Assessing Tools for Use with Webcasts

Elaine G. Toms[1], Christine Dufour[2], Jonathan Lewis[1], Ron Baecker[3]

[1]Faculty of Management
Dalhousie University
Halifax, Nova Scotia
Canada
etoms,jlewis2{@dal.ca}

[2]EBSI
Université de Montréal
Montréal, Quebec
Canada
christine.dufour@umontreal.ca

[3]Knowledge Media Design Institute
University of Toronto
Toronto, Ontario
Canada
baecker@kmdi.utoronto.ca

ABSTRACT

This research assessed the effectiveness of selected interface tools in helping people respond to classic information tasks with webcasts. Rather than focus on a classic search/browse task to locate an appropriate webcast to view, our work takes place at the level of an individual webcast to assess interactivity within the contents of a single webcast. The questions guiding our work are: 1) Which tool(s) are the most effective in achieving the best response? 2) How do users use those tools for task completion? In this study, 16 participants responded to a standard set of information tasks using ePresence, a webcasting system that handles both live and stored video, and provides multiple techniques for accessing content. Using questionnaires, screen capture and interviews, we evaluated the interaction, assessed the tools, and based on our results, make suggestions for improving access to the content of stored webcasts.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**] Web-based interaction.

General Terms

Performance, Design, Human Factors.

Keywords

webcasting systems, interface tools, information tasks, evaluation, ePresence

1. INTRODUCTION

Webcasts were invented for live, real-time-viewing-only Internet-streaming applications. Today that is no longer the case as many applications exist for which the stored webcasts from these sessions are reused such as distance education, libraries of expert presentations, conferences, training events, and archives of historic events. The systems that stream an event are often the same ones used to review the archived event, and we question the appropriateness of this implementation without additional attention to the new tasks. Re-visiting an event or reviewing the record of an event are often done for different purposes than the tasks associated with attending an event. For example, one might wish to extract a direct quote from the presenter, examine parts of the content for informational purposes, or assess the value of the work.

In this research, we examined how typical users interact with stored webcasts to complete a set of information tasks. This study was exploratory, but comprehensive and holistic. Using a mixed method within-subjects design, we assessed how certain interface tools are used to help with certain types of tasks. We wondered if some tools were better suited to certain tasks performed while examining and scanning a stored webcast. In the context of this assessment we examined both procedural and cognitive processes of users. ePresence, an existing open source webcasting system, was used for this purpose.

2. PREVIOUS WORK

A webcast is the Internet audio and/or video stream produced from a live event, or an online simulcast of a broadcast signal [18]. A webcasting system can be classed as a form of multi-media system, and a webcast, thus, is a multi-media object with multiple components. In addition to the video, the webcast usually includes the slides from a presentation, and may include other artifacts. Much of the research that deals with this area has been in the digital video arena, mainly concerning technological issues such as video indexing, capture, compression and storage [9, 14].

More recently, the focus has shifted toward the interface to digital videos and user interactivity with digital video [14]. Since browsing of a video has been thought to be time consuming, much of the effort to augment and improve content-based navigation of digital video collections have centred around surrogates of video objects. Video surrogates are defined as "compact representations of the original video that shares major attributes with the object it represents" [21]. Surrogates are classified by the medium in which they are presented: text, still image, moving image, audio, and multimodal surrogates. To date, several types of surrogates have been assessed: textual surrogates (e.g. [8, 13, 4]), visual surrogates (such as storyboards [6 13, 2]) and keyframes [5, 6, 13]. The evaluation of the surrogates has been done using different task types: selection tasks [20], recognition tasks (object and/or action) (see, for example: [11, 20, 21]), fact-finding tasks [2, 19], and summarization tasks (visual and/or textual) [3, 20, 21]. Audio surrogates, however, have received less attention [20]. Much of this current research has emerged from the work with several digital video repositories such as the Fischlàr Digital Library [9, 14], the Open Video Project [5, 12, 17], and the Informedia project [2, 7, 17].

JCDL'05, June 7-11, 2005, Denver, Colorado, USA.

Video systems use different types of content: informational audio-centric (classroom lectures and conference presentations), informational video-centric (sports, news and travelogues) and narrative-entertainment (news and television dramas) [10]. Li et al. [10] showed that different surrogates are needed for the different types of video content. As few studies looked at the particular context of classroom lectures and conference presentations, additional research is needed to gain a better understanding with this form of content. But different surrogates are needed for different purposes. In particular informational versus entertainment have differing needs [10].

Research on the usefulness of webcasting systems has been much less developed. The existing webcasting systems are mostly commercial systems that have tended not to publish research associated with their development. The research that has been conducted involves three systems – ePresence [1], Berkeley Internet Broadcasting System (BIBS) [12], and an experimental system from Microsoft [8]. These looked primarily at design issues for tools to support the use of stored or live webcasts.

Much of the current research has focused on surrogates as key access points to digital video archives rather than the interactions of people with the digital video objects [15] and their subsequent use of that video. Although attempts are being made to understand and improve access strategies to video repositories, little work has been done to understand how users interact with extended linear content-based video objects. In this research we examine those user interactions in the context of webcasts.

Figure 1. ePresence Webcasting System – 2004 replay interface

3. METHODS

3.1 System Used

ePresence (http:/epresence.tv) is a scalable, flexible, cross-platform, open source webcasting system that also enables the replaying of stored webcasts. In addition to its live view interface, it contains a separate interface for accessing the stored webcasts. The latter interface is illustrated in Figure 1. This interface has several features to aid the user in accessing the content of the system: (1) Powerpoint slides created by the presenter for use in the presentation; (2) a table of contents created in real-time during the presentation that identifies in a metaphoric like way the 'chapters' in the presentation; (3) a timeline view of the presentation identifying the chapters divisions and slides by time; (4) a video window that displays the moving image and controls the audio; (5) search option for searching the text.

