SmartSkip: Consumer level browsing and skipping of digital video content

Steven M. Drucker, Asta Glatzer, Steven De Mar and Curtis Wong
Next Media Research Group
Microsoft Research
1 Microsoft Way
Redmond, WA. 98052

{sdrucker, astag, sdemar, wong}@microsoft.com

ABSTRACT

In this paper, we describe an interface for browsing and skipping digital video content in a consumer setting; that is, sitting and watching television from a couch using a standard remote control. We compare this interface with two other interfaces that are in common use today and found that subjective satisfaction was statistically better with the new interface. Performance metrics however, like time to task completion and number of clicks were worse.

Keywords

Digital video, browsing, skipping, television, user interfaces, user studies, PVR, DVR.

INTRODUCTION

With the advent of DVDs and personal video recorders being used in the home, consumers are starting to have the need to rapidly browse through digitally recorded content while watching television. While there have been many different interfaces devoted to browsing video content for different tasks [1,2,3,5,6,7,8,10,12], few have focused on typical consumer usage and with only a standard remote control (no jog dials etc.).

Existing VCRs have familiarized the average consumer with multiple fast forward speeds and the notion of skipping forwards or backwards in time, but often the mechanical nature of the devices have affected the interface performance. With digital media, there is a great deal more freedom to implement interfaces for controlling the flow of content. It is important to be aware that it is not only consumer preferences that might drive the eventual interfaces that consumers will use. One manufacture of personal video recorders chose not to implement a 30 second skip forward feature because of pressure from the networks since consumers would invariably use it to skip advertisements. Another modified a skip forward time by changing it to skipping forward 25 seconds instead of 30 in order to force the consumer to watch the tail end of a commercial.

Business models aside, we sought to compare existing methods of skipping through content along with a novel interface that we developed to see what consumers preferred and how they performed with three different interfaces.

Specifically, we examined task effectiveness and subjective user preferences for 3 different interfaces on 2 common tasks that consumers often perform on their recorded content: skipping scenes and finding a specific section of a broadcast – in our case, the weather section of a news broadcast.

RELATED WORK

There have been surprisingly few published experiments on effectiveness of remote control operated interfaces for navigating through video content on the television. More work has been done in browsing and skimming digital video on the desktop. Notably, the Informedia project at CMU includes a great deal of work devoted to indexing and searching video including work on video skimming [2] which attempts to compress the playback of a video stream in comprehensible fashion. In a similar vein, the work of Li et al [7] on browsing digital video examined time compression, pause removal, textual indices, and shot boundary frames. Tse et al performed a very similar exploratory study on video browsing interface designs primarily for overall gist determination and for information seeking tasks [12]. Their two primary interfaces were static storyboards and dynamic slideshows and their findings showed that subjects tended to prefer the static story boards which is similar to results that we found in our study.

Work on video surrogates in general is pertinent to browsing and skipping of video. Elliot developed the Video Streamer system [3] which used a stack of still frames with patterns along the edges to assist in user determination of shot boundaries and action. Komlodi and Marchionini looked at key frame preview techniques for video browsing [5]. Srinivasan et al. describe the CueVideo system [10] which uses shot-detected storyboards similar to our SmartSkip system but again, primarily within the desktop context. Arman et al [1] also used shot boundary detection to improve frame selection for storyboards. Lee et al analyze three different dimensions for keyframe based browsing of digital video: a spatial dimension (like a

storyboard), a temporal dimension (a slide show) and a layer dimension (using a keyframe to represent a whole section of video) [6].

Other browsing work includes the Hierarchical Video Magnifier [8] which displayed frames near the current video position to provide context in a similar fashion to the SmartSkip interface.

THE THREE INTERFACES

Two of the interfaces we tested were based on systems already available in over-the-counter personal video recorders: the Replay™ [9] and the TiVo™ [11]. The third interface, SmartSkip was novel and provides visual cues to the scene searching process.

