Browsing Digital Video

Francis C. Li[2], Anoop Gupta[1], Elizabeth Sanocki[1], Li-wei He[1], Yong Rui[1]

[1]Collaboration and Multimedia
Microsoft Research
Redmond, WA 98052
{anoop, a-elisan, lhe, yongrui}@microsoft.com

[2]Group for User Interface Research, EECS Dept.
University of California, Berkeley
Berkeley, CA 94720-1776
fli@cs.berkeley.edu

ABSTRACT

Video in digital format played on programmable devices presents opportunities for significantly enhancing the user's viewing experience. For example, time compression and pause removal can shorten the viewing time for a video, textual and visual indices can allow personalized navigation through the content, and random-access digital storage allows instantaneous seeks into the content. To understand user behavior when such capabilities are available, we built a software video browsing application that combines many such features. We present results from a user study where users browsed video in six different categories: classroom lectures, conference presentations, entertainment shows, news, sports, and travel. Our results show that the most frequently used features were time compression, pause removal, and navigation using shot boundaries. Also, the behavior was different depending on the content type, and we present a classification. Finally, the users found the browser to be very useful. Two main reasons were: i) the ability to save time and ii) the feeling of control over what content they watched.

Keywords

Digital video, video browsing, video indexing, time compression, pause removal, next-generation video playback interfaces.

INTRODUCTION

One of the primary mediums for content creation and distribution is video. However, the way we watch video has not changed significantly since the invention of the analog video-cassette recorder (VCR) in the 1970s and 80s. The VCR makes it possible to watch video on-demand with the additional ability to pause, fast-forward, and rewind.

Today, Internet video streaming and set-top devices like ReplayTV [19] and TiVo [22] are technologies that are defining a new platform for interactive video playback. Unlike traditional VCRs, ReplayTV and TiVo store video in digital form (MPEG-2) on large hard disks. With digital video stored on hard disks and/or as Internet-based streaming media, instant random access into the content is possible. Seeking to a random location was possible with VCRs but had a large delay associated with it due to the use of tape storage. The instant random access facilitates features such as instant replay of just-observed action and rich indices into the content such as the chapter lists found on digital versatile disc (DVD) videos [7]. In addition, as computing costs continue to drop, processing techniques can be utilized to automatically generate indices or increase the playback speed while maintaining intelligibility. Such features potentially allow a viewer to save significant amounts of time watching a video and more effectively filter the content during playback.

Given this emerging new platform for interactive video playback, we explore the following questions in this paper:

- What potentially high-value features can we provide for browsing digital video?
- Will users derive significant benefits from their use and availability?
- How will the benefits vary with the task and type of content being watched?

We designed and implemented a prototype software video browsing application that provides a wide array of features enabled by digital video technologies. In addition to traditional VCR controls, the prototype provides rich indices for navigation (e.g., table of contents and video shot boundaries), speeded-up playback features (e.g., time compression and pause removal), the ability to make personal annotations that are anchored to the video timeline, and other advanced browsing controls.

Some of these features have been studied previously, but primarily in isolation and only for a narrow set of video content types. We evaluated the combined use of these features across a wide range of video content types: classroom lectures, conference presentations, sports, television dramas, news, and travel. This paper quantifies the use of the various features for the different content types and also documents viewers’ subjective experiences. We also present an informal classification of video content types that helps predict the usefulness and applicability of the different browsing features and their impact on the viewing experience.

RELATED WORK

Previous research in browsing digital media has often focused on either audio or video, but not both. The SpeechSkimmer [3,4] provided an interface for selecting time compressed and pause removed audio playback and for jumping back and forward between pre-defined segments of the recording. The Audio Notebook [21] used time-stamps of pen strokes to index audio and allowed time compressed playback.

For browsing video, the Hierarchical Video Magnifier [15] displayed frames near the current video position to provide context. Arman et al [2] improved the frame selection methods by detecting shot boundaries, which were found useful in editing systems [13]. The Classroom 2000 project at Georgia Tech [5] investigated richly indexed videos of lectures, including indexing based on strokes drawn on a black-board. However, none of these systems explore the wide range of browsing techniques and/or video content types explored here.

Christel et al [6] and He et al [8] have discussed techniques for shortening the viewing time of a video based on audio and/or video analysis. Such techniques, used in systems like CueVideo [17], condense the content into a shortened video summary intended to be watched in its entirety. The user does not control what is deleted to create the shortened summary and cannot browse the resulting video, the focus of this study.

The Informedia [9] project at CMU has performed substantial research in indexing and searching video in the context of information retrieval and digital library systems. Companies like Virage [22] and MediaSite [12] are currently providing these services for finding video on the Internet. Others have used domain knowledge to improve these services for specific video content types like news [11]. Such work focuses on query-based searching of collections of video content rather than on browsing an individual video that is the focus of this study.