Some tools are synchronized with others. By selecting a title from the table of contents, the video commences to play at that point in the presentation. The slides load as the video is played but, the user also has the option of using the slide controls to navigate forward and backward through the slides while the video is playing. The timeline has three elements along its continuum: a large vertical line denoting the chapter divisions, smaller lines denoting each slide, and a marker to indicate the point in the video.

3.2 Tasks

To understand user behaviour in the context of webcast use, three tasks based on those defined by Whittaker and colleagues [19] were adapted for the study. Other taxonomies exist, but are too focused on traditional documentary video, and not applicable to webcasts. The tasks used were:

Task-A. *Selection*: Quickly select videos related to a specific topic;
Task-B. *Specific Questions*: Answer three specific factual questions on a video;
Task-C. *Gist*: summarize the main theme(s) or gist of a video.

These tasks were operationalized as three scenarios:

Task-A Scenario: *You have found a few digital videos that are potentially useful for a term paper on "Novel interfaces for knowledge management," and you want to quickly determine which one or ones are relevant enough to merit taking the time to watch the video in detail… indicate why you find each relevant (or not relevant) for your term paper.*

Task-B Scenario: *Find the answers to the three questions using the video X. You have 10 minutes to find all three. a) Why was the Institute for Liberal Arts and Interdisciplinary Studies developed? b) Why is the flying pig significant? c) How many students lived in dormitories?*

Task-C Scenario: *Summarize, for a professor or fellow-student, the main theme(s) or gist of the talk X.*

Time was restricted for each task: 15 minutes for the A and C, and 10 minutes for B. For each task, appropriate video(s) were selected from the Knowledge Media Design Institute (University of Toronto)'s webcast archives (http://www.kmdi.utoronto.ca). The webcasts were selected for their interest to the targeted audience, their delivery by well-known scholars, and their potential applicability for one of the tasks. Each of the webcasts selected was between 40 and 45 minute in length and had previously been delivered as a public lecture, aimed at a heterogeneous audience. Each included a set of slides that illustrated the talk.

3.3 Participants

Sixteen participants were recruited from the University of Toronto. The participants were adults (63% under 30 years old) who were mainly students (88%) with an undergraduate degree (31%) or a master degree (63%). The sixteen had diverse

backgrounds in the humanities, social and natural sciences, but were currently working in areas related to the topics of the webcasts. Nine of the sixteen had never viewed a live or stored ePresence webcast.

3.4 Procedures

The study was conducted using a standard desktop computer with 17 inch monitor. Two applications were simultaneously running in separate Internet Explorer windows: the ePresence webcasting system in one, and a custom-designed research system (modeled after WiIRE [16]) that handled data collection for this study in the second. The latter contained a series of conditional HTML forms that supported data collection, handled the flow of the study, and collected participants' answers to the task questions. A participant maneuvered between the two applications using the icons on the task bar at the bottom of the screen.

Each session lasted about 2 hours and followed these steps:

1. Participants, after consenting to participate, completed a demographic and digital video experience questionnaire,;

2. Participants were given practice time with ePresence using a tutorial webcast, and when they were personally comfortable with the system, the test commenced;

3. For each of the three tasks which were assigned in random order, participants:
   a. responded to questions about their knowledge of the video(s) used and of the topic covered;
   b. worked on the assigned task using ePresence for 10 to 15 minutes; answers to the tasks were typed into open-ended textboxes on the forms in the research system;
   c. on task completion,
      i. responded to questions regarding their perception of the process and the level of accomplishment attained in doing the task;
      ii. while the screen capture of that task was replayed, responded to the following questions using a 'talk-after' semi-structured style:
      What did you understand this task to mean? You started with [timeline, slides, etc.]. Why? What did you think it would achieve (or do for you)? Why did you change from [timeline] to [slides, etc.]? [If off track:] When did you know you were off track? What helped to get you back on track? How could it have been avoided in the first place? What was your biggest challenge/issue/difficulty at this point? How do you navigate through a video? What is your strategy or overall approach? What would simplify this task? How did you decide which video was the more useful? [used only for Task #1] What was the most useful in helping with this task? What else would have helped you to do this task more efficiently?

4. After all tasks were completed, participants were interviewed for more general information concerning their use and assessment of the ePresence system as well as their experience with digital video. Some of the questions in this more structured interview include: *How effective were each of the ePresence tools in helping with the assigned tasks? (use of the soundtrack?) What was the most useful feature? What was the least useful feature? Why? What was your biggest challenge in using the system? Do you have any suggestions on how to improve ePresence access to the stored video? What would you like to see added to ePresence?*

3.5 Data Analysis

The data were collected using three mechanisms: (1) a Web-based database that captured responses to all questionnaires as well as answers to the tasks, (2) digital audio files for all interviews, and (3) video screen-capture software to record the participant's interaction with ePresence while doing the tasks.

