INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

A User Interface for Negotiating QoS for Video

By

Jeffrey Charles Bos. B.A. (Hon)

A thesis submitted to

The Faculty of Graduate Studies and Research

In partial fulfillment of

The requirements for the degree of

Master of Arts

Department of Psychology

Carleton University

Ottawa, Ontario

September, 2000 ©

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-57698-1

Canada

The undersigned recommend to the
Faculty of Graduate Studies and Research
Acceptance of the thesis:

"A User Interface for Negotiating Quality of Service for Video"

submitted by

Jeffrey Charles Bos

in Partial fulfillment of the requirements for
the degree of Master of Arts

Chair

Thesis Supervisor

Carleton University
September, 2000

Abstract

The purpose of this study was to design and evaluate a user interface for adjusting the quality of digital audio and video transmitted across a provisioned network using Quality of Service guarantees. A high fidelity computer prototype was designed for this study and two rounds of usability testing were conducted with 15 Windows users. Ten users participated in the first round of usability testing. The prototype was redesigned based on results of the first usability tests and a second round of usability testing was conducted with five new users and five users who participated in the first usability test. The prototype allowed users to simulate the effects of increasing or decreasing audio and video quality for video presentations. Participants were asked to perform a number of tasks and to evaluate the usability of the interface. Errors and comments were recorded to detect usability problems and to measure the ease of use of the interface. The usability test for the second design iteration indicated that users could control audio and video quality with slider bars. Users were able to compare two QoS presets and they were able to add personal presets with the interface. Users did not like the appearance of the audio controls and were unable to accept their settings and start a presentation. Recommendations for future designs are discussed.

Acknowledgements

I wish to thank my advisor, Dr. Dillon, for sharing his knowledge of user-centered design and providing thought-provoking suggestions and rapid feedback during the course of carrying out this study. I also wish to thank Dr. Wood for sharing her knowledge of user-centered design and for providing excellent advice during the iterations of the user interface design and during the usability testing. I wish to thank Dr. West for providing useful suggestions for improving the usability tests. I wish to thank Ronald Boring for sharing countless hours of his time helping to code the user interface prototypes and for providing assistance during the usability testing. I wish to thank Steven Zikopolous for sharing his knowledge of digital video and data networks. I would also like to thank the students of the CURE lab for providing suggestions and constructive feedback during the design of the user interface. I would like to thank Dr. Stavely and Dr. Matheson for participating on my thesis defense committee. Last, I would like to thank my parents for providing encouragement and support.

Table of Contents

Acceptance Form.....	ii
Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
List of Tables.....	vi
List of Figures.....	vi
Introduction.....	1
Quality of Service Research.....	2
Adjusting Audio and Video Quality.....	6
Judging Audio and Video Quality.....	7
Iteration One.....	8
Method.....	8
Video Content Samples.....	8
Audio/Video Presets.....	8
User Interface Design.....	8
Design Iterations.....	10
Prototype.....	16
Participants.....	17
Test Setting.....	17
Usability Test.....	17
Procedure.....	17
Usability Goals.....	18
Results and Discussion.....	18
Iteration Two.....	25
Method.....	25
Usability Test Revisions.....	25
Video Content Changes.....	26
Prototype Revisions.....	26
Test Protocol Changes.....	28
Rating Scales.....	29
Usability Goals.....	29
Participants.....	29
Results and Discussion.....	30
General Discussion.....	36
References.....	41
Appendices	
A - Scenarios Iteration One.....	43
B - Informed Consent Form.....	46
C - Iteration One Usability Test Introduction.....	47
D - Rating Scales for Iteration One Usability Test.....	48
E - Scenarios Iteration Two.....	49
F - Iteration Two Usability Test Introduction.....	52
G - Rating Scales for Iteration Two Usability Test.....	53

List of Tables

Table 1.	Average Data Rates for Video.....	2
Table 2.	Video Classification Scheme.....	7
Table 3.	Results for Iteration One by Issue and Usability Goal.....	19
Table 4.	Changes to Preset Labels.....	28
Table 5.	Results for Iteration Two by Issue and Usability Goal.....	30

List of Figures

Figure 1.	QoS User Interface for Presentational Applications.....	4
Figure 2.	Mapping User QoS from a Single Parameter.....	5
Figure 3.	The user interface tested in the first round of usability testing.....	9
Figure 4.	The first design iteration of the user interface.....	10
Figure 5.	The video display window used to display two videos simultaneously...	11
Figure 6.	The second design iteration of the user interface.....	12
Figure 7.	The third design iteration of the user interface.....	13
Figure 8.	The fourth design iteration of the user interface.....	14
Figure 9.	Sixth design iteration of the user interface.....	15
Figure 10.	The user interface tested in the second round of usability testing.....	28
Figure 11.	Video Detail and Movement sliders for controlling videos.....	31
Figure 12.	Audio indicator and Audio button.....	33

The purpose of the present study was to design and evaluate a user interface for adjusting the quality of digital audio and video transmitted across a provisioned network using Quality of Service guarantees. Digital audio and video have been broadcast over the Internet for many purposes including video conferencing, education and entertainment. A technology called Quality of Service (QoS) has been developed to guarantee the transmission quality of packet-loss sensitive data such as digital audio and video. Quality of Service technology has three advantages for the end user: guaranteed service cost, guaranteed presentation quality, and guaranteed end-system capability.

First, users may choose to reduce service cost by reducing audio and video quality to the quality they need. The service cost might be determined by the average data rate required for a continuous media presentation for a given period of time. Several audio and video quality parameters affect the average data rate including the size of the video, the number of video frames per second, and the amount of audio and video compression. "Some users may be ready to pay a higher price for obtaining a maximum quality, while others may prefer low-cost presentations with lower quality" (Bochmann, Fischer, and Salem 1997, p.1).

Second, users may adjust the quality for video presentations to suit the type of video content they choose to watch. Research on video acceptability and frame rate (Apteker, Fisher, Valentin, Kisimov, and Neishlos, 1995) has indicated that users may require different audio and video settings to achieve a minimum acceptable level of quality for viewing different types of video content. For example, different video content types with differing levels of sound complexity, image detail and motion may require different audio and video quality settings (Apteker et al., 1995). Accurately representing slow motion video content requires higher video frame rates (number of video frames displayed per second) while fast motion video content requires lower video frame rates. For example, a user might require the highest frame rate to view slow, continuous motion such as a Tai Chi exercise, but might settle for a slightly reduced frame rate for viewing fast, abrupt motion such as a car chase. A user might require CD quality audio when listening to a music video, but might settle for telephone quality audio when listening to a sports event (Apteker et al., 1995).

Third, users might wish to increase or decrease video quality depending on the video processing capacity of the computer they use for displaying video. While most Pentium III class computers and better

are capable of displaying video at maximum quality. some older computers might not be capable of displaying video at high quality. For example, reduced video detail and reduced frame rate settings might be necessary for viewing video on a portable computer equipped with a slow microprocessor or a low resolution video card. Due to the huge amount of data associated with continuous media, especially high quality audio and video, underlying systems are required to support a high data rate to support multimedia applications (Bochmann, Dssouli, and Hafid, 1998). Table 1 summarizes the average rates for various degrees of audio and video quality. In the case of videoconferencing or distance education, a range of quality settings might be required depending on the capabilities of the end systems for different students (Bochmann, et al. 1998, p.1).