The computer software industry has quickly embraced the Internet as a platform for digital video. However, the main focus of industry development has been the creation and distribution of content, not viewing or browsing. As a result, the leading software playback applications such as the Real Networks RealPlayer [18], Apple QuickTime Player [1], and Microsoft Windows Media Player [14] offer relatively few controls for browsing. In addition to the controls found on a VCR, these applications add a seek bar allowing random access via a “thumb” and a table of contents index.

The consumer electronics industry has begun to incorporate more advanced browsing features in the next generation of hardware video playback devices. DVD Video players support random access using a table of contents index. ReplayTV and TiVo set-top boxes offer an index to recorded shows. In addition, they provide the ability to jump forward by 30 or 60 seconds, possibly allowing skipping of commercials, and back 8–10 seconds for “instant replays.” However, these devices do not currently provide features like time compression or shot boundary frames. The user interface design is also quite different as input must be performed using a remote control. Finally, no public data is available on how these devices are actually being used.

PROTOTYPE FEATURES AND FUNCTIONALTIY

Our study used two video browsers: “Basic” and “Enhanced.” The enhanced browser was developed using a modified version of the Microsoft Windows Media Player. The basic browser leveraged the same software, but displayed only a subset of the functionality.

Basic browser controls: The basic controls provide the features typically found on current software video playback applications. They include *Play*, *Pause*, *Fast-forward*, *Seek*, *Skip-to-beginning* of video, and *Skip-to-end* of video. No audio was played during fast-forward as is common with current media players, and seek was accomplished by dragging the seek thumb on the timeline in the interface. Due to limitations of the Windows Media Player, a traditional rewind feature could not be provided.

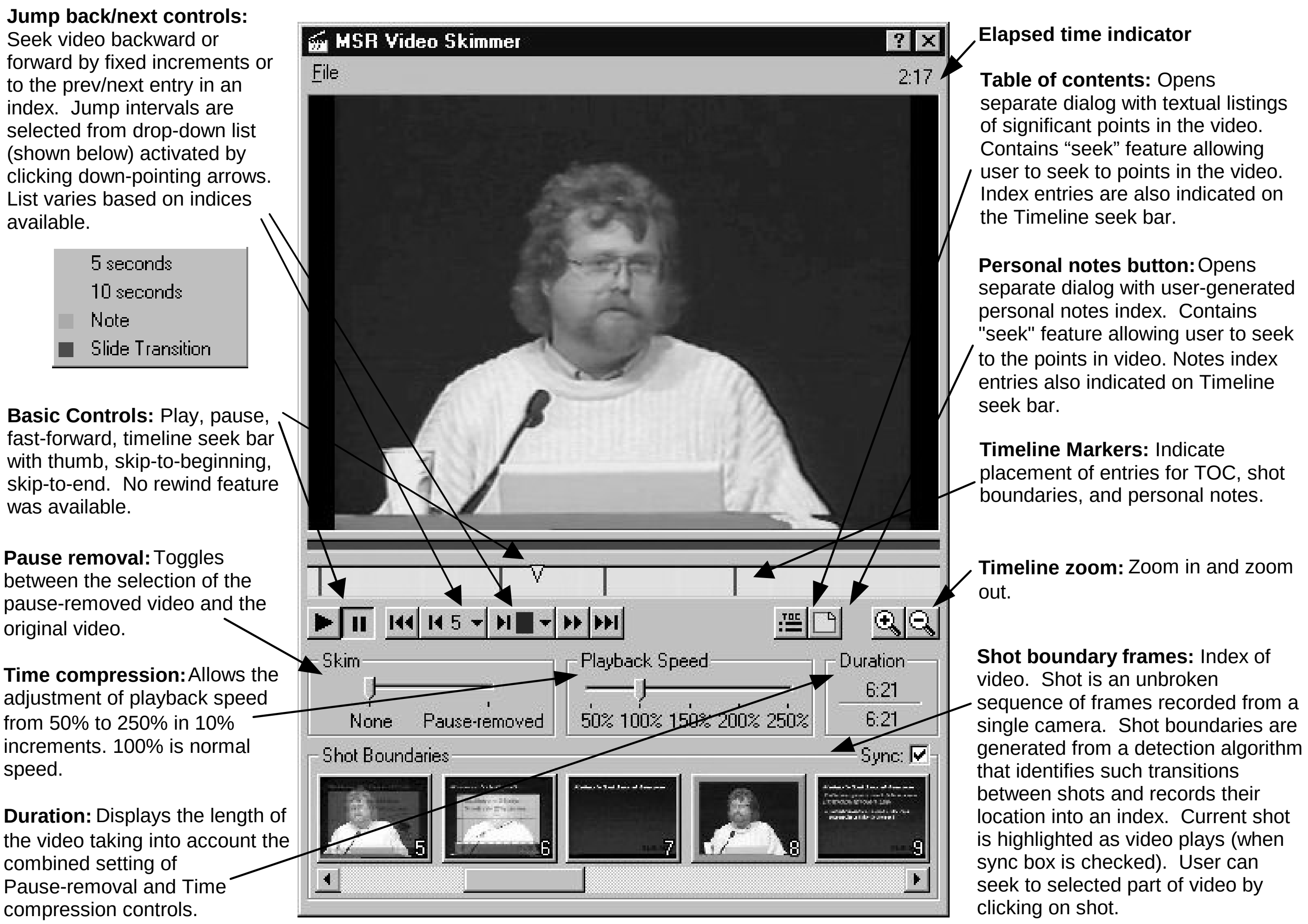
Enhanced browser controls: Figure 1 shows the user interface for the enhanced browser. The following additional controls were provided:

- Speed-up controls: *Time compression (TC), Pause removal (PR)*
- Textual indices: *Table of contents (TOC), Notes*
- Visual indices: *Shot boundary (SB) frames, Timeline markers*
- Jump controls: *Jump-back, Jump-next*

The speed-up controls allow the user to shorten the viewing time of a video. *Time compression* (TC) uses signal processing techniques to increase the playback speed while preserving the pitch of the audio. *Pause removal* (PR) detects pauses and silence in continuous speech and removes both the audio and video segments associated with them.

The textual indices are lists of text entries that describe locations in the video. The user can seek to the location in the video by selecting the associated entry. The *table of contents* (TOC) index is a pre-generated list of entries that cannot be modified. In contrast, the *notes* index is generated from end-user annotations. When the user creates a note, the comment entered by the user is anchored to the current position of the video. We expected that users might use the notes feature to bookmark significant parts of the video for later reference and to record their thoughts regarding the content of the video.

The visual indices are the shot boundary frames and the timeline markers. The numbered *shot boundary frames* (SB) allow the user to visually identify and then seek to a



Jump back/next controls: Seek video backward or forward by fixed increments or to the prev/next entry in an index. Jump intervals are selected from drop-down list (shown below) activated by clicking down-pointing arrows. List varies based on indices available.

Basic Controls: Play, pause, fast-forward, timeline seek bar with thumb, skip-to-beginning, skip-to-end. No rewind feature was available.

Pause removal: Toggles between the selection of the pause-removed video and the original video.

Time compression: Allows the adjustment of playback speed from 50% to 250% in 10% increments. 100% is normal speed.

Duration: Displays the length of the video taking into account the combined setting of Pause-removal and Time compression controls.

Elapsed time indicator

Table of contents: Opens separate dialog with textual listings of significant points in the video. Contains "seek" feature allowing user to seek to points in the video. Index entries are also indicated on the Timeline seek bar.

Personal notes button: Opens separate dialog with user-generated personal notes index. Contains "seek" feature allowing user to seek to the points in video. Notes index entries also indicated on Timeline seek bar.

Timeline Markers: Indicate placement of entries for TOC, shot boundaries, and personal notes.

Timeline zoom: Zoom in and zoom out.

Shot boundary frames: Index of video. Shot is an unbroken sequence of frames recorded from a single camera. Shot boundaries are generated from a detection algorithm that identifies such transitions between shots and records their location into an index. Current shot is highlighted as video plays (when sync box is checked). User can seek to selected part of video by clicking on shot.

Figure 1. Enhanced Browser User Interface

particular shot by clicking on it. As the video plays, the frame corresponding to the currently playing shot is highlighted. The *timeline markers* show the location of the TOC and notes entries with color coded bars. They can be used to judge the location of entries relative to the current position of the video (shown by the thumb).

The *jump-back* and *jump-next* controls seek the video backward or forward, respectively, by a fixed interval or to entries in an index. Users can jump by 5 seconds, 10 seconds, TOC entry, note, or shot boundary. It was hypothesized, for example, that a user might jump back 5 or 10 seconds to repeat parts of the video, whereas the jump next TOC entry control might be used to preview the first few minutes of each consecutive entry in the TOC. Also, it is very difficult to do these operations using the seek thumb. For example, a one-hour video (3600 seconds) spread across roughly 400 pixels (width of our browser) means that moving the thumb one pixel seeks 9 seconds.

Our goal for the prototype was to expose video browsing functionality with a user interface adequate for evaluation. Both the basic and enhanced browsers were instrumented to record the usage of each feature during the study.

USER STUDY DESIGN

The user study was designed to evaluate both feature usage and overall experience with the enhanced browser. Participants were presented with a scenario and browsing task related to one of six video content types: classroom lectures, conference presentations, sports, television dramas, news, and travel.

Each participant completed his or her video browsing task three times. The task was first completed using the basic browser. Then, after a short practice tutorial, the enhanced browser was used for the last two tasks. To encourage browsing behavior, the participants were limited to 30 minutes to browse a 45–60 minute video.

In addition to pre- and post-study surveys, the participants completed a survey after each task. They were asked to describe their browsing strategy and rate their interest in the content of the video, the quality of their experience, and the usefulness of the available features.