With these data, four types of analysis were performed:

a) Responses to questions in the demographic/experience, and pre- and post- task questionnaires were coded and analyzed using primarily SPSS' General Linear Model.

b) Participants' responses to the questions were evaluated for correctness. An expert (one of the researchers) evaluated the accuracy of each of the responses and assigned a grade per question. These grades were normalized across all questions and an overall percentage was assigned each task.

c) Each of the screen capture files recorded the process used by participants. A transaction log file was recreated from these video. To extract the steps each participant took, each screen capture video was analyzed using *HyperResearch* video editing software. Each user action was time-stamped. A pre-defined taxonomy of possible user actions was created at the outset, but modified as the analysis took place to accommodate unpredictable actions. By completion, 50 actions had been noted from "minimize window" to "pause video."

d) The decision making process that participants used, including selected cognitive and emotive responses were extracted from the interviews that took place during the 'talk after' at the end of each task, and at the end of each participant session. Interviews were professionally transcribed, and coded using *Qualrus* qualitative analysis software. Interviews were segmented primarily by question asked in the interview, and coded for response to each question. Although as is the nature of interviews, the talk may stray away from the questions; all user utterances were coded.

4. RESULTS

4.1 Introduction

Results are presented first by task and then by tool. Each task represents a type of use – a purpose – for reviewing a webcast, and thus each is examined individually. While one may aggregate across all tasks for general assessment of ePresence, we were interested in understanding whether the tools and which ones best served the needs of each task.

4.2 Tasks

4.2.1 User Perception and Understanding of the Task

From the pre-task questions, participants indicated level of familiarity with the assigned task and the video(s) used in that task. They were unfamiliar with the webcasts used in this study, having not previously viewed them, or participated in the live session. With respect to each of the topics of the individual webcasts, they tended to be 'somewhat' (middle point on a five-point scale) familiar with the topic of the webcasts. They perceived that the first two tasks were somewhat difficult, but considered Task-C to be closer to 'extremely' difficult.

Hence, previous experience with viewing the webcast was not a potentially confounding variable. However, they did not perceive they had significant expertise in the topics of the tasks, despite being enrolled in degree programs for which the themes covered by the webcasts used in this study are central.

In their response to questions after the task was completed, participants indicated how they interpreted the task. Task-A was interpreted as determining which of the three videos were relevant to the assigned topic. This, however, was identified in two ways: a) was either of the webcasts relevant to the topics? And, b) was anything in either of the videos relevant to the topic?

For Task-B, participants assumed they were finding answers to specific questions using the assigned webcast in the allotted time. Participants compared this task with metaphors like "finding gems," or "treasure hunting."

For Task-C, participants assumed they were summarizing a lecture, which one described as "skimming the lecture to find the main themes."

Thus in general participants had interpreted the task as it was conceived by the researchers, and additionally indicated that these tasks were not outside the norm of tasks that they had done in the past in other information use environments.

4.2.2 User Perception of Task Completion

After each task was completed, participants indicated their perception of doing the task along four dimensions as illustrated in Figure 2. Participants, on average, were not confident about how correctly they completed the task. Overall, they found the tasks difficult, more frustrating than pleasing, and had insufficient time to do the task.

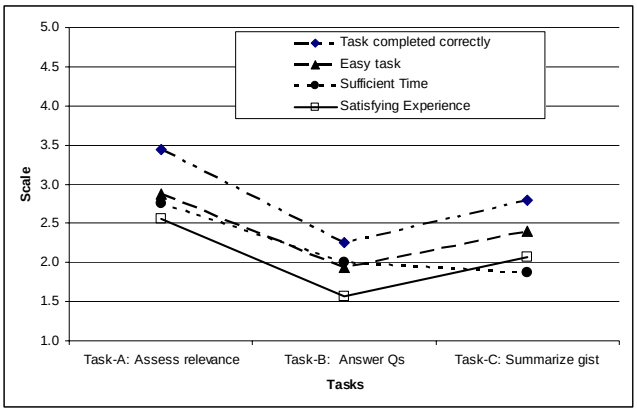

Figure 2. User perception of task completion where 1=worst/most difficult and 5=best/easiest

The response by dimension followed a similar pattern with responses for Task-A scoring the highest and those for Task-B scoring the lowest. However, only task completion (borderline significance: ($F(2,47) = 3.057$, $p=.057$)) and satisfying experience ($F(2,47)=3.606$, $p=.035$) differed by task. Post hoc tests showed that the key differences lay between Task-A and Task-B. These are weak results given the mean lies below the midpoint on the scale. Clearly, participants' found the tasks equally difficult.

4.2.3 Response to the Task

Participants responded to the tasks by recording answers which were subsequently 'graded' by one of the researchers. As illustrated in Table 1, participants varied from performing very poorly (Task-B) to very good (Task-C) on the tasks, based on the percentages in the Accuracy column. On average, participants wrote 135.3 words for Task-A and 86.8 for Task-C. Task-B, specific questions, had many unfilled responses, 3, 13 and 11 for questions 1, 2 and 3, respectively.

From the transaction log files (and from the research system), we determined the amount of time that participants spent on the tasks. Although participants were given a pre-defined amount of time, not all participants used their time allotment (see Table 1) despite indicating that they did not have enough time to do the task (see previous section).

Table 1. Participants mean response to Tasks

Tasks	Accuracy (%)	Time (sec)	# of Words	# of Actions
Task-A: Assess relevance	50.8	768.0	135.3	25.6
Task-B: Answer specific Qs	25.5	522.5	N/A	17.7
Task-C: Summarize gist	80.0	827.7	86.8	17.6

In addition from the transaction log files, we assessed user actions. Each user action constituted one mouse click on some interface object including the browser buttons. Most of the activity took place in conjunction with Task A, which might be expected: in Task A, three webcasts were examined. However, Task B constituted three fact-finding questions, and one would have anticipated a similar level of activity. But Task B is closer in activity level to Task-C.

