Browsing Digital Video

Francis Li
Anoop Gupta
Elizabeth Sanocki
Liwei He
Yong Rui

September 20, 1999

Technical Report
MSR-TR-99-67

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Browsing Digital Video

Francis Li
Group for User Interface Research, CS Dept.
University of California, Berkeley
Berkeley, CA 94720-1776
fli@cs.berkeley.edu

Anoop Gupta, Elizabeth Sanocki, Li-wei He, Yong Rui
Microsoft Research
Redmond, WA 98052
{anoop, a-elisan, lhe, yongrui}@microsoft.com

ABSTRACT

Video in digital format coupled with digital/programmable playback devices presents opportunities for significantly enhancing the user's viewing experience. For example, time compression can shorten the viewing length of a video and shot boundary frames can provide a visual index into the content. Such features have primarily been evaluated in isolation with a narrow set of video content types. We investigated as well as implemented the design of a software video browsing application that combines many such features. In addition, we evaluated its use in watching six different video content types and present the resulting data for analysis and discussion. The participants in the evaluation found the browser to be useful and effective for watching the different types of video in a limited amount of time. Also, the results show that both the experience of using the browser and value of each feature varies depending on the content type.

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 1970-80s. The VCR makes it possible to watch a video with the additional ability to 1) pause the video and 2) fast-forward or rewind the video for skipping or re-watching particular segments. Seeking to a random location is possible, but has a large delay associated with it due to the use of tape storage.

Today, Internet video streaming and set-top devices like ReplayTV [18] and TiVo [21] are technologies that are defining a platform for more interactive video playback. Unlike traditional VCRs, ReplayTV and TiVo devices 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. This allows indices into the content such as the chapter lists of digital versatile disc (DVD) videos [7]. In addition, as computing costs continue to drop, processing techniques can be utilized to automatically generate such indices or shorten the viewing length of a video without losing content. Such features can potentially allow a viewer to save significant amounts of time watching a video as well as more effectively filter the content during playback.

In exploring the ability to browse digital video, we considered the following questions:

- What are potential high-value features that we can 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?
- How does the usage affect the enjoyment and/or other factors of the viewing experience?
- What should the interface be for an application or device that provides these advanced features?

This paper attempts to answer the questions raised above. 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 indexes 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 to the video, and other advanced browsing controls. Many of these features have been studied previously, but primarily in isolation within a narrow set of video content types. We evaluated the combined use of these features using our prototype across six different video content types and tasks: classroom lectures, conference presentations, sports, television dramas, news, and travelogues. The results of this study are presented in this paper for analysis and discussion within the context of the above questions.

In the next section, we discuss related work in browsing digital media. Then, in the following section, we describe the design of our prototype software video browsing application. This is followed by a description of the general experimental method used in the evaluation. We then detail the six video content types and present the results of the study. Finally, we present our conclusions.

RELATED WORK

Previous research in browsing digital media has often focused on either audio or video, but not both. The SpeechSkimmer provided an interface for selecting time compressed and pause-removed audio playback as well as facilitating jumping back and forward between pre-defined segments of the recording [3,4]. The Audio Notebook [20] uses pen strokes to index audio as it is recorded and allows time compressed playback.

For video, the Hierarchical Video Magnifier [14] was designed to provide users with a context of the contents of a video by displaying video frames nearby the current position. Arman, et al [2] improved the frame selection methods by detecting shot-boundaries, useful in editing systems [12]. The Classroom-2000 project at Georgia Tech [5] investigated richly indexed videos of lectures, including indexing based on strokes drawn on a black-board. None of these systems explore the wide range of browsing techniques and/or user scenarios explored here.