The participants were recruited from a pool of non-employees that expressed interest in usability studies at Microsoft. They were screened for two years of computer experience and matching interests with one of the scenarios.

Table 1. Average Ratings of Feature Usefulness. The highest rated Enhanced browser feature for each scenario is highlighted. FF = Fast forward, SB = Shot boundaries, TC = Time compression, PR = Pause removal, Jmp = Jump-back & -next, TOC = Table of contents, Bas = Basic browser, Enh = Enhanced browser. Scale: 1 = strongly disagree, 4 = neutral, 7 = strongly agree.

	Seek		FF		SB	TC	PR	Jmp	TOC	Note
	Bas	Enh	Bas	Enh	Enhanced Browser					
Classroom	4.8	5.6	4.4	4.1	5.0	5.4	5.1	4.8	6.8	3.5
Conference	5.6	4.1	3.6	3.3	4.9	6.9	6.5	5.1	N/A	3.8
Sports	5.2	4.7	5.6	5.9	6.1	5.7	4.3	5.6	5.3	4.5
Shows	5.0	3.6	4.4	4.3	5.1	6.0	4.3	2.8	N/A	2.5
News	5.8	4.9	5.4	4.3	6.4	6.7	6.6	5.6	6.6	4.6
Travel	5.2	5.7	5.4	4.2	6.3	6.6	6.0	6.3	N/A	6.4
Overall	5.3	4.8	4.8	4.4	5.6	6.2	5.5	5.0	6.2	4.1

Table 2. Average Number of Times Feature Used per Participant per Video. The most frequently used Enhanced browser feature for each scenario is highlighted. SB Sk = Shot boundary seek, Jmp Bck/Nxt = Jump Back/Next, TOC Sk = Table of contents seek, Note Sk = Note Seek

	Seek		FF		SB Sk	Jmp Bck	Jmp Nxt	TOC Sk	Note Add	Note Sk
	Bas	Enh	Bas	Enh	Enhanced Browser					
Classroom	21.6	0.0	10.8	0.0	1.5	4.5	2.0	12.5	0.0	0.0
Conference	15.7	0.5	4.2	0.0	2.0	0.5	7.0	N/A	3.0	1.0
Sports	20.0	7.0	12.8	4.5	26.5	0.0	4.0	1.5	2.0	0.5
Shows	14.8	3.0	9.8	1.0	4.5	0.0	11.0	N/A	0.0	0.0
News	34.0	0.5	10.2	0.0	9.5	2.0	10.5	3.5	1.0	0.0
Travel	51.8	3.0	11.0	0.0	55.0	14.5	4.5	N/A	9.5	5.0
Overall	26.3	2.3	9.8	0.9	16.5	3.6	6.5	5.8	2.6	1.1

Five participants per scenario completed the study for a total of 30 participants. Each participant received a Microsoft software product for his or her involvement.

SCENARIOS AND RESULTS

In this section, we describe the browsing scenarios in detail and discuss the corresponding results of the study, but first we present the data that we will reference.

Table 1 presents the average rating of feature usefulness over the participants in each scenario and overall, calculated from surveys completed after each task. Table 2 presents the average number of times features were used by a participant while watching a video. Table 3 shows the average effective playback speed attained using time compression and the combination of time compression and pause removal. Table 4 shows the average percentage of a video watched and decomposes that value into the percentage of video watched only once, exactly twice, and three or more times. Finally, Table 5 shows, on average, what percentage of the task time was spent with a video in different playback modes.

Table 3. Average Effective Playback Speed Attained with the Enhanced Browser. Gain indicates percentage increase over time compressed with no pause removal. The first column is calculated by taking the total length of video watched divided by the total actual viewing time. The effects of pause removal are added by including the length of the deleted pauses into the total length of video.

	Time Comp.	Time Comp. and Pause Removed (Gain)
Classroom	123.4%	137.1% (11.1%)
Conference	122.0%	147.1% (20.6%)
Sports	116.8%	137.1% (17.4%)
Shows	132.6%	146.1% (10.2%)
News	117.8%	138.5% (17.5%)
Travel	132.0%	138.9% (5.2%)
Overall	124.1%	140.8% (13.5%)

Table 4. Average Percentage of Video Watched. This table shows the average percentage of a video watched (%W) and decomposes that value into the percentage of video watched only once (%W1), exactly twice (%W2), and three or more times (%W+). The highlighted entries show that nearly 20% more of a video was watched with the Enhanced browser than with the Basic browser.