In the post-task interview, participants indicated how they responded to the task. Most found Task-B a challenge, noting that the interface tools did not support that kind of task. Responses showed a certain amount of frustration: as one participant stated, there was "no real information to lead to the answers." As the interface offers no tool to "dig" into the video and search it, there is no other way to find the answers. Participants indicated that "…it almost seemed like I was looking for something in a pile rather than stacked, like a library system where I know exactly where to look rather than having to go through a whole bunch of papers and picking up things and kind of randomly hoping you'd find something." Another referred to it as a "grab bag thing."

The time constraint was also perceived by some participants as a hindrance; they would like to listen to more of the lecture (even all of it) to find the answers (although this may not be a realistic response except in exceptional circumstances).

In summary, by examining the participants perception of the task, their responses to the tasks, their navigational pathways, and there statements in the talk-after, it is clear that these tasks were difficult in this context.

4.3 Tools

4.3.1 Overview

From the first set of results, we found that participants had little exposure to ePresence (and thus were not unduly influenced by previous experience), and were familiar with the topical area of the videos, if not with the exact topic of each video. However, they had difficulty completing the tasks and achieved mixed results with the tasks. In this set of results, we examine how the interface helped in the quest for suitable responses to the tasks.

As described in methods and illustrated in Figure 1, participants had four tools to aid in completing the tasks in addition to the usual browser widgets. All tools were always present at the interface and with the exception of the search button did not require any action to activate. This adds a level of complexity to the analysis. The tools were omnipresence; as with direct manipulation style interfaces in general, it is difficult to isolate effects. Thus the results in this section especially those tied to the transaction log responses need to be interpreted carefully.

Participants tended to start the task by browsing the Table-of-Contents "to get a sense of the whole…I'm looking for text… I'm looking for titles that would indicate some more rich information so I can make my choice..." While this comment was made with reference to Task A, the same may also be said for the other two tasks. The strategies taken by the participants in completing the task can be found in [5]; in this section we concentrate on how the tools were used, and how useful they were for the tasks.

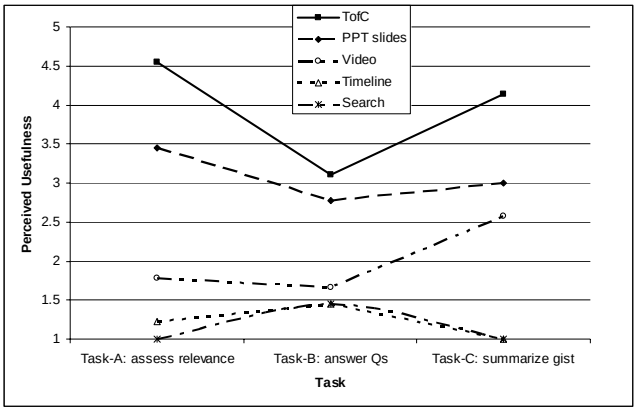

Figure 3. Usefulness of the tools for each task where 1= less useful and 5=most useful

4.3.2 Usefulness of the Tools

In the post-task questionnaire at the end of each task, participants assessed their perceived usefulness of each tool for that task. The results are illustrated in Figure 3. Participants rated the tools on a five-point scale from not-at-all useful to extremely useful. Few tools were rated as "extremely useful" for any task although there were notable exceptions – the Table-of-Contents was so rated by 55% of participants. Some tools were rated as "not at all useful" such as the Search and Timeline. Overall, participants tended to find some tools more useful than others according to the task ($F(2,24)=2.264$, $p=.034$).

In addition, there were within-task effects. The Table-of-Contents was perceived more useful for Task-A and Task-C than for Task-B ($F(2,24)=4.936$, $p=.017$). This result may be anticipated: the Table-of-Contents provides summary information about the webcasts needed in choosing from among webcasts (Task-A) and for writing a gist of a webcast (Task-C). From these results, the Table-of-Contents and Powerpoint Slides are the most useful and perceived as somewhat to extremely useful by participants while the other tools were not as useful.

4.3.3 Use of the Tools

In addition to perceived usefulness, we examined participant actions using the transaction log data. There were significant differences by both cumulative time spent using each tool and the number of instances of tool usage by task. But different amounts of time were assigned per task which can explain some of those differences. Instead of examining those aspects, we focus on how much time was spent per instance of tool use in each task to understand more fully the *intensity* of that use. We speculated that if more continuous time were being allotted to a particular tool, then it was more likely that the tool was receiving more intensive use.

As illustrated in Figure 4, there was an interaction effect. ($F(2,162)=38.177$, $p<.001$). Participants used the Powerpoint Slides ($F(2,160)=17.870$, $p<.001$), Timeline ($F(2,160)=88.112$, $p<.001$) and Video ($F(2,160)=88.112$, $p<.001$) significantly more often than the other tools. These differences were apparent in the tasks. In Task-A, the tools, when used, were all used for about the same amount of time. However, the Video window was used for shorter periods of time per instance than in the other two tasks. Significantly more time was spent with each Powerpoint Slide selection in Task-B, and similarly for the Timeline in Task-C than for all other tasks. Clearly this finding needs to be carefully interpreted. For the both Timeline and the Video, it is likely that the participant was listening to the audio.

