfect transcripts was rejected by all participants, however, requiring transcript editing as part of the course requirements was accepted or somewhat accepted by nine of the twelve respondents.

General User Feedback

Given the particularities of this field study, mainly evaluating a completely new concept (wiki editing of imperfect

transcripts) in a real setting (students attending lectures during an entire semester), participants' feedback is one of the most important aspects of the data collection. Beside various comments related to the web interface (such as video player plugin), most of users' feedback was focused on the editing tool and on the integration of transcripts.

While the shortcomings of the editing tool (and the measures taken to address them) are discussed in the following section, the integration of transcripts into the webcast system generated several suggestions from users. Both our previous study [15] and the analysis of the data collected during this field study show that transcripts are regarded as a necessary addition to the webcast system (even if they are of lower quality). However, the participants indicated that the sheer amount of transcripts make them difficult to be skimmed through. This suggest a more compact representation (either as a summary or another high-level representation of the lecture content) may be more appropriate, especially for lectures that are accompanied by information-rich slides or discussing topics (such as a programming language) that are easily accessible elsewhere.

INTERFACE RE-DESIGN AND RE-EVALUATION

Assessment of Current Design

In terms of the wiki editing, the feedback provided by participants highlighted the need for a more versatile editing tool. Due to the initial recording segmentation process (by pauses of 200ms, as described in the *Enhancing Webcasts with Transcripts* Section) and given the fast speaking rate of the lecturer, some lines of transcript were spanning longer audio segments (15-20 seconds). This resulted in difficulties when correcting the transcripts, and as such, users suggested better playback control for the editing interface.

The second significant suggestion involved users' approach to correcting transcripts. Both through the open-form responses and through specific questions on the final questionnaire, it emerged that our initial model of editing "on-the-fly" is mainly applicable to occasional corrections of the transcripts. However, in the context of the large-scale editing needed to correct entire lectures, an editing tool that incorporates transcript navigation is needed (7 of the 12 participants indicated they prefer to consecutively edit several transcript lines, without switching to lecture watching). Therefore, a second editing tool was integrated with the interface that incorporate multi-line transcript editing with slide navigation and better playback control.

Extended Editing Mode

The "on-the-fly" editing tool is activated by right-clicking on transcript lines during regular webcast viewing, and does not obscure the webcast interface while active (being implemented as a small pop-up window). In contrast, the extended editing tool replaces the webcast viewing mode when activated, as it is designed for longer editing tasks. For this, users can click on the "Edit this slide" button from the slide navigation panel on the right of the webcast interface.

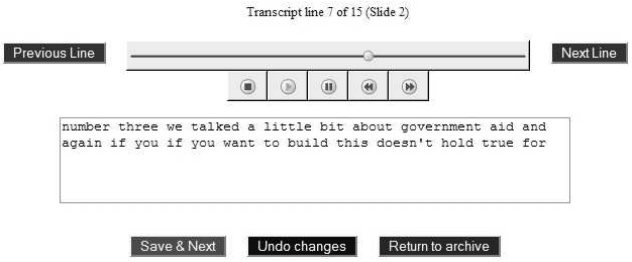

Figure 4. The extended editing mode, allowing for full control of the audio playback and for editing of consecutive transcript lines.

The extended editing mode (Figure 4) allows full editing of a transcript line, similarly to the editing box. In addition, the extended mode provides enhanced features for:

Transcript navigation: users can navigate through all transcript lines corresponding to the current slide.

Enhanced audio playback: the audio segment associated with the line of text being edited is controlled through the visually-intuitive playback panel of the Real Player plugin [19]. Users can play/pause/stop the audio segment, fast-forward through it, or use the slider for more accurate positioning. This allows for better visual synchronization between the length of text and the audio segment, as requested by participants after the first field study.

Enhanced editing: the text area is no longer limited to a single row, facilitating editing of longer transcript lines. Users are also provided with an "Undo" options. The "Save" button will instantly update the changes in the original transcripts stored on the server and, since multiple lines of text can be edited in the extended mode, will also advance to the next transcript line.

Return to archive: exits the extended editing mode and returns the user to the webcast viewing mode.

Locking mechanism: similar to the pop-up editing tool, a server-side locking mechanism prevents simultaneous corrections to the same line. For the extended editing mode the lock is engaged only when the users click inside the text edit area, avoiding unnecessary locks when users do not edit the current line. When advancing to a locked line, the edit buttons (Save and Undo) and text area are gray-shaded and inactive. Clicking on the text area of a previously-locked line or of a line that became locked after the user advanced to it activates a prompt informing the user that the line is currently edited by a different user.

Evaluation of the Re-designed System

The re-designed webcast system (enhanced with the extended editing tool) was deployed for use by students in a different lecture – a fourth year undergraduate Computer Science class. The same data was collected in this follow-up study as in the initial field study, under the same evaluation protocol. However, based on the feedback from the initial study, the motivational scheme was modified: the incentives to correct the transcript were limited to course bonus marks (instead of both marks and financial compensation).