Christel, et al, describe an evaluation of techniques for shortening the viewing time of a video based on both audio and video analysis [6]. Such techniques, used in systems like CueVideo [16], condense the content into a shortened video summary that is 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 [8] 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 [21] and MediaSite [11] are providing these services for finding video on the Internet. Others have used domain knowledge to improve such services for specific video content types like news [10]. 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 [17], Apple QuickTime Player [1], and Microsoft Windows Media Player [13] 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 into the content using a table of contents index. ReplayTV and TiVo set-top boxes offer an index to the shows recorded. In addition, they provide the ability to jump forward by 30 or 60 seconds, primarily intended for skipping commercials, and back 8 – 10 seconds for "instant replays". However, none of these devices 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 device. Finally, no public data is available on how the provided controls 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 the pauses in continuous speech and removes both the audio and video segments associated with them.

The textual indices provide the user with a means to browse the contents of a video in the same manner that they might browse a text document. The user can seek to the location in the video that is associated with a particular entry. The *table of contents* (TOC) is used to provide a pre-generated list of entries that cannot be modified. The *notes* feature allows the user to create their own entries as well as add longer text comments to each entry. When the user creates a note, the video is paused and the title and 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

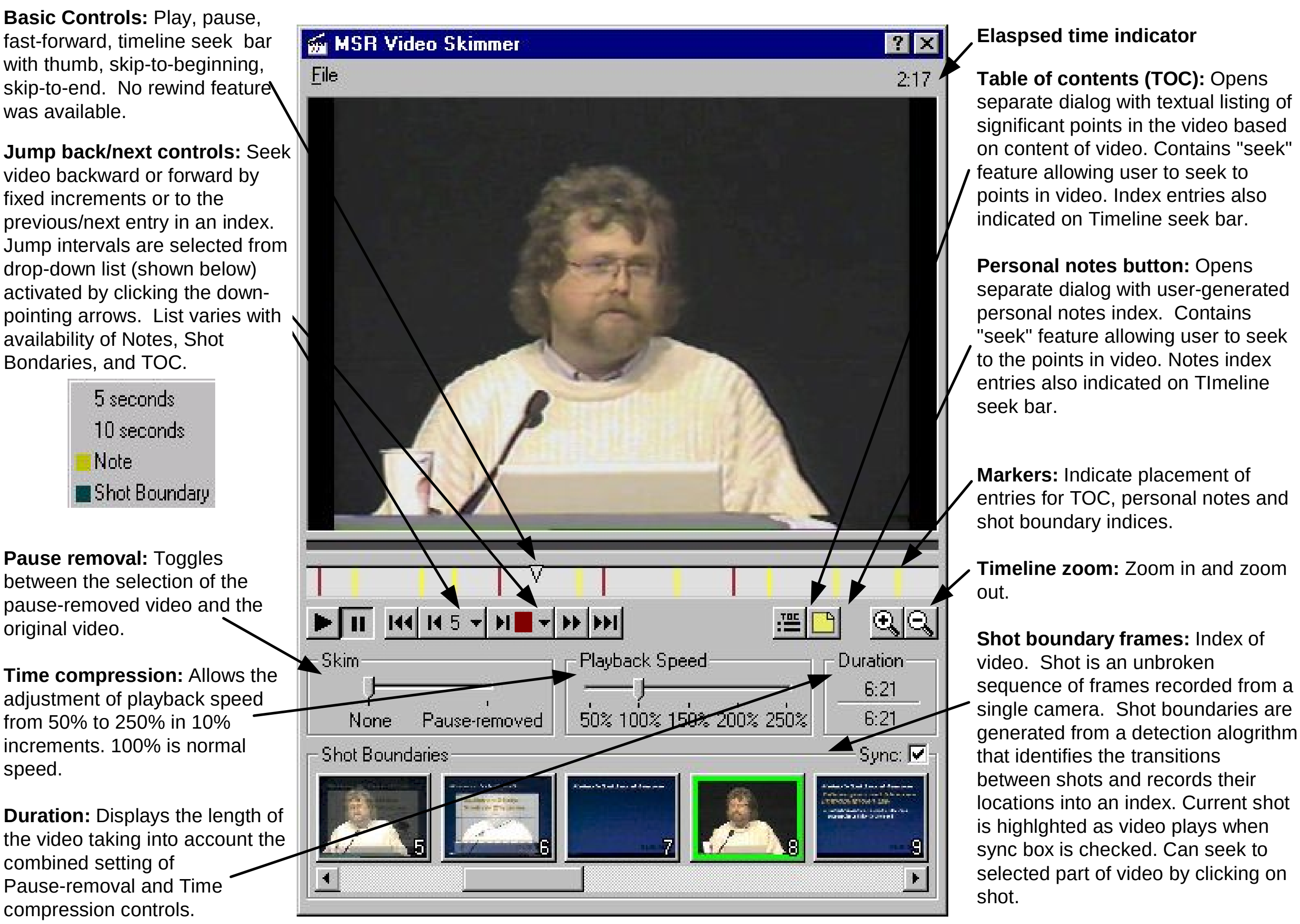

Figure 1. Enhanced Browser User Interface

reference as well as to record their thoughts regarding the content of the video at that location.

The visual indices provided are the shot boundary frames and the timeline markers. The numbered *shot boundary frames* allow the user to visually identify and then seek to a 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 in the video with color coded bars. They can be used to judge the locations 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 by 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 watching the video might use the jump back 5 and 10 seconds controls to repeat significant events just passed 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 the functionality of the browser with a user interface adequate for evaluation. Although not discussed in this paper, the study was also used to evaluate the usability of the interface. Both the basic and enhanced browsers were instrumented to record the usage of each feature during the study.

STUDY DESIGN