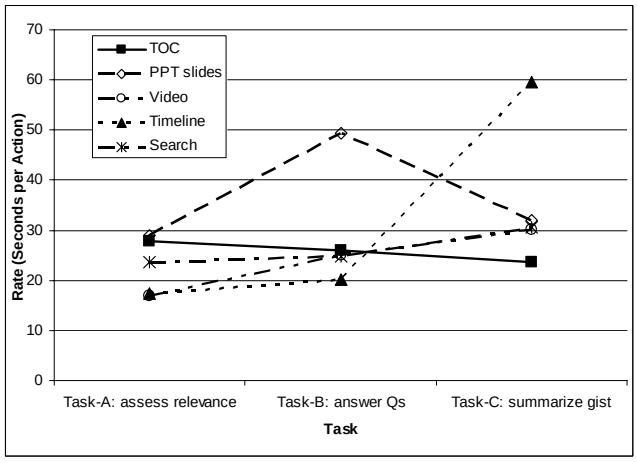

Figure 4. Amount of time spent per user action by task

These objective measures of use provide one approach to understanding how the tools were used within each task. But they tell only half the story. The 'talk-after' interview enabled participants to provide comprehensive, although personal perspectives on the capabilities of these tools to help them in

completing each task. We will address these comments for each tool in this section.

Video Window

The visual presentation – the moving images – tended not to be used by participants – "the video is just not a factor." Although some commented on the uninformative talking head, others noted the persuasiveness of the presenter: "the way they talked and that their body language talked to me" as being a factor in making a decision about a video. This was most noticeable for Task-A when a choice was being made among three videos.

The video window received many comments that related to:
a) window: "the video is just so small";
b) video controls: no rewind button to make it easier to re-listen to sentences, and no pause: "I can't pause it or stop it; I couldn't rewind it"
c) video streaming: slow pace of the buffering – "the little Quicktime ball... persistently stays in front." "I would rather download the video and have it run from my computer"
d) production quality: jerkiness of the video, the lighting, and the *presence* of the presenter.
In terms of doing these tasks, participants commented that they could have done with just an audio track alone, although this was clearly not a unanimous perspective.

Powerpoint Slides

The Slides were heavily used by participants. A technique often applied was rapid forward (and backward) movement for a fast skim of the contents. Sometimes using the Slides was a tactic to avoid using the Table-of-Contents, which loaded and started the Video, a time-consuming operation.

Some slides had uninformative bullet points and graphic examples that were not self-explanatory. They were not easy to interpret by someone new to the topic, or even new to the presentation. Dropping into the middle of the presentation slides proved to be a challenge for comprehension; they were not intended to be read in isolation from the rest of the presentation. In addition, the titles of the slides were perceived as not meaningful.

Some slides had legibility problems – the fonts were too small. Participants did not notice that double-clicking on a slide will open it in a larger window, suggesting that this operation is not intuitive. Slides also lacked consistency in style; these presentations were done by experienced academic researchers and/or professionals, and not by professional presenters. We do no suggest that the presentations were poorly executed; more importantly, the slides were prepared for live presentations not for archiving purposes. That is, they were developed to support a talk by highlighting the main points and/or illustrating some aspect; they were not developed for information searching purposes, or for re-use. These were not transcriptions of the talks, a point noted often by participants seeking a "textual" equivalent to the talk.

Table of Contents

As described in the methods section, the Table-of-Contents (ToC) was semi-automatically created on the fly from the Powerpoint slides, and linked to the video. Thus clicking on a title in the ToC loads the webcast and starts it at that point in the presentation. Because the webcast took a perceptible time to load, participants started to avoid clicking on a title, and instead used it like a text summary, with fast forward and backward scanning movements. Overall, the ToC was heavily used to quickly browse the contents of the presentation, especially in Task-A and Task-C.

The ToC varied significantly among the videos – from single phase titles in one, to a large chunk of text in another, much like the text of the presentation. Thus, each was used slightly differently. Often the audio presentation was stopped or shut off by participants so that they could concentrate on reading the ToC. The ToC was also used to enable focusing in on the text. The ToC showed at least where an overview, outline or conclusions might be contained within the webcast. These items were looked for by participants, especially in Task-A and Task-C to give them a sense of the purpose of the webcast, and thus its pertinence to their task. For those looking for key aspects of the talk, the ToC too enabled them to zero in on those points. But overall, despite its flaws, the ToC became the key navigational map for the webcast, enabling a bird's eye view of the presentation.

Timeline

The Timeline was met with mixed opinions. Some did not understand how it worked – how it corresponded with the Table-of-Contents and the Slides. Some found the density of the lines too crowded to distinguish among the Slides and Chapters.

Participants had expectations about how the Timeline should work. They expected it to be 'dragable.' They also noted that the cursor obscures the titles of the chapter/slides when scanning the Timeline, rendering it useless for that action. While it contains chapter and slide divisions, those divisions are not mapped to the conceptual beginning of a slide and the audio thus sometimes starts mid-sentence.

Search

The search button unfortunately does not work well, as it is at present, the least developed function within ePresence. At the time of this study it searched only the text of the table of contents (but now extends to the Slides as well). We did not warn participants about its limitations. Participants raised on Google had expectations about what the search function should be doing. When the Search button failed, some participants tried the browser "find in text" button, but this could not scan the Slides, a graphic object, and, thus worked only with the Table-of-Contents.

In summary, five different tools were available to assist the participants in accessing the intellectual content of the webcasts. Two of these tools are artifacts of the presenter: Slides and the presentation itself (video and audio); two were designed for the webcasting system: the Timeline, and Search, and one, the Table-of-Contents, was created for the system, but using the presenter's materials. Of the five, the Table-of-Contents and the Slides proved to be the most useful in assisting people with these information tasks. The use and usefulness of the tools was not consistent across all tasks, and extensive use did not always correspond with perceived usefulness. The Timeline was heavily used by those writing a gist of a webcast, but was perceived by participants as not very useful. On the other hand the Slides were the most used by those looking for answers to specific questions, and was perceived as somewhat useful for that task. The Table-of-Contents was used about the same amount of time in all tasks and perceived the most useful of all the tools.