Table 1.

Average Data Rates for Video

Service	Average Bit Rate
Uncompressed CD-Quality Audio	100-200 Kilobits/s
Compressed Video (MPEG-1)	~2 Megabits/s
Uncompressed Standard-Quality Video	140 Megabits/s
Compressed HDTV	128 Megabits/s
Uncompressed HDTV	1.2 Gigabits/s

Note. From "Distributed Multimedia Applications and Quality of Service: A Review," by R. Dssouli, A. Hafid, and G. Von Bochmann, 1998, Electronic Journal on Network and Distributed Processing, 6, p. 5.

Quality of Service Research

Bochmann, Dssouli, and Hafid (1998) proposed the following definition of Quality of Service. "Quality of Service represents the set of those quantitative and qualitative characteristics of a distributed Multimedia system necessary to achieve the required functionality of an application" (p. 1).

Huard and Lazar (1997) have shown that delivering QoS guarantees in distributed systems is an end-to-end issue, i.e., from application-to-application rather than from one network element to another network element. "It is the end-to-end QoS, rather than the per-hop QoS, that is perceived and important to the users" (Huard and Lazar, 1997, p. 1). According to Bochmann, Fischer, and Mohamed-Vall (1997),

Much work on QoS has been done in the context of high-speed networks in order to provide for some guarantee of quality for the provided communication service, which is characterized by the bandwidth of the media stream and the delay, jitter and loss rate provided by the network. More recently, QoS has been considered in a more global context, including also the end systems, such as the user's workstations and database servers (p.1).

From the perspective of users, quality is concerned with achieving the appropriate audio and video quality for the least amount of money. In the case of video conferencing, the video content providers are also the users. Users are responsible for ensuring that the content quality is acceptable. Even when users may not need to directly control QoS, they must be able to specify the level of service they desire.

While a few studies have attempted to design network systems that give users control over QoS, to the author's knowledge, only three studies have proposed a design for a user interface (Bochmann, Fischer, and Salem, 1997; Michelitsch, Reininger, and Welling, 1997; and Antoniadis, Richards, Rogers, and Witana, 1998). These studies were concerned with addressing technical issues for the implementation of QoS technology rather than usability issues. User interfaces designed in these studies were either excessively complex or inadequate. None of these studies followed a user-centered approach for designing the user interface.

According to researchers at the University of Montreal, "the user should define (through a suitable user interface for QoS negotiation) the criteria which are used by the system to select the 'best' system configuration for the application at hand" (Bochmann et al., 1997, p. 2). Figure 1 illustrates the user interface designed by Bochmann, 1997.

Video Profile Editor

Current Profile: **Profile 1** Help

Save as Profile: **Profile 1**

Frame Rate

Frozen **15** TV HDTV

Desired: [Slider from 5 to 15]

Minimum: [Slider from 5 to 15]

Importance: [Three small preview windows]

Colour

B & W Gray scale Full color **3**

Desired: [Slider from 2 to 3]

Minimum: [Slider from 2 to 3]

Importance: [Three small preview windows]

Resolution

Minimum **425** TV HDTV

Desired: [Slider from 300 to 425]

Minimum: [Slider from 300 to 425]

Importance: [Three small preview windows]

Save Save as... Delete Show example Cancel

Figure 1. QoS User Interface for Presentational Applications (Bochmann, Fischer, and Salem, 1997)

The user interface permits users to specify his/her requirements in advance by creating a video quality profile from all the available settings for the multimedia components (frame rate, number of colours, and resolution). In their approach, a single video window is provided for previewing the video settings. Users could change service requirements during a presentation by selecting a different video quality profile.

A QoS project called SuperNOVA (Negotiated Online Video Access) has been conducted by Antoniadis, Richards, Rogers, and Witana (1998) at the Distributed Systems Technology Center in Australia. Figure 2 illustrates the user interfaces designed for the SuperNOVA project.

Figure 2. Mapping User QoS from a Single Parameter (Antoniades, Richards, Rogers, and Witana, 1998)

In the SuperNOVA design, cost was the only parameter used to adjust QoS. Users interacted with the system by adjusting the cost of service on the fly. Users could optionally define a range of preferred values (acceptable and desired) and assign weightings of importance to the video resolution and frame rate parameters (Antoniades et al., 1998). The system attempted to use this information to choose the most appropriate settings to satisfy users' needs (Antoniades et al., 1998). However, Apteker et al. (1998) say that the cost parameter is too narrow for negotiating QoS because different video content types require different audio and video quality settings. According to Bochmann et al. (1998, p. 19), users should have access to profiles for automatically selecting audio quality, video quality, video resolution, and video frame rate to avoid repeating the lengthy process of setting the QoS parameters for each new session.

Adjusting Audio and Video Quality

Adjusting QoS for a video presentation is a complex cognitive task. Part of the reason for this complexity is the way humans perceive video with different content types. Apteker et al. (1995) defined video "watchability" as a composite term that embraces various aspects of user acceptance of a video message, such as clarity and acceptability of audio signals, continuity of visual messages, lip synchronization during speech, and the general relationship between visual and auditory message components. A video classification scheme was created in Apteker's study which identified eight discrete video categories using three dimensions inherent in video messages (a) the temporal nature of the data, (b) the importance of auditory components, and (c) the visual components to understanding the message. According to Apteker et al. (1995),

Temporality describes the dynamic nature of the motion picture. For example, sports footage would be characterized as "high temporal data because of rapid scene changes, while talk show would be characterized as "low temporal data" due to the static nature of the talking heads. Auditory and visual message content describe the importance of sound and image for different video information, based on elementary principles of psychology. Different types of video information imply different user watchability requirements (p. 32).

The classification scheme was developed by showing 20 respondents 80 short video clips, each taken from a wide variety of video types. The respondents were asked to rate the importance of audio, video, and movement for each clip on seven-point Likert scales.

We assessed the reliability among raters using Spearman's rank-order correlation, which yielded reliability coefficients (r_s) ranging from 0.65 to 0.91. These correlations indicated a good to strong reliability among raters, since agreement increases in strength as the coefficient approaches 1. We used the most reliable of these ratings to establish the most discrepant clips (that is the lowest and the highest) on each dimension. We labeled these clips "lo" and "hi" and assembled them in terms of each permutation – A_{lo} or A_{hi} , V_{lo} or V_{hi} , and T_{lo} or T_{hi} – to produce the eight categories (p. 34).

Table 2 provides examples of the eight video content types used in the study by Apteker et al. (1995). Users perceived a reduced frame rate differently for a continuous-media stream depending on the content type.

Table 2.

Video Classification Scheme

	Temporality Low	Temporality High
Audio Low, Video Low	Logo/Test Pattern	Station Break
Audio Low, Video high	Snooker	Sporting Highlights
Audio high, Video Low	Talk Show	Advertisements
Audio high, Video high	Stand-up Comedy	Music Video

Note. From "Video Acceptability and Frame Rate," by Apteker, R. T., Fisher, J. A., Kisimov, V. S., & Neishlos, H. (1995). *IEEE Multimedia*, Vol. 2, No. 3, p. 32 – 40.