Basic Skip

The first implementation, based on a Replay-like interface, allows the consumer to hit a single button which skips the playback forward in time 30 seconds and another single button which skips the playback back in time 10 seconds. A separate mechanism also exists that used the fast-forward and rewind buttons for moving through the content. There is minimal significant feedback on the screen to indicate what button has been pressed. This differs from the existing implementation of a Replay interface in the following ways: Replay does give some small feedback for what button has been pushed; Replay skips back 8 seconds instead of 10, and Replay allows for multiple speeds of fast-forwarding through the content.

Multiple level Fast Forward

The next implementation, based on a TiVo-like interface, allowed the user to press the fast forward button multiple times for various speed levels and an interface popped up on the screen that indicated the speed level, the distance in the entire clip and the length of the entire clip.

Figure 1: The Fast Forward Interface.

The fast forward speeds were 3x, 20x and 60x, which were chosen because they were the same speeds as is currently used by the TiVo system. In addition, when the user hits play, the system jumps back a certain amount based on the reaction time of the user and the speed at which the user is current fast forwarding.

Smart Skip (or Picture Skip interface)

Finally, a novel interface, which we call SmartSkip (but which was called Picture Skip for the purposes of the study to avoid any user bias based on naming), pops up a sequence of thumbnails of the content when the user presses the fast forward or rewind button on a standard remote control (see figure 3). The user can then move the selected thumbnail forwards or backwards and press the play key to start playing from the beginning of the thumbnail. In addition, the user can select a "zoom level" which changes the time between thumbnails, from approximately 10 seconds, all the way up to 8 minutes. Thus the user can quickly find the first segment after a commercial break at the default resolution of 30 seconds between thumbnails, and then zoom down to 10 seconds to get the very first segment. The yellow "thumb" in the center of the picture below helps show both the zoom resolution (the size of the segment reflects the amount of time covered by all the thumbnails) and what thumb is selected from that region. Note that as the user scrolls to the right edge of the screen, the thumbnails page to the right.

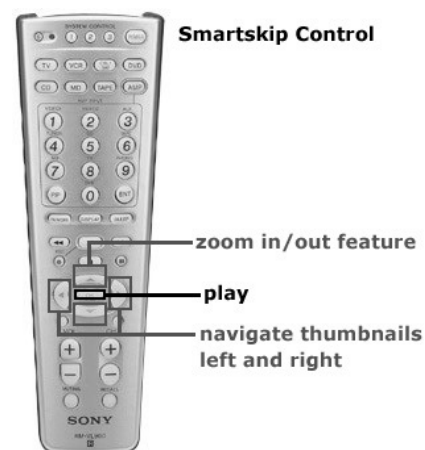**Figure 2: The SmartSkip (or PictureSkip) Interface**

Figure 3: Buttons on a standard Sony Learning Remote Control for controlling SmartSkip (and other interfaces)

Comments on an initial prototype of the SmartSkip system led us to a number of changes: increase in thumbnail size, reduction of the number of thumbnails form 10 to 8 thumbnails per screen, and most significantly, incorporation of shot detection to assist in the selection of thumbnails that represent the beginning of the nearest scene. The initial prototype used a fixed interval between thumbnails, but we suspected that there would be greater utility if the thumbnails were chosen some number of frames after an automatically determined shot boundary. However, basing thumbnail selection only on shot boundaries proved problematic since the spacing between shots can be extremely irregular, and limited testing proved that this was not effective. Users did not like unevenly spaced thumbnails or thumbnails that were spaced evenly but represented significant differences in durations. Instead, the system uses a fixed interval between thumbnails as an initial estimate for the thumbnails, and then modifies the thumbnail to a nearby shot change if there is one within a certain threshold. If there is no shot boundary within the threshold, the default thumbnail is used; otherwise, the shot boundary thumbnail is used. Thus in the following diagram, if a, b, c, d, and e represent shot boundaries, and the interval between thumbnails is 10 seconds, then the thumbnails for 20 seconds, 50 seconds, and 70 seconds will definitely use the thumbnail from the shot boundaries a, c, and e respectively; the thumbnail for 60 seconds may or may not use the thumbnail at shot boundary d based on a system defined threshold; and thumbnails for 10, 30,40 and 80 seconds will all be based purely on absolutely indexing into the content.
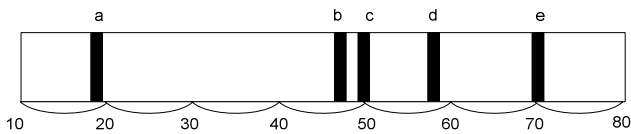

Figure 4: Choice of shot detected thumbnails vs. non shot detected.