5. DISCUSSION

This research examined how 16 people approached three types of information tasks using stored webcasts. The research literature is sparse concerning this complex information object. While it can be classed as a video, it contains multiple parts including support materials used by a presenter such as slides. Unlike the other 'talking-head' videos — newscasts, webcasts typically do not contain close-captioned text or scripts, making access to their intellectual content a challenge. In addition, webcasts tend to have a wide range of speakers with a myriad of accents and speaking styles and who are not professional broadcasters, making voice recognition a complex problem.

The intent of our study was not to find the webcast, but to find information in the webcast and/or to judge its pertinence. Within the information field, most information tasks concern finding an information object, or scanning surrogates to find an object. Rarely is the *use* of that object examined. The tasks in this study involved use of the object, in this case, the webcast, and the tasks are those which previous work found to be pertinent to webcast re-use [19].

Participants found the tasks difficult despite the fact that they were recruited for their background, and thus their ability to comprehend the content of the webcasts. In addition, participants found the tools that should have been helpful rather cumbersome, and the aspects of the experience unsatisfying. Despite the somewhat negative personal perception of the experience, participants performed remarkably well in terms of task success, scoring high on one of the three. The question is why did participants find the tasks so difficult?

There is no text transcription of the presentation, and thus participants had to rely on indirect routes to the pertinent information. This included scanning the Slides and the Table-of-Contents in the hope that some of the words and phrases would provide valuable cues. The Table-of-Contents tended to be used like a map while the Slides provided more specific and contextual information identifying likely zones within the presentation. The slides do not (and likely can never) enable full comprehension of a webcast, since they are supporting 'actors' in the presentation. They do supply a series of access points into the webcast, and depending on the clarity of the author in constructing the slides, can provide useful insights into the content of the presentation.

Interacting directly with the webcast proved to be time-consuming and awkward. Participants found the delay in initiation of the video streaming to be troublesome, and thus tended to avoid manipulating the video. This problem characterizes all video streaming engines and remains a technical matter that requires additional research. Participants had experience in running video on their computers, and had expectations that the streaming would work in exactly the same way. In addition, the video had few of the typical video controls that might have helped in skimming the webcast.

Poor synchronization among the tools proved to be a significant challenge. The system is set to synchronize the Table-of-Contents with the video, and the Timeline with the video, but not the Table-of-Contents with the Slides. The Slides may be scanned forward and backward independently of the video. These actions make sense for some tasks, but not for all. Participants sometimes found this restrictive and developed a mechanism to work around it. This apparent lack of user control was a problem for participants, as the system responded in unexpected (to participants) ways.

Participant needs for synchronization were mixed and complex. Participants found the lack of synchronization between the Table-of-Contents and the Slides to be a handicap, as the Table-of-Contents titles could sometimes not be interpretable without the richer context of the slides. They assumed that clicking on the title in the Table-of-Contents would put that Slide in focus. But notably and more importantly, they would also assume at different times and for different tasks that this would also launch the video at that point, while leaving the Slides untouched. Thus, multiple types of synchronization need to be employed, but must be balanced with a perception of user control. This is a significant design issue that requires investigation. Enabling tight synchronization simultaneously among all the tools reduces user control; enabling user selection of the objects to synchronize adds to the user's cognitive load, and increases interface complexity. We need to find a way of handling both.

One would typically expect the search to be helpful. But because of the nature of video and the graphic representation of the slides, the only textual information at the time of the study was contained in the Table-of-Contents. Many participants reiterated over and over that a searchable transcript would have been the most appropriate device, especially for the specific questions in Task-B. Thus participants were left attempting to zero in on useful nuggets without knowing if the nuggets existed, and if they did exist, without knowing how to easily get to them.

As previously discussed, text transcriptions are difficult to create for webcasts. One also wonders what might be lost from the presentation, if, for example, only a transcription was used for the information tasks used in this study. Some participants found the confidence exuded by the presenter to be a factor in their perception of the value of what was being said. Thus, 'talking head' videos have value in the visual as well as in the audio. In addition, illustrative matter on the slides enhances the verbal presentation providing added value to the audio content. Developing a search system that includes the transcribed text of the audio with the visual representation of both the presenter and the slides is a complex problem that has yet to be solved.

Without useful overview tools, it is difficult to get a sense of the webcast. As participants reminded us, the Slides are most useful *during* the presentation; after the presentation, the Slides when examined independently of the presentation may not provide sufficient context, and thus may not be understood without the thread of the presentation. The Timeline which should have been a useful tool for scanning the contents quickly proved to be less useful in this study. The Table-of-Contents provided the best overview, but was somewhat restricted because it relied on the usefulness of the presenter's slide titles.

Webcasts are, in essence, 'talking head' videos; much of the content lies in the audio. Yet to date limited research has been done to provide good audio surrogates. Had the participants access to good audio clips that were fast to load and with good semantic cues, they may have performed much better on the tasks. This is a direction for future research.

In addition, other mechanisms are needed to provide good surrogates. Some participants suggested simply that the

announcement for the talk would have been a useful device, especially in deciding on the relevance of a video. Other surrogates are needed for fast skimming. For example, integrating the timeline with keyframes in a dynamic presentation including also appropriate audio boundaries might prove to be useful. There are many possibilities for surrogate, some based on selected media used in the presentation, and some that are move multi-media in content.