As this follow-up study took place during the Summer term (when students typically work full-time and enroll in at most

one course) and due to the weaker incentives, enrollment in the user study was lower than in the initial study. Moreover, the overall interest in viewing the lecture archives was significantly lower, in part explained by the very informative lecture slides and the comprehensive reading package. Five students (from a class of 30) participated in the follow-up study, and four of them completed the final questionnaires.

Since the focus of this re-evaluation was to test the changes in the design of the transcript editing mode, we will discuss here the main differences in collected user experience data, in motivation data, and in general user feedback.

User Experience

With respect to the wiki editing of imperfect webcast transcript, data collected from two indicators of user experience was analyzed: *attitudes toward wiki editing* and *perception of indirect benefits*.

In terms of *attitudes toward wiki editing*, the four participants in the follow-up study responded mostly as in the initial study. Compared to the initial data (described in Table 3), a slight shift toward positive attitudes was observed for Q2 and Q3 (no "Disagree" responses for Q2 and only "Agree" responses for Q3). However, the responses for Q1 were divided, one of the four participants indicating disagreement with Q1 and the rest indicating agreement, a difference from the initial study that can be explained by the fewer overall corrections of the lecture transcripts. A more significant difference was observed for Q4 and Q5, where most responses were centered around the neutral answers (a possible consequence of the part-time nature of summer courses, consisting mainly of full-time working students).

The analysis of *perception of indirect benefits* data showed more divided responses than in the initial study. Two of the four participants indicated they saw an overall learning benefit from transcript editing. This difference can also be attributed to the higher availability of lecture preparation materials than during the initial field study.

Motivation

Although we expected students' preference for motivational scheme to be significantly different in the follow-up study (due to the different course setting), participants indicated the same preference (75%) for using the credit earned for transcript editing to gain access to lectures. Moreover, all respondents supported the same scheme, that also allows receiving of course bonus marks for additional editing.

General User Feedback

Our analysis of participants' free-form feedback and of their responses to questions of preference for editing modes indicated a positive response to the introduction of the extended editing tool. Compared to the initial study, the same preference for editing larger ranges of transcript was manifested (three of the four participants indicated a strong preference for this, while one response indicated a preference for a mixed mode). One participant explicitly commended it for allowing fast transcript corrections. No other negative comments were collected regarding the editing modes, suggesting that providing both alternatives for editing ("on-the-fly" and extended) is the appropriate solution for facilitating transcript editing.

DISCUSSION AND 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 [15]. 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. We have evaluated this tool through a field study carried out in the context of a real classroom, and have shown that this is a feasible solution for alleviating the ASR errors of webcast lecture transcripts.

The editing tool was evaluated iteratively by integrating it with the other educational resources available to the students of two Computer Science courses. We have analyzed not only the improvements in transcript quality brought by the editing tool, but also looked at how students are making use of transcripts in general and of the wiki-editing tool, and what is their attitude toward such enhancements of webcast systems. We have found that wiki-editing is well received by webcast users, and that students are willing to contribute to the improvement of lecture transcripts. Our study revealed that access to transcript editing must be facilitated by providing both the option of "on-the-fly" editing and that of mass-editing.

One of the research questions that remains open is determining the appropriate motivational method. Combining lecture access limited to wiki contributors with academic and financial incentives yielded sufficient contributions from students even in a smaller-sized class. However, weaker incentives coupled with comprehensive lecture materials (slides, readings) resulted in both significantly less contributions and reduced interest in archived lectures viewing. Our future work will look at evaluating the wiki editing concept in a larger-sized class with reduced availability of course materials.

The ePresence system is used not only for webcast lectures, but for general-interest presentations, which are archived and available through the `ePresence.tv` media portal. Many of these archives are accessed by several thousand viewers. As the large number of users can overcome the issue of reaching a critical mass of contributors, and since such communities typically offer intrinsic motivations to contribute [12], we will investigate the integration of the wiki editing interface into the ePresence portal.

Providing financial incentives for transcript editing may not always be feasible. However, research evidence suggest further WER reductions are possible in subsequent lectures when manual transcripts of earlier lectures in a series are

used to re-train an ASR system [9]. Therefore, we are beginning work on developing ASR training methods that exploit users' corrections of the early lectures in a course.

Our editing interfaces allow users to either correct the existing text or edit from scratch by deleting, through a simple text selection, the entire line. We plan to conduct an evaluation that establishes the transcript quality where one alternative would be easier for users than the other.

While several studies (as well as our previous experiments [15]) suggest the importance of transcripts for webcast systems, the current in-situ study showed that, in an educational context, a higher-level (but more comprehensive than the table of contents) text-based representation of the lecture content is also needed. Future work will look at enhancing the current interface with other, more compact, textual projections of the webcast lecture, such as automatically-generated summaries.

ACKNOWLEDGEMENTS

This research was funded by the NSERC Canada Network for Effective Collaboration Technologies through Advanced Research (NECTAR). The authors wish to also thank our undergraduate research collaborator Yuecheng Zhang for her contributions to the implementation of the editing tool.