Shot detection was done using code from Wei et al [13] and is based on taking a histogram of the image and looking at rapid overall changes in the color and brightness. This operation can also be easily done in the compressed domain. The implementation of the algorithm that we used can run at approximately 120x real time. This algorithm, however, does not interpret subtle histogram changes from fades or dissolves as a scene change. For details on the shot detection method used in SmartSkip, see [13]. We set the thresholds so that we would err on the side of not detecting shot boundaries. Thus in effect, many of the thumbnails that were selected were based on strict time duration. Further experiments need to be done to compare the effect of changing this threshold on user satisfaction and quantitative performance.

Incorporating shot detection in this manner was just one possible way to include shots in the interface. We emphasized keeping the temporal distance between shots approximately uniform since our early testing showed that users preferred this representation as opposed to varying the spatial layout of the thumbnails to represent actual time between shots.

There were several interesting aspects of the SmartSkip interface to note. First, the overall image is paused when the interface is up. This permits the user to move at whatever rate they desire to select the appropriate thumbnail. This may have contributed to the longer average time the user spent finding the desired scene. Second, since the user can observe thumbnails of the commercials, they are not skipped completely which helps address to a certain extent some of the controversial aspects of commercial skipping. Finally, while automatic commercial detection systems have been implemented, there seems to be a constant battle between commercial detectors and advertisers trying to circumvent detection; this method uses human assistance for determination of what is in a shot's content.

We were curious about both subjective perceptions of the 3 different interfaces in addition to quantitative task performance.

RESEARCH METHOD

Each participant received all experimental treatments ((Skip / commercial skip; Fast Forward / commercial skip; SmartSkip / commercial skip; Skip / weather-finding; Fast Forward / weather-finding; SmartSkip / weather-finding) in a repeated-measures design. The order of the 6 interface-task combinations was randomized.

Hypothesis 1: Subjective satisfaction will be different amongst the 3 different interfaces.

Hypothesis 2: User performance (measured by speed and accuracy with which they complete the tasks) will be different amongst the 3 different interfaces.

Experimental Design
Independent Variables:
- Interface Display (3 treatments): Skip; Fast Forward; and SmartSkip
- Tasks: (2 treatments): Commercial skipping and weather finding from a news broadcast.

Dependent Variables:
- Task performance: Accuracy, time to task completion, # of clicks to task completion
- Subjective satisfaction: ranking, ease of use, ease of learning, frustration, and fun.

Participants:
Eleven males and nine females participated in the entire study. They were recruited from the Microsoft's Research Division, but with a variety of positions that were primarily non-technical and administrative support. All participated on a voluntary basis and each received a coffee or desert coupon as compensation. Ages ranged from 21 years old to 53 years with a mean age of 35. Only one of the participants noted that they had previous experience with personal video recorders. 42% of the participants said that they watched between 10 and 15 hours a week of

television, 16% said they watched 5-10 hours per week, 26 % watched less than 5 hours per week, and 16% said they watched no TV during the week.

Software:

The system was developed using Macromedia Director with QuickTime as a plug-in which allowed interface elements to be overlaid on top of the moving video. Shot boundaries were detected automatically for the SmartSkip interface using the algorithm described in the previous section. The only interface required for operating the software was a Sony Learning TV Remote with a fairly standard layout for fast forward, rewind, play and cursor buttons. Unfortunately, the Sony Learning TV Remote did not allow for buttons to automatically send repeated signals which would have allowed users to press and hold a button as opposed to repeatedly pressing it.