The user study was designed to evaluate feature usage and the experience with the enhanced browser via observation, subjective surveys, and comparison with the basic browser. In addition, we chose to conduct the study across a broad range of content types, ultimately choosing six such "browsing scenarios". The scenarios included watching classroom lectures, conference presentations, sports, television dramas, news, and travelogues. We detail the scenarios with the presentation of results in the next section.

Participants were recruited from a pool of non-Microsoft employees that expressed interest in participating in a usability study at Microsoft. In addition, the participants were selected and assigned to a scenario based upon

	Seek		FF		SB	TC	PR	Jmp	TOC	Note
	Bas	Enh	Bas	Enh	Enhanced					
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
Average	5.3	4.8	4.8	4.4	5.6	6.2	5.5	5.0	6.2	4.1

SB = Shot Boundaries, Jmp = Jump-back and Jump-next

Table 1. Qualitative Ratings of Browser Features

matching interests with the scenario content. Five participants per scenario completed the study for a total of 30 participants. Each participant received a Microsoft software product for their involvement in the study.

The participants were assigned tasks related to their browsing scenario. Each participant first completed their task watching a video using the basic browser. Then, after completing a practice task to learn the enhanced features, they watched two more videos using the enhanced browser. To encourage the use of the browsing features, they were limited to ½ hour for watching each 45 min. – 1 hour video.

In addition to pre- and post-study surveys, the participants completed a survey after watching each video. The participants were asked to describe their browsing strategy as well as rate their interest in the contents of the video, the quality of their experience, and the usefulness of the features available in each condition.

SCENARIOS AND RESULTS

We present the results with the descriptions of the six browsing scenarios, including quantitative data on what features were used most and a qualitative analysis of the users' browsing experience.

First, we present tables that we will be using in discussing the results for scenarios. Table 1 presents the users' qualitative ratings of various features provided in the basic and enhanced interfaces on a scale of 1…7, where 1 is "not useful at all" and 7 is "very useful". Table 2 shows the frequency of use of the features across the scenarios. Table 3 shows the overall playback speed using time-compression and pause-removal across the scenarios. Table 4 shows what fraction of content the users watched 0, 1, 2, or 3+ times. Finally, Table 5 shows how the users utilized their time during the study, i.e., with the video paused, playing at normal speed, fast-forward (FF), playing time-compressed (TC), pause-removed (PR), or both (TC, PR).

Classroom Lecture

Increasing resource demands on education have led to the adoption of video offering of courses by many institutions. Stanford University, for example, offers hundreds of courses each year, live and on-demand, via television broadcast, videotape, and Internet delivery [19]. The classroom lecture scenario simulates a student taking a traditional live course with a video archive. 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 the lecture video and summarize the main points in preparation for the quiz.

