

# Exploring User Interaction with Digital Videos

Christine Dufour<sup>1,3</sup>

Elaine G. Toms<sup>1,3</sup>

Joan Bartlett<sup>1</sup>

Joseph Ferenbok<sup>1,3</sup>

Ron Baecker<sup>2,3</sup>

Faculty of Information Studies<sup>1</sup>, Department of Computer Science<sup>2</sup>, Knowledge Media Design Institute<sup>3</sup>  
University of Toronto

## *Abstract*

This poster presents initial results from a study that explores user interaction with digital videos. While much of the interface research to date examines the types of surrogates that one needs to locate a video and to determine its pertinence, our work takes place at the level of the video and the tools that enable access to specific parts of its content. In this study, 16 participants performed three types of tasks using ePresence, a webcasting system that handles both live and stored video. Several data collection mechanisms were used: questionnaires, screen capture and interviews, to assess the interaction holistically. The analysis is currently in progress, although preliminary results show differences in interface tool use and by task type.

*Key words: digital videos, user interaction, ePresence.*

## **1 Introduction**

Research to date in digital video has dealt with many of the technical issues required to build a system. Much less emphasis, however, has been placed on the interface to such systems, or indeed to how a digital video is skimmed and scanned, and its content comprehended. In this study, we assess how ePresence [1], a webcasting system, facilitates access to the content of a digital video, and the pathways that people take in navigating a video for explicit purposes.

## **2 Previous Work**

Changes in technology over the past 10 years have led to the viable development in digital video projects such as the Open Video Project, the Informedia Project, and the Fischlar system, which subsequently lead to increased research and real world interest in large-scale online digital video information systems. Although digital video collections present a number of technical challenges from compression to storage, they also present unique opportunities for retrieval and interface design [3, 5, 10].

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.” [10] Surrogates are classified by the medium in which they are presented: text, still image, moving image, audio, and multimode surrogates. Text surrogates generally involve textual metadata such as bibliographic information or automatically summarized audio extracts [see, for example, 7]. Still image surrogates include storyboards, thumbnails or slideshows, and key frame displays [4]. Moving or dynamic surrogates include video skims and fast

forwards [8]. Audio surrogates use extracted audio information and have been used in combination with other multimodal approaches [2], but also present interesting possibilities in light of research into the skimming of recorded speech. Other surrogates include timelines and representation of the presentation such as its slides.

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 [6] and their subsequent use of digital video. To date, no study has attempted a holistic understanding of the strategies used by people in skimming and scanning lengthy (over one hour) digital video objects. So, although, attempts are being made to understand and improve access strategies to video libraries, little work has been done to understand how users interact with extended linear content-based video objects and how surrogates may facilitate their browsing. This work is an exploratory study that examines those user interactions in the context of ePresence.

## **3 Methods**

### **3.1 System used -- ePresence**

The stored video function of ePresence, a webcasting system that also handles live webcasts, was used for this study (<http://www.kmdi.utoronto.ca/epresence>). In addition to reviewing a video, the system has several features: (1) the slides used in the presentation, (2) a table of contents created in real time from the slides that identifies in a metaphoric like way the ‘chapters’ in the presentation, (3) a timeline of the presentation illustrating the chapters divisions as well as the slide coordination.

### **3.2 Participants**

Sixteen participants were recruited from the University of Toronto. The participants were adults (63% under 30 years old). They were mainly students (88%) with an undergraduate degree (31%) or a master degree (63%). While they already experienced online video, they were modest users of online videos. Almost half of them were familiar with ePresence. Very few were familiar with the videos used while around half of them were familiar with the topics of those videos.

### **3.3 Tasks**

The participants were assigned, in a random order, three different tasks, inspired by Whittaker et al. [9]:

- Quickly select videos related to a specific topic;
- Answer three very specific factual questions on a video;
- Summarize the main theme(s) or gist of a video.