Text file transaction logs were automatically recorded for all user actions with each interface to allow for the quantitative analysis of the time and number of clicks it took to perform the tasks. Subjects also completed a web-based survey immediately following the experiment.

Video Materials:

Video clips were digitally encoded using a Sony miniDV camcorder with line-in from a cable TV tuner. They were then transferred via IEEE-1394 interface into a PC and reencoded in QuickTime format. The video clips that were used for the commercial skipping task were rebroadcasts of two different episodes of the situation comedy "Friends" and a rebroadcast of an episode of "The Cosby Show". The video clips used for the weather finding task were 3 different broadcasts of local news on different local stations.

Experiment setting and Hardware:

The participants sat in a couch 7 feet away from a 32 inch television monitor powered by a Windows XP based Pentium III computer running at 800 MHz. The resolution of the display was set to 640x480 and approximated a TV viewing experience in a living room.

Procedures:

1. Participants were briefed on the goals of the study.
2. All 3 interface designs were explained and demonstrated. The participants were allowed as much time as they liked to experiment with each interface design. This part of the experiment lasted 15 minutes at most and usually lasted less than 10 minutes.
3. After the participants expressed familiarity with each of the interfaces the experiment began.
4. For the first part of the experiment, the participant was told that a video clip would be brought up and that as soon as a commercial came on, the participant was to skip the commercial and start the video playing again as close as possible to the segment immediately following the commercial break.
5. Before the video clip came up, the type of interface that was to be used was named, and the experimenter prompted the participant on what the controls were for that interface.
6. The video commenced at 30 seconds prior to the commercial break in all instances.
7. All of the participant's button clicks were logged by the software for later quantitative examination.
8. This process was repeated 2 times for each interface.
9. The second part of the experiment was then described to the participants as finding the beginning of the weather section of a local news broadcast.
10. Before each video clip of the news was brought up, the type of interface was again displayed to the participant and the experimenter reminded the participant as to the controls for the interface. The order followed was the same order as the participant received in the first experimental section.
11. The participants found the weather section once with each type of interface.
12. The participants then filled out a survey with demographic information, television viewing habits and a subjective evaluation of the interfaces. The survey questions included the following information: age; sex; job position, experience with personal video recorder devices; rankings of each interface; how easy each interface was to learn; how easy was each interface to use; how easy was it for each interface to skip commercials; how easy was it for each interface to find the weather; how fun was each interface; and how frustrating was each interface.

RESULTS

Questionnaire Results

After they had completed all tasks, each participant was given a web based questionnaire that asked them to rank their preferences and give their overall reactions to each of the interfaces.
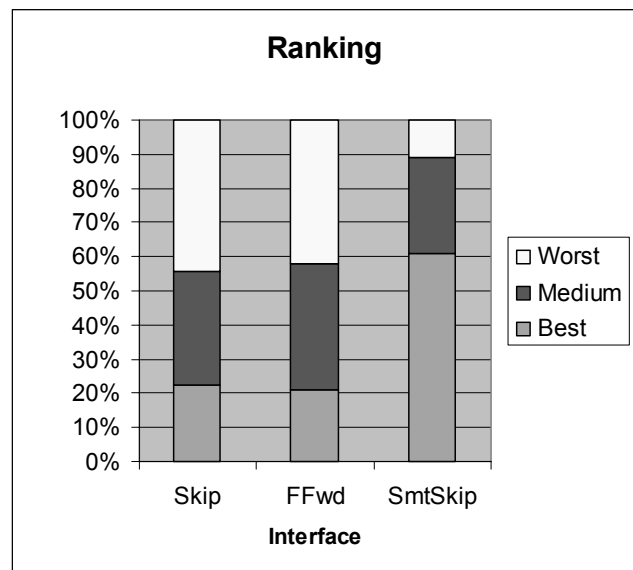
Figure 5: Rankings for each interface: each subject chose a best, medium, and worst interface. The histograms are shown: SmartSkip had the most 'best' rankings and the least 'worst' rankings.