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. Figure 2 shows they used the seek thumb an average of 21 times in the half hour, or roughly once every 1.5 minutes.

The enhanced browser provided a TOC generated from slides used in the lecture. Participants in this scenario used the TOC to seek the video with greater frequency than any other scenario (avg. use 12.5 times vs. 2 times overall, Table 2). They reported that they "used the TOC to jump to the main parts of the lecture rather than guessing". They also made considerable use of TC and PR. This increased the fraction of content they watched once or more from 35% to 48% (Table 4), corresponding to a combined speed-up factor of 1.37 (Table 3). The TOC, TC, and PR were the top-three valued controls, with ratings of 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

	Seek		FF		SB	Jmp Bck	Jmp Nxt	TOC Sk	Note Add	Note Sk
	Bas	Enh	Bas	Enh	Enhanced					
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
Average	26.3	2.3	9.8	0.9	16.5	3.6	6.5	5.8	2.6	1.1

SB = Shot Boundary Seek, TOC Sk = TOC Seek, Note Sk = Note Seek

Table 2. Avg. Feature Use per Participant per Video

[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.

	Time Comp.	Time Comp. And Pause-Removed (Gain)
Classroom	126.6%	136.8% (10.2%)
Conference	122.0%	150.4% (28.3%)
Sports	116.9%	137.3% (20.4%)
Shows	132.3%	144.6% (12.3%)
News	123.3%	142.2% (18.9%)
Travel	139.9%	149.5% (9.6%)
Average	126.8%	143.4% (16.6%)

Table 3. Avg. Playback Speed in Enhanced Conditions

watch televised courses. The initial findings of this scenario suggest that significant benefits could be gained by providing a TOC for courses, if not also TC and PR.

Conference Presentation

Conference and seminar presentations are valuable for keeping up with contemporary work in various academic and professional fields. Electronically accessible on-demand presentations can provide added flexibility of anytime, anywhere viewing. However, the ability to browse a presentation can potentially be of great value when time is limited.

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 the missed presentation and summarize the main points in preparation for the meeting.

The videos for the study ranged between 40 to 50 minutes and were selected from the ACM 97 presentations of "The Next 50 Years of Computing". 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 provided for enhanced browser.

Using the basic browser, the participants used the seek thumb and the 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, a combined speed-up of 1.5 was used by the participants (Table 2) and, as compared to the basic browser, they covered 86% of the content instead of 68% (Table 3). 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 effectively 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 the location of the note. 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 bookmark feature that does not require typing a title for a note or a logging feature that automatically marks the portions of the video skipped.

TC and PR made it possible to watch significantly more of the video, as in the classroom scenario. However, without a TOC, both the shot boundaries and the notes were utilized to effectively browse the video.

Sports

These days, twenty-four hour networks bombard us with a wide array of sports programming. However, the time available to watch sports has not increased. The sports scenario gave participants the chance to browse sports events. Each participant reported that they watched sports or sports news shows on a regular basis.

The specific 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 was ample opportunity to skim the video. As an aid, a table of contents was provided in the enhanced condition indexing the top and bottom of each inning in the video (~6 entries).

Using the basic browser, most of the participants started out by using the fast forward button to skip the commercials and dead time between plays. The participants spent nearly 40% of their time watching the game in fast-forward (Table 5), higher than any other scenario. Play highlights can be identified visually, so the lack of audio was insignificant. Fast-forward, however, was not enough to skim the game in ½ hour. As a result, the participants also made considerable use of the seek thumb (~15 times in 30 minutes).