	Basic				Enhanced			
	%W	%W1	%W2	%W+	%W	%W1	%W2	%W+
Classroom	33.0	32.5	0.5	0.0	48.2	41.1	6.3	0.8
Conference	64.4	62.6	1.6	0.2	86.1	75.2	10.0	0.9
Sports	21.8	20.7	1.1	0.0	41.3	34.1	5.6	1.6
Shows	40.5	40.3	0.2	0.0	53.8	53.0	0.8	0.0
News	35.0	33.5	1.5	0.0	51.9	46.7	4.8	0.4
Travel	26.5	20.1	5.7	0.7	57.2	30.9	11.6	14.7
Overall	36.9	35.0	1.7	0.2	56.4	46.8	6.5	3.1

Table 5. Average Percentage of Task Time Spent in Playback Modes. Using the Enhanced browser participants spent considerably less time watching video at normal playback speeds than with the Basic browser. Playing = Normal playback speed, FF = Fast forward, TC = Time compressed, PR = Pause removed.

	Paused		Playing		FF		TC	PR	TC&PR
	Bas	Enh	Bas	Enh	Bas	Enh	Enhanced		
Classroom	15.7	12.5	74.4	33.4	9.9	0.4	24.7	5.6	23.4
Conference	14.4	16.1	83.4	9.8	2.2	0.0	13.4	2.6	58.1
Sports	9.1	4.5	53.2	35.8	37.7	10.7	21.1	6.5	21.4
Shows	15.7	4.8	72.0	21.5	12.3	0.9	33.0	0.2	39.6
News	10.8	10.0	73.9	10.8	15.3	0.0	22.5	12.3	44.4
Travel	14.3	21.5	67.2	21.6	18.5	0.0	23.9	1.1	31.9
Overall	13.3	11.6	70.7	22.2	16.0	2.0	23.1	4.7	36.5

Classroom Lecture

Many educational institutions videotape courses for archival and rebroadcast. Stanford University, for example, offers hundreds of courses each year, both live and on-demand, via television broadcast, videotape, and Internet delivery [20]. In the classroom lecture scenario, the participants were asked to imagine they were taking a C programming class. A quiz was going to be administered in ½ hour but they did not attend the previous one-hour lecture. The task was to watch an archived video of the lecture and summarize the main points of the content.

The time constraint ensured that the participants would not be able to watch the entire video. However, the participants were selected based upon previous programming experience in a language other than C. Since many programming concepts are similar across different languages, it was presumed that the participants could effectively skim the video based upon previous knowledge.

Using the basic browser, though, the participants had a difficult time skimming the video. The participants fast-forwarded through topics and skipped topics using the seek thumb. However, with no indication of the position of topic changes, the participants made random guesses to seek. Table 2 shows they used the seek thumb an average of 21.6 times in the half hour, or roughly once every 1.5 minutes.

Using the enhanced browser, the participants in this scenario used the TOC to seek the video with greater frequency than any other scenario (avg. use 12.5 times, Table 2). The TOC was generated from slides used in the lecture. In addition, they also made considerable use of TC and PR. This increased the fraction of content they watched once or more from 33.0% to 48.2% (Table 4), corresponding to an effective playback speed of 137.1% (Table 3). The TOC, TC, and PR were the highest rated controls (6.8, 5.4, and 5.1 respectively, Table 1).[1]

The basic browser interface is not unlike that of the VCRs that many Stanford students use in campus libraries to watch videotaped courses. Providing a TOC index as well as TC and PR could make video a more useful and efficient tool for reviewing courses.

Conference Presentation

Conferences and seminars are valuable means for keeping up with the latest trends. Electronically accessible on-demand presentations provide the flexibility of anytime, anywhere viewing. We believe browsing capabilities can potentially be of great value when time is limited and as the number of presentations increases.

The participants were asked to pretend they had ½ hour before attending a meeting with co-workers to discuss a conference they had attended. The participants did not attend the same presentations as their co-workers, but would still like to take part in the discussion. The task was to review a video of a missed presentation and summarize the main points in preparation for the meeting.

The videos were selected from the ACM 97 presentations of "The Next 50 Years of Computing" and ranged in length between 40–50 minutes. Participants were recruited based upon background interests in the future of computing and education. Unlike the classroom lecture scenario, the contents of videos were not technical or highly structured, so a TOC was not generated for the enhanced browser.

Using the basic browser, the participants used the seek thumb and fast forward to skim the video much like in the classroom scenario.

Using the enhanced browser, the highest rated controls were TC and PR (6.9 and 6.5, Table 1). On average, an effective playback speed of 147.1% was attained by the participants (Table 3) and, as compared to the basic browser, they covered 86.1% of the content instead of 64.4% (Table 4). Shot boundary frames were used twice on average, usually to skip lengthy introductions as the transition between the host and the speaker could be seen in the frames.

Although the average rating was neutral (3.8, Table 1), personal notes were used by several participants. Two of the five participants used notes to mark interesting locations in the video. One of them included the shot boundary frame number in the title of her notes, providing a visual indicator for their location. Both participants used their notes to review the main points of the video for their summary. A third participant used the notes feature to bookmark the start and end of video segments he skipped to review them later if time allowed. This behavior suggests the need for a quick bookmark feature that does not require typing a title for a note and/or a logging feature that can automatically marks the portions of a video skipped.