Judging Audio and Video Quality

According to Bailer, Berts, and Persson (1998, p. 28), "The most basic method of subjective assessment of digital video is playing two video sequences next to each other to determine which one has the best quality." A study has shown that viewers are significantly more confident making comparisons viewing two digital images simultaneously rather than sequentially in terms of decision confidence (Carboni, 1989, p. 94). "Eighty-three percent of subjects said they would prefer to use simultaneous mode in a job where accuracy was crucial" (Carboni, 1989, p. 95).

Iteration One

Method

Video Content Samples

Video samples of five video content types of the eight identified by Apteker et al. (1995) were included in this study: tai chi, talk show, lecture, sporting highlights, and action movie. Video samples of three other video content types defined by Apteker et al. (logo/test pattern, station break, and advertisements) were not included because it was assumed consumers were not interested in paying to watch these content types.

Audio/Video Presets

The user interface included Quality of Service presets for five of the eight video content-types (talk show, seminar, arts, crafts and games, stadium sports event, and music video) identified by Apteker et al. (1995) because they were appropriate for watching the majority of video content types where image quality is the most important factor. Presets were provided because there were 64 possible combinations of audio and video settings, but acceptable quality might be achieved with only a few settings. Presets were designed to help users quickly configure the quality settings because each user's perception of video quality depended on a number of subjective factors. The preset audio quality, video detail, and movement levels were selected based on the video content categories identified by Apteker et al. (1995).

User Interface Design

The first part of the study was to design a user interface for controlling QoS. During the design of the user interface several design principles were followed to improve the usability of the interface. Some design principles included reducing interface complexity, reducing hierarchy, reducing visual noise, providing user feedback, providing visual cues, providing context sensitive information, and providing consistency.

Figure 3 illustrates the user interface tested during the first round of usability testing. To reduce complexity, the QoS user interface was displayed in a single window to eliminate the need for navigation between windows. During video previews, the current (right) and proposed (left) video windows were presented simultaneously so users could adjust the QoS controls while they compared two video settings.

In the first iteration, the location of the two videos (current and proposed) was chosen based on the order in which users view the interface (from left to right).

Figure 3. The user interface tested in the first round of usability testing.

To provide feedback, light grey indicators appeared on and below slider bars and on the right side of the radio buttons to highlight the current settings and the current preset. The positions of the grey indicators were immediately updated after users pressed the Accept button to apply changes. If such changes to the QoS were made during the preview, the quality of the video would change immediately. Captions above the videos indicated the name of the selected preset. The proposed setting changed as users selected a preset content type or moved the video or audio quality sliders. The current setting changed

when users clicked the Accept button. The name of the current setting changed immediately after changes were applied to the current setting.

The videos looped continuously until users clicked the Stop button or the OK button. When the Stop button was clicked, both videos froze on the first frame allowing users to judge the video detail based on the quality of a frozen still image. If no changes were made, a caption, "Move sliders to adjust" was displayed in the left window. To reduce visual noise the interface had as few buttons as possible. Graphics were only used to indicate the audio source and to indicate the current positions of the quality settings. Icons and graphics were kept as small and simple as possible. Black text is used on either a grey or white background to make the text easier to read.

In this study the number of frames per second, audio quality, and image quality were investigated. Network delay, jitter, and loss rate were not manipulated in this study because it was not possible to simulate them. Video resolution and video display size were not manipulated in this study because we proposed to address these parameters in a separate user interface to be designed in a future study. This was the first study on QoS conducted in the CURE (Computer User Research and Evaluation) lab. We plan to address hardware and network constraints in future studies.

Design Iterations

The user interface used in iteration one and shown in Figure 3 was designed using an iterative process. Figure 4 illustrates the first design iteration presented to Dr. R. F. Dillon and J. Dr. Wood.

Figure 4. The first design iteration of the user interface.

Two videos were displayed in a separate window in the second design iteration (see Figure 5).

Figure 5. The video display window used to display two videos simultaneously.

The first design iteration was discussed and an expert evaluation was conducted to identify potential usability problems. The user interface was iterated and redesigned (see Figure 6). Pull-down menus for adjusting quality were replaced by slider bars, and radio buttons were added for selecting Quality of Service presets.

Figure 6. The second design iteration of the user interface.

Dr. Dillon and Dr. Wood conducted a second expert evaluation. The design was iterated a third time as shown in Figure 7. Graphics were added above the slider bars to indicate the current settings and many labels on the interface were changed. Two videos were displayed in a separate window in the third iteration of the design as shown in Figure 3.

Figure 7. The third design iteration of the user interface.

Changes were presented to Dr. Dillon, Dr. Wood, and six graduate students in the CURE laboratory. Potential usability problems were identified and recommendations were made to improve the usability of the design. Graphics were added to indicate the level for each of the quality settings, and up and down arrow buttons were added to adjust each QoS parameter (see Figure 8). The cost parameter was indicated graphically using a slider bar and pictures of gold coins. The bottom half of the interface (used for adjusting quality) could be rolled up and hidden to reduce complexity. Two videos were displayed in a separate window as in earlier versions of the design (see Figure 5).

Figure 8. The fourth design iteration of the user interface.

The prototype was then iterated a fifth time and the video window size parameter was eliminated. The plus and minus buttons replaced the arrow buttons. The caption for the Compare button was changed to Preview.

The prototype was iterated a sixth time (see Figure 9). Apple Quicktime 4.0 streaming video technology was used to replace RealVideo G2 streaming video technology. The Apple Quicktime Pro video player ActiveX control was embedded within the Visual Basic application. This improvement made it technically possible to display both videos in the main window and provided higher quality video synchronization for playing two videos simultaneously. Slider bars with four steps were implemented to replace the buttons for adjusting quality. Buttons with audio icons were located on each of the video windows to select the active audio source. Highlighted grey slider bar graphics were created to represent

the positions of the current quality settings. The captions above the videos were added and the videos were renamed Current Setting and Proposed Setting. The elements of the user interface referring to service cost were removed because cost decisions were beyond the scope of this study. We proposed to address the cost parameter in a future study. The setting for adjusting video size was removed from the interface to simplify the QoS negotiation process. We proposed to investigate video size in a future study. We proposed that the size of the video window should be flexible and that users should be permitted to adjust size during video presentations by dragging the video window to change its size at standard sizes. A change in window size is a request for a change in QoS.

Figure 9. Sixth design iteration of the user interface.

Prototype

The study was conducted using a high fidelity QoS computer prototype programmed in Visual Basic 6.0 (see Figures 3). The prototype was run on a PC running Microsoft Windows NT with a Pentium III 450 MHz processor, 256 megabytes RAM (random access memory), 8 megabytes of video RAM and a 19" (48.26 cm) Sony monitor with the display resolution set to 1024 by 768 pixels.