A repeated measures ANOVA on the ranking showed that there was a statistically significant difference between the 3 interfaces, $F(1,18) = 7.23$ ($p<.01$). Pair wise comparisons showed that there was no difference between the Skip and Fast Forward interfaces (the means were: 2.21 and 2.21 respectively), and statistically significant differences between both the Smart Skip and Skip interfaces ($p < .05$) and the Smart Skip and Fast Forward interfaces ($p < .05$).

Table 1 lists the questions used for self-reported reactions to the interfaces and figure 6 shows the means and standard deviations. Note that users reported the most satisfaction with the SmartSkip interface, though only responses to questions 5 (fun) and 6 (frustration) were significantly better for SmartSkip (higher for fun, and lower for frustration) than for the other two interfaces.

#	Questions
1	I found this interface easy to use.
2	I found this interface easy to learn.
3	I could skip commercials easily with this interface.
4	I could find the weather section easily with this interface.
5*	I thought this interface was fun to use.
6*	I thought this interface was frustrating to use. (this is the only question where lower numbers indicate more satisfactory experience).

Table 1: Subject response: * indicates statistically significant results.

The subjective responses were also analyzed through a repeated measures ANOVA. Question 6 (frustration) yielded overall marginally significant results ($F(1,17) = 2.32$; $p = .128$, two-tailed) and pair-wise comparisons yielded a marginally significant difference between both the skip and fast-forward condition (means were 3.2 and 3.9 respectively, $p = .08$) and the smart skip and fast forward conditions (means were 2.7 and 3.9 respectively, $p =.06$. Higher numbers indicate greater levels of frustration.

The only subjective measure that yielded complete statistical significance was whether the interface was 'fun' to use. Here, a repeated measures model yielded ($F(1,17) = 13.99$; $p < .001$). Pair-wise comparisons yielded a statistically significant difference between both the SmartSkip and Skip (means were 6.2 and 4.3 respectively, $p < .001$) and the SmartSkip and Fast Forward conditions (means were 6.2 and 4.2 respectively, $p < .001$). There was no difference between the Fast Forward and the Skip conditions.

Behavioral Results

Recall that we predicted there would be a difference in performance as participants used each of the three interfaces, but we did not know which would lead to better performances. We assessed differences on 3 performance measures: time to complete the tasks, number of clicks to complete tasks and accuracy in performance.

On the first measure, time, we found that it took the least amount of time to skip a commercial segment with the Skip interface, and the most amount of time with the SmartSkip interface. The behavioral measures of performance showed that there was a statistically significant difference in the time it took to skip commercials with the SmartSkip system being the slowest.
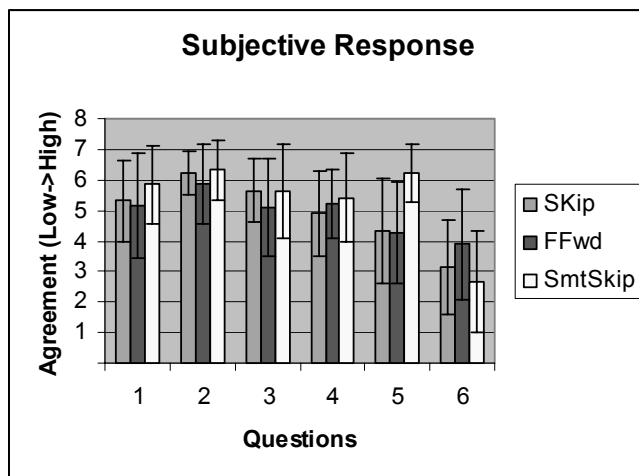