Although much of this discussion has surrounded items external to the webcast, there are items intrinsic to a webcast that also need addressing. Webcast are presentations, and despite the criticism of Slides, slides in general tend to be a significant part of a presentation. All webcasting systems allow for them. Yet, the creation of slides is a personal matter, and to date, there are no protocols even for formal presentations, although guidelines exist. If webcasts are to become educational objects, then some standards will be needed for accompanying material. Notably, in the case of this task, participants had expectations about the introduction and conclusion, assuming that the presenter would provide an overview at the beginning and a summary at the end. Furthermore, participants made judgments about the webcast based on the mannerisms and carriage of the presenter. This aspect of authority based on visual characteristics is rarely referenced, although clearly pertinent in this case.

The ePresence system services live events very effectively, as attested by its use in many organizations. The challenge is how to provide effective access to stored webcasts. Reviewing a 40-45 minute video is much like scanning a book. One does not expect to get a sense of a book by reading it from cover to cover, and likewise one should not have to view a 40 minute video to get a gist of its contents, or to decide its suitability for a particular work task.

In this presentation the table of contents was an extremely useful device despite the fact that it is created semi-automatically during the presentation. Participants had expectations about textual tools and maximized their use of those text tools; one wonders if they are text-oriented when it comes to information, and thus find it difficult to process in other modes.

The tasks, however, were especially challenging in this environment. It is not that the tasks were especially difficult, as they are clearly consistent with a typical text environment and with some digital video systems. However, performing these tasks with webcasts is quite another matter. Webcasting systems have not yet adapted to the idea that webcasts are information objects with intellectual content that may be consulted and used for a myriad of purposes that were not intended or foreseen by the original presenter or webcasting system developers. The challenge now becomes how to effectively provide for this additional, but unexpected use. Webcasts will eventually be included in digital libraries, archives and other types of information systems. We need methods for interacting with them in effective ways; this is a new area for both development and research.

6. RESEARCH & DESIGN IMPLICATIONS

From this study of webcast re-use, we have learned how people interact with stored webcasts and how they use a set of tools to facilitate access to the content of the webcast. This analysis has resulted in a number of both design and research implications.

a) *Text transcripts*

The need for text transcripts as a source for search cannot be underestimated, if webcasts are to be components of digital libraries/archives. How to create those transcripts is a research problem for voice recognition. While the transcripts may not be 90-100% accurate given the variety of voices and accents, it is likely that less accurate transcripts may provide a suitable source for search. In this case, inaccurate transcripts should be hidden from the user so that trust issues with the content do not emerge and exacerbate the problem. Instead the threshold should be set such that the probability of a match is fairly high, and inaccurate words are not included in the database. Even a partial transcript will be better than none to facilitate better search. How partial the transcript can be remains a problem for further research.

b) *Synchronization and complexity*

The need for more (and less) synchronization was a key outcome from this study. At times participants needed everything in sync, while at other times, they needed two to four of the possible tools to be in sync. Enabling complete user flexibility may add to interface complexity. This problem demands a creative solution.

c) *Timeline*

To date many videos have timelines for a linear time-based representation of the video. In the case of ePresence, we attempted to add functionality by including the titles of slides and chapters which proved to be awkward to use. A webcast timeline is different from a typical video or audio timeline, and needs re-thinking with regard to a webcasting application. A timeline is a useful scanning device for dipping into the presentation, and scanning its contents. The design problem is thus how to represent a lengthy webcast in a condensed form.

d) *Surrogates*

Most text and video systems provide surrogates to illustrate the content or provide good overviews or previews of the content. For text-based systems, these are abstracts or snippets of text such as those present in search results. For digital video, these are often based on keyframes. But what makes a good surrogate for a webcast? One solution is audio surrogates: snippets of audio that are fast to load and represent key areas of the text. Part of the research and the design, thus, is how to identify these key audio snippets, and how to represent them to users. Audio has no visual cues like moving/still images and text. In addition, a webcast is a complex object, and we have yet to see good surrogates for a multi-media object, as surrogates tend to be developed for a single form of media. How could one integrate a ‘talking-head’ video, with its audio, and in this case slides?

e) *Notes and Bookmarks*

For the types of tasks that we used (and likely also with live webcasting), users needed the ability to annotate the contents as well as bookmark parts of the video for easy return. Figure 1 contains an image of the stored webcasts interface. Adding new functionality will be a challenge from a complexity perspective.

7. DEVELOPMENTS WITH ePRESENCE

ePresence is an open source research project as well as an operational system in use by many organizations. Since this study

a number of notable improvements have been made including work toward improving the synchronization among the tools, the ability to search through both the slides and the table of contents, and improvements to the video controls. In addition several research projects are in progress: a) creating partial to full text transcripts of the audio which will significantly improve access to the intellectual content of the presentation, b) rapid initiation of the video playback to reduce the video load problem, c) user-controlled synchronization of the tools to provide for flexible content browsing, and d) re-inventing the webcast timeline to enhance both access to the content and rapid scanning of the presentation.

8. CONCLUSIONS

Webcasting systems are migrating into the digital library/archive venue by including also access to stored webcasts. One can equally imagine a digital library/archive that includes webcast objects within its collections. However, unlike typical informational objects such as articles and videos, webcasts are complex with multiple components in different media.

In this study, we examined how people interact with a webcast to perform classic information tasks. Although the tasks had equivalents in the text world, users were challenged in performing those tasks and we conclude that the tools provided were not sufficient for undertaking these tasks. Reviewing a webcast is different from attending a webcasting event, and requires modifications to existing tools and some additional tools.