Five different high quality video clips were recorded from DVD (Digital Video Disc or Digital Versatile Disc) sources and digitized to AVI (Audio Video Interleaved) format using loss less hardware compression with CD quality audio. The clips were captured at a resolution of 320 by 240 pixels using Microsoft VidCap32 with a Winnov Videum video capture card. Five video clips were chosen to represent the five different types of video content used by Apteker et al. (1995). The video content types included figure skating, a car chase, an interview, a tai chi lesson, and a lecture. The video source clips were digitally degraded using Adobe Premiere to produce a range of 16 video quality settings. Each video clip was compressed at four video detail settings using Sorenson video compression CODEC. Video was compressed at 1%, 25%, 50%, and 100% of maximum video quality. With Sorenson Video, 0% was the lowest possible quality compressed video and 100% was the highest possible quality compressed video. Video compression reduces the average bandwidth allocated to video without affecting the frame rate. Each detail setting was encoded at four frame rate settings (1 frame per second, 3 frames per second, 9 frames per second, and 29.95 frames per second). These detail settings were chosen based on the judgment of the experimenter.

Audio was recorded on a PC from a DVD source and digitally degraded at four quality settings using the demo version of QDesign2 Pro streaming audio CODEC for Apple Quicktime 4.0. Audio was presented at four different quality settings (8kb/s, 32kb/s, 64kb/s, and 128kb/s). One audio track and two video windows were presented using the Skylight Quicktime ActiveX control for Visual Basic. Using the prototype, users were able to select and compare any of 16 different digitally degraded versions of each video clip combined with any one of the four available audio quality settings.

Participants

10 Windows PC Users were recruited to participate in the first usability test. All participants were Microsoft Windows 95/98/NT users. Experience viewing Internet streaming video applications was not a requirement to participate. Users were paid to participate.

Test Setting

Testing was conducted in a testing room in the CURE lab. Two experimenters were present at all times during testing. Videotape was used to record the subjects' comments and to record video of the users' actions on the computer screen.

Usability Test

A usability test, shown in Appendix A, was designed to evaluate the first iteration user interface. Based on feedback from expert evaluators and from experience using the prototype, several potential usability problems were identified. Scenarios and tasks were created to evaluate potential usability problems identified during the expert evaluations. The scenarios were designed to investigate the following usability issues (a) are users able to adjust audio and video quality, (b) do they use the two videos to compare settings, (c) can they use the video content type presets, and (d) do the labels and examples for the video content type presets make sense to users.

Procedure

Prior to testing, all users were required to complete an informed consent form shown in Appendix B. In the introduction to the usability test, shown in Appendix C, each user was assured that the purpose of this study was not to test his or her ability. Rather, the purpose of the study was to evaluate the user interface. There were no right or wrong answers to the questions. Users were instructed to speak out loud during the test to explain what they were thinking while performing each task. Users were told that higher quality costs more money. Users were instructed to select the minimum acceptable quality to try and reduce the cost. Each usability test took approximately one hour to complete.

Each user was asked to perform nine tasks shown in Appendix A. One experimenter read instructions to users and the second experimenter recorded all user actions, comments, errors, and specific areas of confusion. At the beginning of each task, users were provided any contextual information necessary to complete the task. First the experimenter asked users to verbally explain the required steps for

completing the task. Then the experimenter instructed users to perform the task using the prototype. At the end of the first task, the first experimenter asked users to assign a rating on a Likert scale (shown in Appendix D) for ease of use of the interface. The first experimenter also asked users to assign a rating on a Likert scale for satisfaction for the user interface in terms of usability. At the end of the second task, the first experimenter asked the user to assign a rating on a Likert scale to rate the ease of judging video quality using the user interface. Before the third task, users were informed about how the application worked and areas of uncertainty were discussed. All users were asked to perform each scenario. The presentation order of the video clips was counterbalanced using a Latin square.

Usability Goals

The ease of use of the interface was measured according to the number of users able to successfully complete each task without making an error or requiring a hint. Each task was evaluated according to the number of user errors. For the purpose of this study, errors were defined as inappropriate or unintentional actions that delayed or prevented completion of a specific task. The user interface issue evaluated in a scenario passed if more than 7 of 10 users were able to complete the scenario without making an error and without requiring a hint. For the design to pass, seven of 10 users had to successfully complete all scenarios. Ratings for ease of use and user satisfaction were recorded for the first scenario. The first scenario user interface issue passed if 7 of 10 users rated the ease of use three or less on a seven-point scale (one being very easy and seven being very difficult). Additionally, 7 of 10 users had to rate user satisfaction three or less (one being very satisfied and seven being very dissatisfied) for the first scenario. The user interface passed if 7 of 10 users were able to successfully complete all scenarios.

Results and Discussion

The usability test for the first iteration was designed to evaluate the user interface in terms of thirteen specific issues: (1) Do users need one or two videos, (2) Can users use slider bars to adjust video quality, (3) Can users adjust audio quality, (4) Can users determine which audio track is playing, (5) Can users switch between audio tracks, (6) Can users use the presets to adjust audio and video quality, (7) Are the preset names and exemplars correct, (8) Can users compare two presets, (9) Do users need personal presets, (10) Do users need both slider bars and presets, (11) Can users preview their settings, (12) Can users accept their proposed settings and start a presentation, and (13) Do users understand the purpose of

the ghost sliders. Results for iteration one are summarized in Table 3. Design decisions implemented in the second iteration are presented for each issue.

Table 3.

Results for Iteration One by Issue and Usability Goal

Issue	Number of Users with Initial Understanding Before using Interface	Number of Users able to use the Interface	Design Passed?
1. Two videos	3	6	Failed
2. Use video sliders	4	10	Failed
3. Use audio sliders	9	10	Passed
4. Which side has audio	0	Not Evaluated	Failed
5. Switch audio	0	6	Failed
6. Use presets	1	3	Failed
7. Preset names OK	5	Not Evaluated	Failed
8. Compare presets	Not Evaluated	2	Failed
9. Need personal presets	7	9	Passed
10. Need sliders and presets	Not Evaluated	6	Failed
11. Preview settings	9	3	Failed
12. Accept settings	3	4	Failed
13. Ghost sliders	1	10	Failed

For the purpose of this analysis, only the user performance data was examined. Design decisions were based on user errors observed during the usability test. User comments and rating scale data was not considered because this data was often inconsistent with the user behaviors and reactions observed during the usability test. For example, users rated the interface high for ease of use and satisfaction despite being unable to use the interface to carry out the tasks.

1. Do users need one or two videos?

QoS was adjusted by comparing the videos in the two windows. One of the primary issues evaluated in the usability test for the first iteration was whether or not the purpose of the two videos was clear to users and whether they used both videos. Two videos were initially confusing to users. Before

performing any tasks. 7 of 10 users did not understand that the left video was for making QoS adjustments and the right video displayed the most recent QoS adjustment the user had negotiated. The captions (Current Settings and Proposed Settings) were not adequate to communicate to users the purpose of having two videos. Even after the distinction between the videos was explained, four users looked at only one video. One user repeatedly demonstrated difficulties distinguishing between the two videos. However, once users completed all tasks and understood how to use the two videos for comparing quality, eight users indicated that use of two videos was effective during the preview.

Decision: Use two videos in the second iteration of the prototype but change design to clarify the purpose.

Rationale: The purpose of the videos was initially unclear to users, but users indicated that the two videos were effective after they understood how to use them.