REFERENCES

1. L. von Ahn and L. Dabbish. Labeling Images With a Computer Game. *Proc. ACM CHI*, pp. 319–326, 2004.
2. B. Arons. Speechskimmer: A System for Interactively Skimming Recorded Speech. *ACM Transactions on Computer-Human Interaction*, 4(1):3–38, 1997.
3. R. Baecker. A Principled Design for Scalable Internet Visual Communications with Rich Media, Interactivity, and StructuredArchives. *Proc. CASCON*, pp. 83–96, 2003.
4. K. Crowston, H. Annabi, J. Howson, and C. Masango. Effective Work Practices for Software Engineering: Free/Libre Open Source Software Development. *Proc. ACM WISER*, pp. 18–26, 2006.
5. C. Dufour, E. G. Toms, J. Lewis, and R. Baecker. User Strategies for Handling Information Tasks in Webcasts. *Proc. ACM CHI*, pp. 1343–1346, 2005.
6. A. Forte, and A. Bruckman. From Wikipedia to the Classroom: Exploring Online Publication and Learning. *Proc. ICLS*, pp. 182–188, 2006.
7. F. Fügen *et al.* Advances in Lecture Recognition: The ISL RT-06S Evaluation System. *Proc. Interspeech*, pp. 1229–1232, 2006.
8. S. Furui. Recent Progress in Corpus-Based Spontaneous Speech Recognition. *IEICE Transactions on Information and Systems*, 88(3):366–375, 2005.
9. J. Glass *et al.* Recent Progress in the MIT Spoken Lecture Processing Project. *Proc. Interspeech*, pp. 2553–2556, 2007
10. K. Goldberg, B. Chen, Solomon R., and S. Bui. Collaborative Teleoperation Via The Internet. *Proc. IEEE ICRA*, pp. 2019–2024, 2000.
11. A. Hauptmann *et al.* Informedia at TRECVID 2003: Analyzing and Searching Broadcast News Video. *Proc. (VIDEO) TREC*, 2003.
12. S. Kuznetsov. Motivations of Contributors to Wikipedia. *ACM Computers and Society*, 36(2), 2006.
13. E. Leeuwis, M. Federico, and M. Cettolo. Language Modeling and Transcription of the TED Corpus Lectures. *Proc. IEEE ICASSP*, pp. 232–235, 2003.
14. S. Li and D. Coleman. Results of CSCW Supported Collaborative GIS Data Production: An Internet-based Solution. *Proc. ISPRS SIPT*, pp. 1–66, 2002.
15. C. Munteanu *et al.* The Effect of Speech Recognition Accuracy Rates on the Usefulness and Usability of Webcast Archives. *Proc. ACM CHI*, pp. 493–502, 2006.
16. C. Munteanu, G. Penn, and R. Baecker. Web-Based Language Modelling for Automatic Lecture Transcription. *Proc. Interspeech*, pp. 2353–2356, 2007
17. A. Park, T. J. Hazen, and J. R. Glass. Automatic Processing of Audio Lectures for Information Retrieval: Vocabulary Selection and Language Modeling. *Proc. IEEE ICASSP*, 2005.
18. B. L. Pellom. Sonic: The University of Colorado Continuous Speech Recognizer. Technical Report #TR-CSLR-2001-01, University of Colorado, 2001.
19. RealNetworks. Introduction to Streaming Media with RealPlayer. `www.realnetworks.com/support/ /education/production.html`, 2004.
20. I. Rogina and T. Schaaf. Lecture and Presentation Tracking in an Intelligent Meeting Room. *Proc. ACM (IEEE) ICMI*, pp. 47–52, 2002.
21. N. Sawhney and C. Schmandt. Nomadic Radio: Speech & Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.
22. L. Stark, S. Whittaker, and J. Hirschberg. ASR Satisficing: The Effects of ASR Accuracy on Speech Retrieval. *Proc. ICSLP*, pp. 1069–1072, 2000.
23. E. G. Toms, C. Dufour, J. Lewis, and R. Baecker. Assessing Tools For Use With Webcasts. *Proc. ACM/IEEE JCDL*, pp. 79–88, 2005.
24. T. Volkmer, J. Smith, and A. Natsev. A Web-Based System for Collaborative Annotation of Large Image & Video Collections. *Proc. ACM MM*, pp. 892–901, 2005.
25. M. Wald, K. Bain, and S.H. Basson. Speech Recognition in University Classrooms. *Proc. ACM SIGACCESS*, pp. 192–196, 2002.
26. W. Ward and S. Issar. The CMU ATIS System. *Proc. ARPA WSLT*, pp. 249–251, 1995.
27. M. Weintraub, K. Taussig, K. Hunicke-Smith, and A. Snodgrass. Effect of Speaking Style on LVCSR Performance. *Proc. Interspeech*, pp. 16–19 (Addendum), 1996.
28. S. Whittaker *et al.* Scanmail: A Voicemail Interface that Makes Speech Browsable, Readable and Searchable. *Proc. ACM CHI*, pp. 275 – 282, 2002.
29. S. Whittaker and J. Hirschberg. Look or Listen: Discovering Effective Techniques for Accessing Speech Data. *Proc. British HCI*, pp. 253–269, 2003.