Figure 6: Responses to questions in table 1. Note: higher levels of satisfaction are associated with higher levels in all but question 6 where lower levels indicate greater satisfaction.
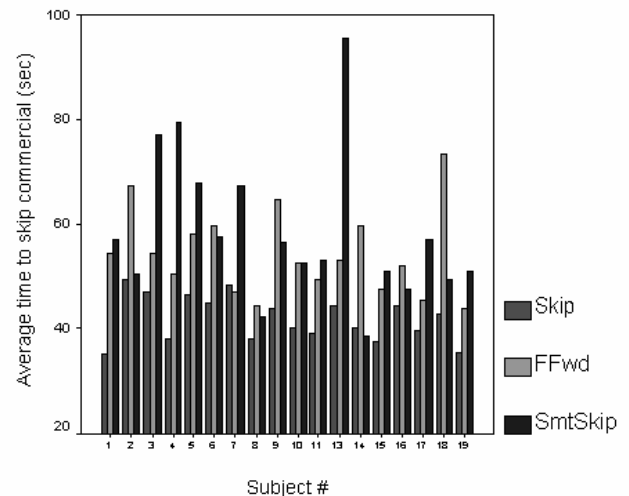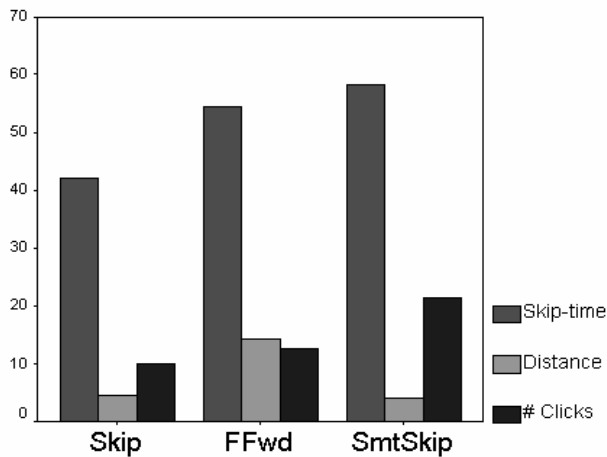

Figure 7: Individual time to completion for the three different interfaces. Note that 4 users got lost while using the SmartSkip interface.

Figure 8: average time to skip commercials, accuracy to final target frame, and number of clicks required for each interface. Time and Distance are both measured in seconds. #Clicks are measured as an absolute number.

Figure 8 shows the average time to skip a commercial break (the breaks ranged from 60 seconds to 4 minutes), the accuracy of the final resulting frame (distance) and the # of clicks that it took. The mean times for commercial skipping for each interface (Skip, Fast Forward and SmartSkip) were 41.9, 54.3, and 58.4 respectively. There was a statistically significant difference ($p < .001$) between Skip and Fast Forward, and between Skip and SmartSkip but not between Fast Forward and SmartSkip.

The means for number of clicks for Skip, Fast Forward and SmartSkip were 9.9, 12.6, and 21.4 respectively. There was a marginal difference between Skip and Fast Forward ($p = .06$) and significant differences between both Skip and SmartSkip and Fast Forward and SmartSkip ($p < .001$). There was no significant difference between the accuracy of the final frame found by the participant.
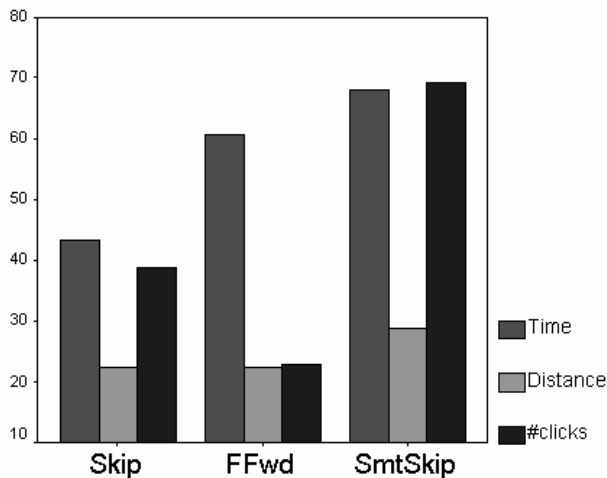
Figure9: For the weather segment finding task, the average time to find the commercial, the distance from the final target frame, and the average number of clicks for each interface. Time and Distance are both measured in seconds. #Clicks are measured as an absolute number.