2. Can users use slider bars to adjust video quality?

Another important issue was whether or not users were able to use the slider bars to adjust video quality to the lowest acceptable level. Video quality was adjusted by moving the Video Detail and Movement slider bars shown in Figure 3. Before performing any tasks, users were asked to look at the user interface and answer specific questions to evaluate if they understood how to adjust video quality. Six users did not initially understand the purpose of the Video Detail and Movement slider bars. With practice, users were able to use the slider bars to adjust the quality of the proposed video. Eight users said that adjusting video quality was not complex.

Decision: Include Video Detail and Movement slider bars in the second iteration of the prototype but change the design to clarify the purpose.

Rationale: The purpose of the video sliders was initially unclear to users, but users indicated that the video sliders were not complex to use after they understood how to use them.

3. Can users adjust audio quality?

Another important issue was whether or not users were able to adjust audio quality. Audio quality was adjusted by moving the Audio Quality slider bar. Before performing any tasks, users were asked to explain the purpose of the Audio Quality slider. Nine of 10 users were able to correctly explain the purpose of the Audio Quality slider. During the usability test, all users were able to use the Audio Quality slider to adjust audio quality.

Decision: Add audio slider bars under each of the videos in the second iteration of the prototype.

4. Can users determine which audio track is playing?

Another important issue was whether or not users were able to determine which of the two audio tracks was playing. When the audio track for the left video was playing, a red X was displayed over the right audio button. A red X was displayed over the left audio button when the audio track for the right video was playing. Before using the prototype, users were asked to explain what the icons on the speaker buttons represented. No users initially understood that the icons indicated which of the audio tracks was playing. Three users initially said the icons indicated the sound was on and three users said the icons controlled volume. During the second task, one user said, "It is difficult to determine which of the two videos the audio slider refers to." After using the interface, one user said the audio buttons should be positioned below the pictures. Another user said captions should be added below the videos to indicate which of the two audio tracks was playing.

Decision: Add arrows and an Audio On/Off caption below the audio buttons in the second iteration of the prototype to indicate which audio track is playing.

Rationale: The speaker icons looked like volume controls to many users. Arrow shaped buttons may more clearly indicate which video has sound.

5. Can users switch between audio tracks?

Another important issue was whether or not users were able to use the speaker buttons to switch the audio between the two audio tracks. The audio track for the video on the left was switched on by clicking the speaker button on the left video. The audio track for the video on the right was switched on by clicking the speaker button on the right video. No users initially understood that the speaker buttons were used to switch between the two audio tracks. During the test, users had to use the speaker buttons when they were asked to compare the audio quality for two presets. Only six users were able to use the speaker buttons to listen to and compare the audio tracks. Four users clicked each of the presets and noted the changes in the slider bar positions rather than using the speaker buttons to switch between audio tracks. After the usability test, seven users were able to explain how to use the speaker buttons to switch audio tracks.

Decision: Find a better way to switch audio between the video windows.

Rationale: No users initially understood the purpose of the speaker buttons. Arrow shaped buttons may clarify the relationship between audio controls and the two videos.

6. Can users use the presets to adjust audio and video quality?

Another important issue was whether or not users were able to use the Content Type Presets to adjust video quality to the lowest acceptable level. Clicking the radio button beside each preset changed the audio quality, video detail, and movement slider bar settings to a preset setting. Before performing any tasks, users were asked to explain the purpose of the Content Type Presets. Only 1 of 10 users initially understood the purpose of the content type presets. Three of 10 users were able to use the presets to adjust QoS during the first task. The same three users used the presets again during the second task to further adjust QoS. After the purpose of the presets was explained, users were instructed to use the presets to get the desired settings for each of the video content types. All users were able to use the presets to adjust audio and video quality.

Decision: Rename the presets Audio and Video Quality Presets in the second iteration of the prototype.

Rationale: The name Content Type Presets may have implied to users that the presets were used to change the media content rather than adjust audio and video quality.

7. Are the preset names and exemplars correct?

Another important issue was whether or not the preset names and exemplars were correct. Before using the user interface, five users incorrectly expected that the content type presets would change the video content rather than change the quality of the audio and video. After using the prototype and receiving an explanation of the purpose of the presets, users were asked to try each preset and evaluate if the names were correct. One user commented that News Anchor Desk (from the talk show category) and Weather Report (from the seminar category) represented the same type of content. Individual users suggested including the following additional categories: Non-stadium sports, Cartoons, Movies, and Opera.

Decision: Rename the presets using names that describe the settings rather than the video content.

Rationale: The exemplar names may have implied to users that the presets were used to change the media content rather than to adjust the audio and video quality.

8. Can users compare two presets?

Another important issue was whether or not users were able to compare two presets. Comparing

two presets was done by clicking on a preset, clicking on the Accept button, and clicking on another preset. During the usability test, users were instructed to compare two presets side by side. Eight users were unable to compare two presets. Five users required a hint and three users were not able to compare the two presets. The two successful users said comparing two presets was a trick they would never think of trying if they were using the interface on their own. After an unsuccessful user was shown how to compare two presets he said, "I would never have been able to figure this out on my own." After using the interface, two users commented that comparing two presets was not intuitive.

Decision: Add separate presets to control left and right videos in the second iteration of the prototype.

Rationale: Comparing presets was complex and not intuitive to users. Adding two sets of presets simplified the process of comparing presets.

9. Do users need personal presets?

The usability test evaluated whether or not users needed personal presets and if they understood how to add a personal preset. Clicking the pull-down list labeled "Add a Preset" added a new entry to the personal preset pull-down list. A unique preset name could be typed so the settings could be recognized and used again at a later time. Before performing any tasks, seven users were able to correctly explain how to add a preset. After performing all tasks, nine users said personal presets are desirable. Two users said they would also prefer to be able to change the names of existing presets or delete presets.

Decision: Include a text field and an Add button to add personal presets in the second iteration of the prototype.

Rationale: The Personal Preset pull-down list was a non-standard use of a standard Windows user interface component.

10. Do users need both slider bars and presets?

Another important issue was whether or not users need both slider bars and presets. After completing all the tasks in the usability test, seven users said they preferred to use the slider bars to the presets. However, six users said that the slider bars alone (without the presets) would not be acceptable. After using the presets one user commented, "It is much easier to use the presets, as opposed to adjusting the slider bars, because there are fewer choices to try out." Another user said, "I wouldn't know what I was doing without the presets."

Decision: Use slider bars and presets in the second iteration of the prototype but change the design to clarify the purpose.

11. Can users preview their settings?

Another important issue was whether or not users were able to preview their settings when they made changes. Clicking the Preview button started live videos and audio. When the live videos played, the label on the Preview button changed to Stop. Clicking the Stop button stopped the live videos and audio. Before performing any tasks, users were asked to explain how to preview the videos. All users initially expected it was possible to play the videos. Nine users correctly expected that the purpose of the Preview button was to play the videos. However, four users incorrectly expected that the purpose of the OK button was to play the videos.