Overall, as in the classroom scenario, TC and PR made it possible to watch substantially more of the presentation in the limited time available. In the absence of a TOC, shot boundaries and notes were utilized to effectively mark and jump to locations in the video.

Sports

Sports programming is one of the most popular forms of video entertainment. In the sports scenario, we wanted to see how participants would react to the added ability to browse and skim events. Each participant reported that they watched sports or sports news shows regularly.

The task was to find highlights in a baseball game to discuss with friends at the health club in ½ hour. A single baseball game was divided into three one-hour segments and presented in order to the participants. Since baseball can have long periods of little or no scoring activity, it was expected that there would be ample opportunity to skim the video. As an aid, a TOC was generated for the enhanced

[1] Although "Seek" is rated high in Table 1, notice that it is used zero times in the enhanced browser (Table 2). The high rating is due to the fact that the participants thought of TOC also as a seek mechanism.

browser indexing the top and bottom of each inning in the video (about 6 entries per video).

Using the basic browser, most of the participants started out by using fast-forward to skip commercials and dead time between plays. The participants spent an average of 37.7% of their time watching the game in fast-forward (Table 5), more than any other scenario. Play highlights can be identified visually, so the lack of audio during fast-forward was not a deterrent. Fast-forward, however, was not enough to skim the video in ½ hour. As a result, the participants also frequently used the seek thumb (average 15.7 times, Table 2).

Using the enhanced browser, the participants most frequently used the shot-boundary frames to seek the video (average 26.5 times, Table 2) and rated it highest in surveys (average 6.1, Table 1). Using the five frames at the bottom of the browser, the participants could determine the outcome of the current play. By scrolling the frames ahead, the participants could preview and seek to successive plays. In contrast, the TOC inning index was only used once or twice, mainly to skip the ads at the end of an inning.

TC, PR, and fast-forward were also very popular in the enhanced browser. Unlike other scenarios, fast-forward was still useful as it allowed greater speed-up than time compression, and key information was in the video channel anyway. TC and PR combined resulted in an effective playback speed of 137.1% (Table 3).

In this scenario, we saw the development of more sophisticated strategies over time. For example, when watching the second video using the enhanced browser, two participants chose to watch the home team at bat while *completely* skipping the visitors. Another two participants used the notes feature to bookmark interesting plays for later reference. Both strategies demonstrate the user being in control over the game, unlike watching a set of highlights from a news show.

When asked if the availability of the enhanced browser would affect how they watched television, the average response increased from 4.2 (neutral) using the basic browser, to 6.0 (agree) after the second use of the enhanced browser (scale of 1–7, 7 being strongly agree). Similarly, when asked about the quality of their experience, ratings increased from 4.8 to 6.0 (scale of 1–7, 7 being best).

The participants in this scenario found the enhanced browser to be a useful tool and enjoyed being able to browse the game. Features that support skimming visually, such as shot boundaries, were more useful here than in previous scenarios. TC and PR continued to be of high value too.

Shows

As in the sports scenario, we wanted to see how participants would react to the ability to browse a one-hour television show. Each participant regularly watched at least one weekly television show. They were asked to pretend that the final episode of their favorite show was airing in ½ hour, but they still needed to watch the previous episode that they had recorded.

The task was to review the major events in the show before watching the final episode. Each participant watched a full episode of "E.R.," "Ally McBeal," and "Babylon 5" (including commercials). We knew that the browsing features would be used to skip commercials. However, how each participant might choose to browse the content of the shows might depend heavily upon personal preference.

Using the basic browser, it was not possible for the participants to watch the entire show in ½ hour even if they skipped commercials. The seek thumb was used 14 times on average, or roughly one seek every 2 minutes (Table 2). The participants reported that they seeked randomly.

Using the enhanced browser, TC was the highest rated feature (6.0, Table 1). It was used to increase the amount of the show watched from an average of 40.5% in the basic condition to 53.8% over the enhanced conditions (Table 4). The second highest rated feature was shot boundaries (5.1, Table 1). By scrolling the shot boundary frames, the participants could instantly and accurately skip commercials. The average use of 5 shot boundary seeks corresponds roughly to the number of commercials in a one-hour show (Table 2). Otherwise, the participants did not exhibit any particular browsing behavior.

When asked to rate satisfaction of coverage of the show, the participants reported an increase from 3.4 using the basic browser to 5.4 after the second use of the enhanced browser (scale of 1 to 7, 7 being best coverage). However, unlike the sports condition, the participants did not agree that the availability of a video browser would affect the way they watched television, reporting an average 3.6 for the basic browser and 4.3 for enhanced browser (scale of 1 to 7, 7 being strongly agree). The participants all reported that they would not regularly watch a show under such time constraints. One participant complained that watching a show with TC and PR was "mentally fatiguing."