With the enhanced browser, the participants most frequently used the shot-boundary frames to seek the video (~27 times in 30 minutes, Table 2) and rated it highest in

	Basic				Enhanced			
	%W0	%W1	%W2	%W+	%W0	%W1	%W2	%W+
Classroom	64.2	32.8	2.6	0.2	51.9	41.1	6.6	0.9
Conference	32.2	64.5	2.7	0.8	14.0	73.9	11.0	0.9
Sports	78.2	20.8	1.0	0.0	58.8	33.3	6.3	1.7
Shows	59.4	40.4	0.0	0.0	46.2	52.9	0.9	0.0
News	63.4	33.8	2.4	0.0	48.5	46.3	5.1	0.4
Travel	66.8	25.3	7.3	0.5	42.9	30.8	11.5	15.1
Average	60.7	36.3	2.7	0.3	43.7	46.4	6.9	3.2

Table 4. Percent of video not watched (%W0), watched once (%W1), twice (%W1), 3 or more times (%W+)

	Paused	Playing	FF	TC	PR	TC, PR
Classroom Basic	6.0	86.4	11.0			
Enhanced	11.1	27.5	0.3	19.3	5.5	26.4
Conference Basic	10.2	90.0	2.2			
Enhanced	14.7	10.2	0.1	13.5	2.4	54.6
Sports Basic	8.6	54.0	38.0			
Enhanced	4.6	35.9	10.7	21.2	6.4	22.2
Shows Basic	15.4	72.2	12.6			
Enhanced	4.6	18.8	0.9	29.6	0.2	36.0
News Basic	7.4	79.4	17.0			
Enhanced	9.6	10.8	0.0	23.6	12.4	44.5
Travel Basic	11.8	53.8	15.2			
Enhanced	19.3	17.5	0.0	19.7	1.3	27.7

Table 5. Percentage of Study Time Spent

surveys (6.1, Table 1). Using the five frames at bottom of 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 remained quite attractive as it allowed greater speed-up than time compression and key information was in the video channel anyway. TC and PR combined offered a speed-up of 1.37 allowing more of the game to be watched.

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 exemplify the user 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 participants' responses increased from 4.2 using the basic browser (neutral), to 6 after the second use of the enhanced browser (agree, scale of 1 – 7, 7 being strongly agree). Similarly, when asked about the quality of their experience, ratings increased from 4.8 to 6 (scale of 1 – 7, 7 being best).

The results show that having the ability to browse and skim a baseball game can potentially be very appealing. Features that support skimming visually, such as shot boundaries, TC, and PR, are far more useful than others in this scenario.

Shows

Every day, millions of viewers watch the countless number of sitcoms, soap operas, dramas, and other shows that fill the airwaves. The VCR has proven to be an indispensable aid in allowing viewers to skim and browse their recordings, primarily through skipping advertisements. How would they react to the features in our application?

Each participant regularly watched at least one weekly television show. They were asked to pretend that they wanted to watch the final episode of their favorite show 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).

Few expectations were made regarding the browsing behavior of the participants in this scenario. It was an absolute certainty that the features would be used to skip commercials. However, how each participant might choose to browse the content of the shows could 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 guessed randomly when seeking.

In the enhanced conditions, time compression was the highest rated feature of the browser (6, Table 1). It was used to increase the amount of the show watched from an average of 40% in the basic condition to 54% 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 (Table 2) corresponds roughly to the number of commercials in a one-hour show.

When asked to rate their satisfaction with their 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 – 7, 7 being best). However, unlike the sports condition, the participants did not agree that the availability of a video browser would affect the way they watched television (3.6 in basic, avg. 4.3 over enhanced). The participants all reported that they would not regularly watch a show under such time constraints. One participant called time compression and pause removal "mentally fatiguing".

News

Like sports, news is available 24 hours a day, 7 days a week on television and the Internet, but the time for watching them has not increased. The participants in this 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 program in the ½ hour before dinner and summarize the news for discussion at the table. Participants were selected based upon a prerequisite of watching 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. To facilitate this behavior, a table of contents was generated for the enhanced conditions to index the beginning of the news summary and each story.

Using the basic browser, the seek thumb was used heavily (35 times, Table 2) to skip. 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 (37% watched with basic vs. 52% 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 (rated 6.4, 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 on a scale of 7, 7 being best, versus an average rating of 4.8 in the basic condition. When asked if a video browser would affect the way they watched television, the participants were even more enthusiastic than those in the sports scenario, rating an average of 6.9 on a scale of 7 (strongly agree).