The means for the total time to find the weather segment for the Skip, Fast Forward, and SmartSkip were 43.3, 60.8, and 68.1 respectively. There was a statistically significant difference between the Skip interface and both the Fast Forward and Smart Skip interfaces ($p < .01$) but no statistically significant difference between the Fast Forward interface and the Smart Skip interfaces.

The means for the number of clicks to find the weather segment for Skip, Fast Forward, and SmartSkip were 38.7, 22.9, and 69.1 respectively. There was a statistically significant difference between each pair of interfaces ($p < .01$ between Skip and Fast Forward, and $p < .001$ for all the other pairs).

There was no significant difference in the accuracy for the participants to find the target frame between each of the interfaces.

In summary, while there was no significant difference in the accuracy of each interface, the SmartSkip interface did take a significantly longer time and significantly more clicks to skip a commercial segment. Likewise, in the weather segment finding task, the SmartSkip interface took significantly longer than the Skip interface, though not significantly longer than the Fast Forward interface. The Smart Skip interface did take significantly more clicks than either of the other 2 interfaces, and the Fast Forward interface took significantly fewer clicks.

DISCUSSION:

It is curious that the subjective opinions of the interface differed so much with the quantitatively calculated results. That is, the SmartSkip interface was rated the most fun, the least frustrating, and as easy to use and learn as all the other interfaces. However it also took the longest to skip commercials and the longest to find the weather section. It also took the most clicks for both of these tasks.

This brings up an important point: it is often not clear what is most important for a consumer level device: subjective satisfaction or quantitative performance. Indeed, what is a good metric for evaluating quantitative performance? We assumed that it would be time to task completion, but perhaps some other metric, involving the amount of attention that the interface demands from the user would have a better correlation with user preference.

From post experiment conversations with the subjects, we found that the disparity between user preference and task performance might be attributed to the amount of attention needed to perform the task with each interface. With the Fast Forward style interface, the user must attend very closely to the picture being displayed on the screen and react exactly when the appropriate section comes up. Likewise, when skipping forward, the user must also attend reasonably closely to the image that comes up immediately after skipping. However with the SmartSkip method, the main screen is frozen and the user can move around the thumbnails at leisure looking for the section that they're interested in. This very lack of time pressure might have

been why they rated it fun and not frustrating while at the same time taking longer to perform the task.

The # of clicks was the least for the fast forward interface since the user primarily places the interface in fast forward mode and then stops it when the video is near the desired target. There is some overshoot and rewinding common for this interface, which marginally correlates with the level of frustration. If the remote control used allowed for automatically repeating a command by holding down the button, we suspect that the number of clicks would have been far less for both the skip and smart skip interfaces than they were in the experiment.

While the users did say that each of the interfaces were equally easy to learn and to use, there was not much time for a user to get used to each interface in this study, so there probably would be some effect if a participant used the interfaces over a significant amount of time. This might include changes in the quantitative results, or the subject rating of 'fun' for the SmartSkip interface might decrease due to any novelty factor that might go away over time. We did find it mildly encouraging that the SmartSkip interface, with its added complexity was rated as easy to learn and use as the other two interfaces.

Informal experiments with experienced users showed that performance on the weather finding segment increased significantly as the user became more comfortable with the zooming aspect of the interface. During the initial experiment however, 4 of the novice users lost their place when zooming; they zoomed up to the level of the entire program, changed the selected thumbnail, and then were not able to find out exactly where they started from. Even with this difficulty, one of those users still preferred the SmartSkip interface over the other interfaces. It is possible that an interface that had pictures in it without multiple zoom levels might have tested better overall for casual users. On the other hand, many users specifically requested more zoom levels to allow finer control of the distance between thumbnails. A redesigned interface includes an indication of the frame that the user initially started on.

The situation for skipping commercials was somewhat artificial in comparison to a real home setting since the subjects had varying degrees of familiarity with the content though the structure of the experiment (comparison within subjects) helps minimize those effects. Still, many subjects commented that their performance would have been different if they were more familiar with the content.