News

The participants in the news scenario were asked to pretend that they were forced by family members to spend less time watching the news. The task was to watch a one-hour news show in the ½ hour before dinner and summarize the program for discussion at the table. Each participants reported that they watched at least ½ hour of news daily.

The participants were presented three consecutive airings of "The News Hour with Jim Lehrer" which consists of a general news summary followed by five in-depth reports. Since the content is highly structured into discrete story segments, we expected that the participants would want to choose the stories they were interested in watching. A TOC was generated for the enhanced browser to index the beginning of the news summary and each story.

Using the basic browser, the seek thumb was used heavily (34.0 times, Table 2). The participants had to make many guesses to find the beginnings of stories in the video.

Using the enhanced browser, participants were able to use TC and PR to watch more of the video (35.0% watched with basic vs. 51.9% with enhanced, Table 4). In addition, the TOC made it possible for participants to "select which one [story to watch] or in which order I watched them". Like the classroom scenario, TC, PR, and the TOC were the highest rated features (6.7, 6.6, 6.6, respectively, Table 1).

Unlike the classroom scenario, though, shot boundary frames proved to be a useful preview feature for the participants as they watched the video (average rating of 6.4 versus 5.0 for classroom, Table 1). Participants would scroll the frames to get an overview of the contents of a story, using the jump-next button or clicking on a frame to skip ahead.

Ultimately, all the participants felt that they could better cover the news program using the enhanced browser, with an average satisfaction of coverage rating of 6.6 versus an average rating of 4.8 in the basic (scale of 1–7, 7 being best coverage). When asked if a video browser would affect the way they watched television, the participants were more enthusiastic than those in the sports scenario, rating an average of 6.9 (scale of 1–7, 7 being strongly agree).

Overall, as in the sports scenario, the participants enjoyed the additional control the browser afforded them. News is a very rich video content type, and browsers can take advantage of both textual and visual indices for searching as well as TC and PR for saving time.

Travel

Travel videos are often used to preview destination getaways. The participants in this scenario were asked to form a five minute summary of a travel video by identifying the begin and end points of interesting clips. The summary would be used as a potential travel itinerary to convince their families where they wanted to go on their vacation.

Each participant reported an interest in travel as well as having planned or taken a vacation in recent years. The travel videos contained tourist points of interest and used narrator voice-overs to describe the scenes.

Using the basic browser, the seek thumb was used nearly twice as often as in the next most used scenario (average 51.8 times vs. 34.0 for news, Table 2). The greater accuracy needed for finding the begin and end points of clips required many adjustments using the seek thumb.

In the enhanced condition, the participants relied on the shot boundary frames to navigate the videos, using them to identify interesting looking destinations. They used the shot boundary frames to seek an average of 55.0 times versus an average of 16.5 over all the scenarios (Table 2) and rated it the third most useful feature (6.3, Table 1).

The notes were invaluable for marking the start and end points of clips, receiving its highest rating across the scenarios (6.4 versus 4.1 overall, Table 1). An average of 9.5 notes were added by each participant versus 2.6 overall (Table 2). They often positioned their notes by hitting the jump-back button after noticing an interesting landmark. Jump-back was also used the most (14.5 times, Table 2) and rated the highest in this scenario (6.3, Table 1).

Ultimately, the participants rated TC the highest in this scenario (6.6, Table 1). TC and PR made it possible to watch almost twice as much of the video, increasing from 26.5% using the basic browser to 57.2% using the enhanced (Table 4). When asked to rate the quality of their itinerary summaries, the participants reported an increase from 4.4 using the basic browser to 5.8 by the second use of the enhanced browser (scale of 1–7, 7 being best).

As for other scenarios, TC and PR were used very effectively. However, the rich visual content of the videos also allowed for effective shot boundary based navigation. In addition, with the added requirement of precise positioning, the participants found notes and the jump buttons very useful.

CONCLUDING REMARKS

Overall, the participants exhibited substantially different viewing behavior when they used the enhanced video browser. For example, the traditional seek and fast-forward controls were almost never used. Instead, features like the table of contents and shot boundaries were used to more accurately jump to locations in the video. In addition, the participants spent a substantial amount of time watching the video with time compression and pause removal, increasing the amount of video they watched by 20%.

In the sports and news scenarios, the participants enthusiastically agreed that having enhanced browsing features would affect the way they watched television. In these scenarios, we observed content-specific browsing behavior using the enhanced features such as watching only the home team of a sports game or choosing the order of stories to watch in a news show.

Based on such patterns of feature use and experience across the scenarios, we can also informally classify our six video content types into different categories: informational audio-centric, informational video-centric, and narrative-entertainment.

Informational audio-centric videos like classroom lectures and conference presentations contain most of their content in the audio channel and usually have little visual activity. As such, visual browsing features like shot boundary frames provide minimal cues. For structured content, a TOC provides a valuable index, although users can take advantage of notes and shot boundaries to form their own ad-hoc index when it is unavailable.