During the first task, seven users failed to preview the videos when they changed the Audio Quality and Movement settings. Three users clicked the OK button rather than the Preview button to preview the videos. Four users accepted new Audio Quality and Movement settings without watching or listening to them. After the user interface was explained, eight users were able to correctly use the Preview button to test their Audio Quality and Movement settings. After performing all tasks, eight users said they liked to play the videos when they made changes.

Decision: Remove the Preview button, and play the videos automatically in the second iteration of the prototype.

Rationale: Playing the videos automatically reduces the risk of error.

12. Can users accept their proposed settings and start a presentation?

Another important issue was whether or not users were able to accept their proposed settings and start a presentation. Clicking the Accept button caused the proposed settings to replace the current settings. Clicking the OK button started the real presentation and closed the QoS window. Before using the prototype, users were asked to look at the user interface and answer specific questions to evaluate if they understood the purpose of the OK button and the Accept button. Before attempting any tasks, seven users were not sure whether the purpose of the OK button was to play the videos or accept the proposed settings. Four users said the OK and Accept buttons did the same thing.

During the first task, seven users were unable to accept their proposed settings and six users were

unable to start the presentation. However, during the second task, all 10 users were unable to start the presentation or accept their proposed changes. Even after the distinction between the OK and Accept buttons was explained, three users repeatedly demonstrated difficulties distinguishing between the OK and Accept. Once they understood the distinction between the two buttons, two users indicated that they would prefer to be able to accept either of the two videos.

Decision: Remove the Accept button and the OK button and add a button labeled Use This Setting on both videos in the second iteration of the prototype.

Rationale: Removing the OK button reduces the number of steps required to start the presentation.

Removing the Accept button and adding two Use This Setting buttons simplifies the model for comparing two videos.

13. Do users understand the purpose of the ghost sliders?

Another important issue was whether or not users were able to understand the purpose of the ghost sliders displayed on the slider bars. The ghost sliders were light grey graphical overlays that indicated the current settings. Initially, only one user correctly said the ghost sliders represented the recommended settings and another user said the ghost sliders were bugs in the prototype. However, after using the prototype, all users were able to correctly explain the purpose of the ghost sliders.

Decision: Remove the ghost sliders in the prototype for the second iteration.

Rationale: Users did not initially understand the purpose of the ghost sliders. Ghost sliders were not required in the second iteration of the prototype because two sets of sliders were used instead.

Iteration Two

Method

Usability Test Revisions

During the first round of testing, several problems were observed with the usability test. Several improvements were made to the usability test, shown in Appendix E. First, despite the usability problems in the first iteration, users gave high ratings for ease of use and user satisfaction. During the introduction, shown in Appendix F, instructions were added to encourage subjects to be critical in their comments. Subjects were told, "We want you to identify problems we can fix. Don't say something is good if you don't like it. Criticism helps us to improve the user interface."

A minimal scenario was provided before introducing the user interface to demonstrate the think-out-loud protocol. The experimenter demonstrated the think-out-loud protocol by verbally describing each step out loud while creating a new document in Microsoft Word. A practice task was also added before the start of the test to encourage subjects to talk out loud. To practice the think-out-loud protocol, subjects were instructed to verbally describe each step while searching for the price of the Cannon Elura video camera using Windows Explorer.

After users adjusted the videos for the first time, specific follow-up questions were added to the usability test including, "Are you comparing the videos? When are you comparing the two videos? Are you comparing the videos before or after you decide on your settings?" During the preset selection task, subjects were also asked, "Are presets useful?" The following questions from the first iteration of usability test were eliminated in the second iteration as they were no longer applicable, "What does the video on the left represent? What does the video on the right represent? What does Proposed Setting mean? What does Current Setting mean? How does Undo differ from Cancel? How does OK differ from Accept? Can you play these videos? If so, how?"

Video Content Changes

Two changes were required due to problems with the video content used in the first iteration of the prototype. First, the audio quality was too high for the low quality setting in the "Lecture" content type. A U-Law 8khz audio CODEC was used to produce a poorer sounding audio sample. The "Talking head" video content type was also replaced because the narrator's voice was difficult to understand. A video sample of a news broadcast was recorded and digitally degraded with the same process used for iteration one video samples.

Prototype Revisions

Several changes were applied to the second iteration of the prototype based on the results of the first round of usability testing. The revised user interface is shown in Figure 10. The Accept button and the OK button were eliminated. A button labeled Use This Setting was added on both videos to allow users to select either of the two videos. This design change presented a much simpler model to the user because either video could be selected at any time. The distinction was eliminated between which video is current and which is proposed. Two sets of Video Quality and Movement slider bars were used to allow users to

change either video at any time. Adding two sets of sliders also eliminated the need for the ghost sliders in the second iteration. The labels next to the video detail slider were changed to blurry and clear. The labels next to the movement slider were changed to choppy and smooth. The title of the presets was changed to Audio and Video Quality Presets to make it clear the presets do not change the video content. The presets were renamed using names that described the settings rather than the video content to further support the message that the presets could be used to adjust audio and video quality. The labels used in the prototype tested in the first round of testing were changed to: (a) Talking without much action (News), (b) Talking with some action (Seminar), (c) Slow action with talking (Arts), (d) Fast action with talking (Sports), and (e) Fast action with music (Movies). The changes to the preset labels are summarized in Table 3.

Table 4.

Changes to Preset Labels

Preset Labels in First Iteration	Preset Labels in Second Iteration
Talk Show (e.g. Video Phone, News Anchor Desk, etc.)	Talking without much action (News)
Seminar (e.g. Lecture, Weather Report, etc.)	Talking with some action (Seminar)
Arts, Crafts, and Games (e.g. Pool, Pottery, Calligraphy, etc.)	Slow action with talking (Arts)
Stadium Sports Event (e.g. Hockey, Football, Auto Racing, etc.)	Fast action with talking (Sports)
Music Video (e.g. Concerts, Movies, Documentaries, etc.)	Fast action with music (Movies)

Drop-down combo boxes were used instead of radio buttons for presets to conserve screen space. Presets were added below the right video to make it easier to compare two presets side by side. A text field and an Add button were added for entering and saving personal presets. The videos started playing automatically to eliminate the risk that users could accept settings without previewing them first. An audio slider was added under each of the videos to make it easier to compare and change the audio settings for either of the two videos. Arrow shaped audio buttons were positioned below each video to make switching between the audio tracks more intuitive. The arrow below the video with audio switched on was highlighted in orange.

Figure 10. The user interface tested in the second round of usability testing.

Test Protocol Changes

Several changes were made to improve the testing procedure. Prior to testing, two short video demonstrations were shown to users to show users an example of a low quality Internet streaming video (Microsoft NetShow format video at 56.6 kb/s with 320x240 pixel resolution) and a high quality video clip (uncompressed AVI video format with 320x240 pixel resolution). Subjects were coached to think out loud during the test (e.g. "Why did you do that?"). Two levels of responding were recorded (a) what they did and (b) why they did it. The experimenters watched to see if users initially set the sliders in the same position for the two videos as a basis for comparison. A tabletop microphone was used to record the voices of the experimenters and the subjects. During the first iteration many subjects commented that they wanted to watch the videos full screen before they would accept the setting. In the second iteration subjects were

told that there are no full screen previews with this prototype. We were only interested in videos of the size shown here at this time.