### 3.4 Procedure

Each participant session took about 2 hours and followed these steps:

- After consenting to participate, participants filled a demographic and digital video experience questionnaire;
- Participants had practice time with the system;
- For each of the three tasks, participants:
  - answered questions about their knowledge of the video(s) used and of the topic covered;
  - worked on the task using ePresence for 10-15 minutes;
  - answered questions regarding their perception of the process and the level of accomplishment attained in doing the task;
  - annotated the session using a 'talk-after' style while the screen capture of that session was replaying;
- When the three 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.

The data were collected using three mechanisms: (1) a Web-based database that captured responses to all questionnaires, (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.

### 3.5 Data analysis

The analysis of the questionnaires and the interviews involves statistical as well as content analyses. A video editing software is being used to analyze the screen-capture files to extract the process used by each participant.

## 4 Preliminary results

Of the three tasks, "answering specific questions" was the less well executed and received from the participants the lowest values for the correctness of their answers, the easiness of the task and their satisfaction. Participants suggested ways of helping with this task including adding a searchable textual transcript of the lectures and having a more precise control over the video playback.

The participants' use of the different tools varied as well: the most used and useful tools were the slides and the table of contents, and the less used and useful were the search button and the timeline. In general, only slight differences were found between the tools perceived usefulness by task. The exception is the table of contents for which a significant difference existed by task type. It was perceived as being significantly less useful for helping to find specific answers than for the two other tasks (quickly select videos and summarize the gist of a video).

## 5 Discussion and Conclusions

This research explored how people interact with a digital video, so that we can design more useful tools to support video use. While initial results suggest certain types of

tools are more useful for certain types of tasks, it also has indicated design flaws with current implementation, e.g., the timeline. Data analysis is still in process. We are analyzing screen capture files to identify user behaviour in their response to the tasks. This will be complemented by the rich information from interviews concerning participants' decisions, problems and expectations. The results will be used to identify new tools to aid in the completion of such tasks.

### Acknowledgments

Research funded by Heritage Canada sub-grant to E.G. Toms.

### References

- [1] Baecker, R.M. (2003). A principled design for scalable Internet visual communications with rich media, interactivity, and structured archives. In Proceedings of CASCON 2003. 83-96.
- [2] Ding, W., et al. (1999). Multimodal surrogates for video browsing. In Proceedings of Digital Libraries 99. New York, ACM Press: 85-93.
- [3] Geisler, G., et al. (2001). Interface concepts for the Open Video Project. In Proceedings of ASIST2001. Medford, Information Today Inc.: 58-75.
- [4] Hauptmann, A., G., et al. (2002). Video retrieval with multiple image search strategies. In Proceedings of JCDL2002. Portland, ACM: 376.
- [5] Lee, H. & Smeaton, A. (2002). Designing the user interface for the Fischlár Digital Video Library. *Journal of Digital Information* 2(4). <http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Lee/>
- [6] Syeda-Mahmood, T. & Ponceleon, D. (2001). Video retrieval and browsing: Learning video browsing behavior and its application in the generation of video previews. In Ninth ACM international conference on Multimedia. Ottawa, ACM Press: 119-128.
- [7] Wactlar, H., et al. (1999). Lessons learned from building a terabyte digital video library. *Computer* 32(2), 66-73.
- [8] Wildemuth, B. M., et al. (2003). How fast is too fast?: Evaluating fast forward surrogates for digital video. In Proceedings of JCDL2003. Houston, IEEE Computer Society: 221-230.
- [9] Whittaker, S., et al. (1999). SCAN: Designing and evaluating user interfaces to support retrieval from speech archives. In Proceedings of SIGIR '99. Berkeley, ACM Press: 26-33.
- [10] Yang, M., B. M. Wildemuth, et al. (2003). Measuring user performance during interactions with digital video. In Proceedings of ASIST2003. Medford, Information Today Inc.: 221-230.