A number of user comments have suggested some other design improvements that we are currently in the process of implementing. This includes increasing the size of both the thumbnails and the selected thumb (by decreasing the number of thumbnails from 8 to 5). A number of users also expressed a desire to be able to zoom into a scene at a greater resolution of 10 seconds between thumbnails. A new prototype has thumbnail to thumbnail intervals of a single second. Some users suggested that it might also be desirable to show motion in the selected thumbnail and we

are currently experimenting with this as a possible interface enhancement. Also, as mentioned previously, an indication of the frame that was up when the user initially presses the SmartSkip button is now given.

Figure 9: Redesigned smart skip with larger thumbnails and better zoom factor indicator.

CONCLUSION

Consumer level control of skipping and browsing of digital video has only been studied marginally to date, primarily for information finding and video editing tasks. As more devices based on digital video come into the home; more rigorous studies on consumer preference and performance need to be done. This study examines three different kinds of skipping and browsing interfaces for a couch based television interaction; 2 based on those currently used in personal video recording devices; and a third, novel interface.

While the performance based on time to task completion and number of clicks was the worst in the last, novel interface, the user satisfaction was significantly better with this interface. This study highlights the necessity for designing interfaces that consumers enjoy using as well as ones that are easy to use and get the job done effectively.

There may be some inverse relation between the amount of attention that the interface requires and the amount of satisfaction that a user has with that interface.

The study also suggests many possible improvements in each of the interfaces and we have already begun to develop new prototypes based on these suggestions.

REFERENCES

1. Arman, F. Depommier, R, Hsu, A. and Chiu, M. "Content-based browsing of video sequences." In Proceedings of the second ACM international conference on Multimedia '94, 1994.

2. Chistel, M. Smith, M. Taylor, C, and Winkler, D, "Evolving video skims into useful multimedia abstractions", In Proceedings of CHI, '98. ACM Press. 1998.

3. Elliot, E. Watch, grab, arrange, see: Thinking with motion images via streams and collages, MSVS thesis, Massachusetts Institute of Technology, Cambridge, MA. 1993.

4. Informedia, http://www.informedia.cs.cmu.edu/

5. Komlodi, A. and G. Marchionini, Key frame preview techniques for video browsing, In Proceedings of the ACM Digital Libraries (DL '98). pp. 118-125. 1998.

6. Lee, H. Smeaton, A., Berrut, C., Murphy, N. Marlow, S, O'Connor, N.E. Implementation and analysis of several keyframe-based browsing interfaces to digital video. Proceedings of Research and Advanced Technology for Digital Libraries. 4th European Conference, ECDL 2000.

7. Li, Frances C., A. Gupta, E. Sanocki, L. He, Y. Rui, Browsing Digital Video, Proceedings of CHI 2000. pp. 169-176. 2000.

8. Mills, M. Cohen, J. and Wong, Y. A magnifier tool for video data, in Proceedings of CHI '92, ACM Press, 93-98. 1992.

9. Replay Networks ReplayTV, http://www.replaytv.com/

10. Srinivasan, S. D. Ponceleon, A. Amir, D. Petkovic, "What is that video anyway?": in search of better browsing. Proceedings of IEEE International Conference on Multimedia.

11. TiVo Inc. http://www.tivo.com/

12. Tse, Toney, S. Vegh, G. Marchionini, B. Shneiderman, An Exploratory Study of Video Browsing User Interface Designs and Research Methodologies: Effectiveness in Information Seeking Tasks. ASIS'99. Proceedings of the 62nd ASIS Annual Meeting. 1999

13. Wei Qi, Lie Gu, Hao Jiang, Xiang-Rong Chen and Hong-Jiang Zhang "Integrating Visual, Audio And Text Analysis For News Video",(Invited Paper), 7th IEEE Intn'l Conference on Image Processing (ICIP 2000), Vancouver, British Columbia, Canada,10-13 September 2000