In the case of this study, one might conclude that ePresence has usability issues (and the study did identify some areas for its improvement), and thus this was the reason for user performance. This, however, is not the most significant contribution from our work. Our study is one of the first, if not the first, to examine the problem of user interactivity with webcasts for informational purposes, to illuminate that process, and to provide guidance for future research and development in webcasting systems. Doing information tasks with webcasts is a new area for both research and development.

9. ACKNOWLEDGMENTS

This work was funded by Heritage Canada, NSERC (Natural Science and Engineering Research Council of Canada) and Canada Research Chairs Program grants to the first author. The authors thank the 16 participants for their time, research assistant, Joan Bartlett, who conducted the interviews, Peter Wolf, technical support for ePresence, and the anonymous reviewers for their helpful comments.

10. REFERENCES

[1] Baecker, R.M. A principled design for scalable internet visual communications with rich media, interactivity, and structured archives. In *Proceedings of CASCON 2003* (Toronto, ON, Canada, October 6-9, 2003). IBM Press, 2003, 83-96. [Note: see additional works concerning ePresence at http://kmdi.utoronto.ca/rmb/publications.html]

[2] Christel, M. G., Winkler, D., Taylor, R., and Smith, M. Evolving video skims into useful multimedia abstractions. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '98)* (Los Angeles, CA, USA, April 18-23, 1998). ACM Press, New York, NY, 1998, 171-178.

[3] Ding, W., Marchionini, G. and Soergel, D. Multimodal surrogates for video browsing. In *Proceedings of the fourth ACM conference on Digital libraries (DL'99)* (Berkeley, CA, USA, August 11-14, 1999). ACM Press, New York, NY, 1999, 85-93.

[4] Ding, W. and Marchionini, G. *A Study on Video Browsing Strategies.* Technical Report UMIACS-TR-97-40, University of Maryland, College Park, MD, 1997.

[5] Dufour, C., Toms, E.G., Lewis, J. & Baecker, R. User strategies for handling information tasks in webcasts. In *Proceedings of the ACM Special Interest Group on Computer-Human Interaction Conference (CHI 2005)* (Portland, Oregon, April 2-5, 2005). In press.

[6] Geisler, G. et al. Interface concepts for the open video project. In *Proceedings of the 64th ASIST Annual Meeting (ASIST2001)* (Washington, DC, USA, November 3-8, 2001). Information Today Inc., Medford, 2001, 58-75.

[7] Hauptmann, A.G., Papernick, N., and Christel, M. Video retrieval with multiple image search strategies. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries (JCDL2002)* (Portland, OR, USA, July 14-18, 2002). ACM Press, New York, NY, 2002, 376.

[8] He, L. et al. Comparing presentation summaries: slides vs. reading vs. listening. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '00)* (The Hague, The Netherlands, April 1-6, 2000). ACM Press, New York, NY, 2000, 177-184.

[9] Lee, H. & Smeaton, A. Designing the user interface for the Físchlár Digital Video Library. *Journal of Digital Information, 2*, 4 (2002). http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Lee/

[10] Li, F. et al. Browsing digital video. *CHI Letters: Human Factors in Computing Systems, 2*, 1 (2000), 169-176.

[11] Moraveji, N. Improving video browsing with an eye-tracking evaluation of feature-based color bars. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries (JCDL'04)* (Tuscon, AZ. USA, June 7-11, 2004). ACM Press, New York, NY, 2004, 49-50.

[12] Rowe, L.A., Harley, D., and Pletcher P. BIBS: a lecture webcasting system. (2001) http://bmrc.berkeley.edu/ research/ publications/2001/160/bibs-report.pdf

[13] Slaughter, L., Marchionini, G. and Geisler, G. Open Video: a framework for a test collection. *Journal of Network and Computer Applications, 23*, 3 (2000), 219-245.

[14] Smeaton, A. F. et al. The Físchlár Digital Video System: a digital library of broadcast tv programmes. In *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries (JCDL'01)* (Roanoke, Virginia, USA, June 24-28, 2001). ACM Press, New York, NY, 2001, 312-313.

[15] Syeda-Mahmood, T. and Ponceleon, D. Video retrieval and browsing: learning video browsing behavior and its application in the generation of video previews. In *Proceedings of the ninth ACM international conference on Multimedia (MM'01)* (Ottawa, ON, Canada, September 30-

October 5, 2001). ACM Press, New York, NY, 2001, 119-128.

[16] Toms, E.G., Freund, L. and Li, C. WiIRE: a web interactive information retrieval experimentation system prototype. *Information Processing & Management, 40*, 4 (2004), 655-675.

[17] Wactlar, H., Hauptmann, A., Gong, Y. and Christel, M. Lessons learned from building a terabyte digital video library. *Computer, 32*, 2 (1999), 66-73.

[18] Webopedia. http://isp.webopedia.com/TERM/W/Webcast.htm

[19] Whittaker, S. et al. SCAN: designing and evaluating user interfaces to support retrieval from speech archives. In *Proceedings of SIGIR'99* (Berkeley, CA, USA, August 15-19, 1999). ACM Press, New York, NY, 1999, 26-33.

[20] Wildemuth, B. M., et al. How fast is too fast? evaluating fast forward surrogates for digital video. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries (JCDL'03)* (Houston, Texas, USA, May 27-31, 2003). IEEE Computer Society, Washington, DC, USA, 2003, 221-230.

[21] Yang, M. et al. Measuring user performance during interactions with digital video. In *Proceedings of 66th ASIST Annual Meeting (ASIST2003)* (Long Beach, CA, USA, October 19-22, 2003). Information Today Inc., Medford, 2003, 221-230.