Rating Scales

Despite the fact that users were unable to complete most tasks in the first iteration, they gave inflated ratings for ease of use and user satisfaction. In an effort to prevent ratings from clustering at the high end of the scales, a broader range of scales (from 1 to 100 shown in Appendix G) were used in the second iteration to capture ease of use and user satisfaction (e.g. 100 being very easy and 1 being very difficult) feedback. In the first iteration, users may have been confused about the values used on the rating scales because low values on a numeric scale were used to represent positive ratings and high values on a numeric scale were used to represent negative ratings (e.g. seven being very difficult and one being very easy). This was corrected in the second iteration by using high values on a numeric scale to represent positive ratings and low values on a numeric scale to represent negative ratings. Before each task, users were reminded that higher quality costs more money and that they should use the lowest acceptable settings. In the first iteration, the rating scales were all presented to users on one page. The rating scales in the second iteration were printed on separate pages.

Usability Goals

The number of errors was recorded for each scenario. The scenario passed if 7 or more of 10 users were able to complete the scenario without making an error and without requiring a hint. Ratings for ease of use and user satisfaction were also recorded for each scenario. The user interface issue evaluated in a scenario passed if 7 of 10 users rated the ease of use 70 or more on a scale from 1 to 100 (1 being very difficult and 100 being very easy). Additionally, the first scenario user interface issue passed if 7 of 10 users rated user satisfaction 70 or more (1 being very unsatisfied and 100 being very satisfied). The user interface passed if 7 of 10 users were able to successfully complete all scenarios.

Participants

Ten users (comprised of five users who participated in the first usability test and five new users) participated in the second usability test. All participants were experienced Windows users. Users were paid to participate and were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 1992).

Results and Discussion

The usability test for the second iteration was designed to evaluate the user interface in terms of ten major issues (1) Do users need one or two videos. (2) Do users need two sets of sliders for adjusting video quality for two videos. (3) Do users need two sliders to adjust audio quality for two videos. (4) Can users switch between audio tracks for the two videos. (5) Do users like the appearance of the audio controls. (6) Can users use the presets to adjust audio and video quality. (7) Are the preset and exemplar names correct. (8) Can users compare two presets. (9) Can users add personal presets. (10) Do users need both slider bars and presets. and (11) Can users accept their settings and start a presentation. Results for iteration one are summarized in Table 4. Result summaries are presented for each issue.

The decision to only consider performance data, the issues that should be included in any next design, and issues that are still unresolved are elaborated in the discussion section.

Table 5.

Results for Iteration Two by Issue and Usability Goal

Issue	Number of Users with Initial Understanding Before using Interface	Number of Users able to use the Interface	Design Passed?
1. Two videos	7 (4 New, 3 Repeat)	8 (4 New, 4 Repeat)	Passed
2. Two video sliders	9 (5 New, 4 Repeat)	9 (5 New, 4 Repeat)	Passed
3. Two audio sliders	8 (5 New, 3 Repeat)	9 (4 New, 5 Repeat)	Passed
4. Switch audio	7 (4 New, 3 Repeat)	10	Passed
5. Which side has audio	Not Evaluated	3 (1 New, 2 Repeat)	Failed
6. Use presets	8 (3 New, 5 Repeat)	0	Failed
7. Preset names	Not Evaluated	7 (3 New, 4 Repeat)	Passed
8. Compare presets	Not Evaluated	10	Passed
9. Add personal presets	Not Evaluated	10	Passed
10. Need sliders and presets	Not Evaluated	7 (3 New, 4 Repeat)	Passed
11. Accept settings	4 (2 New, 2 Repeat)	5 (2 New, 3 Repeat)	Failed

1. Do users need one or two videos?

As in the first iteration, QoS was adjusted by comparing the videos in the two windows in the second iteration. One of the primary issues evaluated in the usability test for the second iteration was whether or not users understood the purpose of the two videos and whether they used both videos. Before performing any tasks, seven users (three repeat and four new) were able to explain the purpose of the two videos. After using the user interface to adjust the videos for the first time, seven users (three repeat and four new) said they compared the two videos when adjusting quality and eight users (four repeat and four new) understood the purpose of the two videos. After completing all the tasks in the usability test, seven users (three repeat and four new) said comparing the videos was not useful. Six users (two repeat and four new) said they did not look at both videos once they were familiar with the user interface. Six users (four repeat and two new) said they would prefer just one video.

Summary: Users are able to compare two videos, but most users prefer only one.

2. Two sets of sliders for adjusting video quality for two videos

The user interface for the second iteration had two Video Detail slider bars and two Movement slider bars illustrated in Figure 11. The Video Detail slider and Movement slider on the left controlled the video on the left and the Video Detail and Movement slider bar on the right controlled the video on the right. Two related issues evaluated in the usability test for the second iteration were, (a) can users use the Video Detail and Movement sliders to control video quality, and (b) can users use two sets of sliders to adjust video quality for two videos.

Figure 11. Video Detail and Movement sliders for controlling left video (left arrows) and Video Detail and Movement sliders for controlling the right video (right arrows) used in the user interface tested in the second round of usability testing.

Before performing any tasks, 9 of 10 users (four repeat and five new) initially understood the purpose of the Video Detail slider bars. Four of 10 users (one repeat and three new) did not initially understand the

purpose of the Movement slider bars. Nine of 10 users (four repeat and five new) initially understood that the Video Detail and Movement sliders on the left controlled the video on the left and the Video Quality and Movement sliders on the right controlled the video on the right. During the tasks, 9 of 10 (five repeat and four new) users were able to use the slider bars and presets to adjust the videos. After having completed the video quality adjustment tasks, all users said that adjusting video quality was not complex.

Summary: (a) Users are able to use the Movement and Video Detail sliders to adjust QoS and (b) users are able to use two sets of sliders to adjust video quality for two videos with the user interface provided.

3. Two sliders for adjusting audio quality for two videos

There were two audio sliders on the user interface. The slider on the left controlled audio quality for the left video and the slider on the right controlled audio quality for the right video. Two other related issues evaluated in the usability test for the second iteration were, (a) can users use the Audio Quality sliders control audio quality, and (b) can users use two sets of sliders to adjust audio quality for two videos. Before performing any tasks, eight users (three repeat and five new) correctly explained the purpose of the Audio Quality sliders. Two repeat users initially said the audio quality sliders controlled volume. All users were able to explain that the Audio Quality sliders on the left controlled audio for the left video and the Audio Quality slider on the right controlled audio for the right video. During the tasks, nine users were able to use the two sliders to adjust audio quality. During the first task, one new user tried to use the audio quality sliders to adjust lip synchronization.

Summary: (a) Users are able to use the Audio Quality sliders to adjust audio quality and (b) users are able to use two sets of sliders to adjust audio quality for two videos with the user interface provided.