With *informational video-centric* content like travel and sports videos, the rich video content makes shot boundary frames an effective navigation tool. When combined with notes and the jump-back button, it was possible to accurately position the video. News can fall equally into both the informational audio-centric and informational video-centric categories, and can take advantage of a combination of the different indices for effective browsing.

When watching *narrative-entertainment* like television dramas, the viewing experience was affected when the participants were forced to use browsing features like TC and PR. One participant succinctly stated the general sentiment: "I saved time but I would seldom want to watch a show in a fast version."

However, when watching news and sports, the participants reported the opposite response. A sports participant remarked that "anything to remove excess time from viewing is positive." A news participant went further to say that "saving time isn't the best part—being in control is". The features provided the ability to "move to what interested me most and then return to the other segments as time permitted."

In the travel scenario, the participants believed that the enhanced features could be useful for editing. When asked about the technology, one participant responded: "It's exciting. I think editing home movies would be fun." Another remarked, "I would buy this software in a minute if it would allow me to edit video."

In general, we are greatly encouraged by the participants' positive reaction to the enhanced browser. However, this study is only the first step in evaluating the potential of these advanced browsing and skimming features. We need to look at the results from a larger number of subjects, ideally in more natural task environments where we can measure usage over a longer period of time. Having performed this broad study across the six different content types, we can now design for specific video browsing tasks with a better understanding of how these features can be applied and evaluated.

ACKNOWLEDGMENTS

We would like to thank Asta Glazer and Cindy Solomon for helping with the initial design of the browser prototype.

REFERENCES

1. Apple QuickTime Player, Apple Corporation Inc., http://www.apple.com/quicktime/
2. Arman, F., Depommier, R., Hsu, A., and Chiu, M.-Y. "Content-based browsing of video sequences." In *Proceedings of the second ACM international conference on Multimedia '94* , 1994, Page 97
3. Arons, B. "SpeechSkimmer: A System for Interactively Skimming Recorded Speech." *ACM Transactions on Computer Human Interaction, 4,* 1, 1997, 3-38.
4. Arons, B. "Techniques, Perception, and Applications of Time-Compressed Speech." In *Proceedings of 1992 Conference,* American Voice I/O Society, Sep. 1992, pp. 169-177.
5. Brotherton, J. A., Bhalodia, J. R., and Abowd, G. D. "Automated Capture, Integration, and Visualization of Multiple Media Streams." In the *Proceedings of IEEE Multimedia '98*, July, 1998.
6. Chistel, M. G., Smith, M., Taylor, C. R., and Winkler, D. B., "Evolving video skims into useful multimedia abstractions". In *Proceedings of CHI '98* (Los Angeles, CA, 1998), ACM Press, 171-178.
7. DVD Video Group, http://www.dvdvideogroup.com/
8. He, L., Sanocki, E., Gupta, A., and Grudin, J., "Auto-summarization of audio-video presentations". In *Proceedings of the Conference on ACM Multimedia 99*, 1999, Pages 489-498.
9. Informedia, http://www.informedia.cs.cmu.edu/
10. Komlodi, A. and Marchionini, G. "Key frame preview techniques for video browsing." In *Proceedings of the third ACM Conference on Digital libraries*, 1998, Pages 118 – 125.
11. Low, C. Y., Tian, Q., and Zhang, H. "An automatic news video parsing, indexing and browsing system." In *Proceedings ACM Multimedia '96*, 1996, Page 425.
12. MediaSite, http://www.mediasite.com/
13. Meng, J. and Chang, S. "CVEPS - a compressed video editing and parsing system." In *Proceedings ACM Multimedia '96*, 1996, Page 43.
14. Microsoft Windows Media, Microsoft Corporation Inc., http://www.microsoft.com/windows/windowsmedia/
15. Mills, M., Cohen, J., and Wong, Y. Y., A magnifier tool for video data, in *Proceedings of CHI '92*, 1992, ACM Press, 93-98.
16. Omoigui, N., He, L., Gupta, A., Grudin, J., and Sanocki, E., Time-Compression: Systems Concerns, Usage, and Benefits, in *Proceedings of CHI '99* (Pittsburgh, PA, 1999), ACM Press, 136-143.
17. Ponceleon, D., Strinivasan, S., Amir, A., Dragutin, P., and Diklic, D. "Key to effective video retrieval: effective cataloging and browsing." In *Proceedings of the 6th ACM international conference on Multimedia,* 1998, Pages 99 – 107.
18. Real Networks RealPlayer, http://www.real.com/
19. Replay Networks ReplayTV, http://www.replaytv.com/
20. Stanford Online, http://stanford-online.stanford.edu/
21. Stifelman, L. "The Audio Notebook: Paper and Pen Interaction with Structured Speech" *Ph.D. dissertation, MIT Media Laboratory,* 1997.
22. TiVo Inc., http://www.tivo.com/
23. Virage, http://www.virage.com/