Overall, the results indicate that news is a very rich video content type and that browsers can take advantage of both textual and visual indices for searching as well as TC and PR for saving time. And, the overwhelmingly positive subjective results show that many people could immediately benefit from such features.

Travel

Travel videos can be effective previews for destination getaways. The participants were asked to construct a five minute summary of a travel video that outlined a potential itinerary for a vacation. The summary would be used to convince their families where they wanted to go on their vacation. The travel video scenario was designed to evaluate the use of the advanced features in a simple editing task- identifying and assembling clips into a sequence.

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 contents of the scenes.

Using the basic browser, the seek thumb was used nearly twice as often as in the next most used scenario (64 times vs. 34 for news, Table 2). The greater accuracy needed for defining 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 the video an average of 55 times versus an average of 16.5 over all 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 vs average 4.1, Table 1). An average of 9.5 notes was added by each participant versus 2.6 overall (Table 2). They 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 here across the scenarios (6.25, Table 1).

Ultimately, the participants rated TC the highest in this scenario (6.6, Table 1). TC and PR made it possible to watch nearly 25% more of the video, increasing from 33% with the basic browser to 57% with 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).

These early results indicate that casual users find the combination of different features very useful in a simple editing task. Yet some of these features have yet to be found even on professional editing packages.

CONCLUDING REMARKS

The widespread adoption of Internet streaming video and the development of devices like ReplayTV and TiVo present an unprecedented opportunity to provide new features for browsing digital video. We investigated and implemented the design of a software video browsing application that included features like time compression, pause removal, and different forms of content indices. In addition, we evaluated this design using six different video content types and presented the resulting data for analysis and discussion.

Using the enhanced browser, the participants viewed nearly 20% more of the video using TC and PR (average video watched for basic browser condition = 39% vs. average watched for enhanced = 57%, Table 4). They extensively used the shot boundary frames (overall average 16.5 times, Table 2) in place of the seek thumb to advance in the video. Based on the individual results of each scenario, we can informally classify our six video content types into different categories: informational audio-centric, informational video-centric, 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 in the content. 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 when it is unavailable.

With i*nformational video-centric* content like travel and sports videos, the video contains significant information and the shot boundary frames become indispensable. When combined with notes and the jump-back button, it was possible to accurately bookmark locations in 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 could identify the editing nature of the task. 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”.

The results also show that the availability of such advanced features could be immediately beneficial to users. Immediate plans for future work include refining the browser interface based upon observed usability problems. Long term plans include performing in-depth evaluations of the browser features with greater numbers of participants over a much longer period of time. Given the increasing production of video content in different contexts, we feel that there is no shortage of applications for new technology in this area.

ACKNOWLEDGMENT

We thank the Microsoft Usability Labs for use of their facilities and Gayna Williams and JJ Cadiz, who reviewed the paper and gave us valuable suggestions for improvement.

REFERENCES

1. Apple QuickTime Player, 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. Informedia, http://www.informedia.cs.cmu.edu/
9. 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.
10. 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.
11. MediaSite, http://www.mediasite.com/
12. Meng, J. and Chang, S. “CVEPS - a compressed video editing and parsing system.” In *Proceedings ACM Multimedia 96*, 1996, Page 43.
13. Microsoft Windows Media, http://www.microsoft.com/windows/windowsmedia/
14. Mills, M., Cohen, J., and Wong, Y. Y., A magnifier tool for video data, in *Proceedings of CHI '92*, 1992, ACM Press, 93-98.
15. 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.
16. 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.
17. Real Networks RealPlayer, http://www.real.com/
18. Replay Networks ReplayTV, http://www.replaytv.com/
19. Stanford Online, http://stanford-online.stanford.edu/

20. Stifelman, L. "The Audio Notebook: Paper and Pen Interaction with Structured Speech" *Ph.D. dissertation, MIT Media Laboratory*, 1997.

21. TiVo Inc., http://www.tivo.com/

22. Virage, http://www.virage.com