4. Can users switch between audio tracks?

Another important issue evaluated in the usability test for the second iteration was whether or not users were able to switch the audio between the two videos. Users could switch on the audio for the right video by clicking the right video, the right audio slider, or the grey arrow button positioned below right video as illustrated in Figure 12. The orange audio indicator pointed at the video that had sound. Users could switch on the audio for the left video by clicking the left video, the left audio slider, or the grey arrow button positioned below the left video. Clicking Audio On/Off button turned the sound on or off for both videos. Before performing any tasks, nine (four repeat and five new) users initially understood the purpose

of the Audio On/Off button. Six users (three repeat and three new) understood the purpose of the orange audio indicator. Seven users (three repeat and four new) understood the purpose of the grey audio selection button.

Figure 12. Audio indicator (left arrow) and Audio button (right arrow) used in the user interface tested in the second round of usability testing.

During the tasks, all users were able to use the audio source controls to compare two audio quality settings. After switching the audio source one user said, "I like this." After using the user interface, nine users were able to explain how the audio source controls worked.

Summary: While users were initially unable to use the audio buttons, with experience they were able to discover how to switch between audio tracks using the user interface provided.

5. Do users like the appearance of the audio controls?

Another issue evaluated in the usability test for the second iteration was whether or not users liked the appearance of the audio controls for switching audio between the two videos illustrated in Figure 12. After the first task, one new user said, "The non-standard audio buttons make the user interface look more strange." After completing all tasks, users were asked if they liked the colour of the activated audio control. Six (three repeat and three new) users did not have comments about the colour. Three (two repeat and one new) users liked orange colour and one new user said the audio controls should be black.

Summary: Users do not like how the audio controls look.

6. Can users use the presets to adjust audio and video quality?

Another important issue was whether or not users were able to use the presets to adjust video quality to the lowest acceptable level. Clicking the pull-down menu revealed a list of QoS presets. Clicking on one of the presets moved the Audio Quality, Video Detail, and Movement slider bar settings to a preset setting. Before performing any tasks, 8 of 10 users (five repeat and three new) correctly explained the purpose of the presets. However, no users attempted to use the presets to adjust QoS during the first

and second tasks. During the second task, one user opened the presets and said, "I would never use these." After the purpose of the presets was explained, users were instructed to use the presets to get the desired settings for each of the video content types. All users were able to use the presets to adjust audio and video quality.

Summary: Although many users understood the purpose of the presets, users did not initially use the presets to adjust QoS with the user interface. When users were instructed how to use the presets, all users were able to use the presets.

7. Are the preset and exemplar names correct?

Another important issue evaluated in the usability test for the second iteration was whether or not the names and exemplars used for the audio/video presets were correct. After users were informed about the purpose of the presets, for the third task users were asked to select the most appropriate preset for four video content types they had not seen before during the test. After choosing each preset, users were asked if the preset name and exemplars were correct and if any presets were missing. Seven of 10 users (four repeat and three new) said all the category and exemplar names were correct. However, two new users commented that the Arts exemplar (slow precise movement with music category) was unclear because there are many kinds of visual and musical arts, each with different audio and video quality requirements. Two repeat users commented that the user interface should provide an Audio-Only preset (if no video is desired by users). Two new users said a "Live News" setting was required and two users (one repeat and one new) said a "Cartoons" setting was required.

Summary: Users generally appear to understand the preset names provided with the user interface, but the descriptions require improvement. The Arts exemplar may be a problem.

8. Can users compare two presets?

Another important issue evaluated in the usability test for the second iteration was whether or not users were able to compare two settings using the presets. Users could compare two presets by selecting presets for either video by clicking the down arrow and selecting one of the items in the pull-down lists under either video. During the usability test, users were instructed to use the interface to compare the Talk Show setting to the Arts, Crafts, and Games setting. All users were able to use the interface to compare

two presets side by side. Two new users said they preferred to use one preset pull-down menu to compare presets.

Summary: Users were able to compare two presets with the user interface provided.

9. Can users add personal presets?

The usability test for the second iteration evaluated whether or not users can add a personal preset. Clicking the Add button saved the selected QoS settings and an entry was added to the personal preset pull-down list. A unique preset name could be entered so the settings could be recognized and used again at a later time. During the usability test, all users were able to use the user interface to add a personal preset. After completing all tasks, eight users (five repeat and three new) said personal presets are desirable. One new user said, "I also want a Delete Preset button."

Summary: Users can add personal presets with the user interface provided.

10. Do users need both slider bars and presets?

After completing all the tasks in the usability test, users were asked to answer follow-up questions to determine whether they needed both the slider bars and presets. Eight users (four repeat and four new) said it was reasonable to use both the slider bars and presets. Seven users (four repeat and three new) indicated that the slider bars alone without the presets would not be acceptable. Three users (one repeat and two new) said, "It takes too much time to adjust without the presets."

Summary: Users need both slider bars and presets.

11. Can users accept their settings and start a presentation?

Buttons labeled, "Use This Setting" were positioned on both of the two videos as illustrated in Figure 10. Clicking the Use This Setting button closed the QoS prototype and started the real presentation in a different window. Users could watch the settings for the right video by clicking on the Use This Setting button on the right video or users could watch the settings for the left video by clicking on the Use This Setting button on the left video. Another important issue evaluated in the usability test for the second iteration was whether or not users were able to accept their settings and start a presentation. Before attempting any tasks, four users (two repeat and two new) were initially unable to explain the purpose of the Use This Setting buttons. However, when users were asked why there are two buttons both labeled, "Use This Setting" nine users (five repeat and four new) were able to explain that there were two Use This

Setting buttons to choose which settings to watch. During the first task, five users (two repeat and three new) were unable to accept their proposed settings and start the presentation. During the second task, seven users (three repeat and four new) were unable to accept their changes and start the presentation. After the purpose of the Use This Setting buttons was explained, eight users (five repeat and three new) were able to accept their changes and start the presentation. Two new users said, "The labels on the buttons were not intuitive. Zero ease of use!"

Summary: The label, "Use This Setting" is inadequate.

General Discussion

This study involved the design of a user interface to negotiate quality of service for digital video. The study was conducted to address three user requirements: (a) to provide users a way to determine the minimum quality they need, (b) to enable users to achieve the minimum acceptable quality for viewing different types of content, and (c) to design a way to allow users to negotiate the quality of service in keeping with the video processing capacity of their computers.

The prototype that was evaluated in the first iteration revealed a number of usability problems. The second iteration was designed to address them. In the second iteration most of the problems identified in the first iteration were addressed, but also some new problems emerged.

Several usability problems were resolved by the end of the study. Results of the usability tests for the second design iteration indicated that users could control audio and video quality using slider bars. Users were able to use QoS presets to compare two QoS settings. Users needed both slider bars and presets for adjusting QoS.

The study identified specific user interface problems. The usability problems that remain unresolved at the end of this study and require future research, are presented below:

Most users indicated that they preferred only one video. This issue remains unresolved as users did indicate that two videos were useful when they were using the prototype for the first time. Users' responses changed only after they had had experience using the prototype. Further research to investigate the effectiveness of interfaces with only one video is required.

Our study revealed that users were able to use the audio and video quality sliders to adjust quality. However, research is required to determine which is more appropriate, the use of one or two sets of sliders.

Research is also required to define appropriate levels for the audio quality, video quality, and movement parameters as the four levels used for Video Detail and Audio Quality in the study were not distinguishable to users. Study should also be conducted on the merit of dynamically moving the range of quality and the distance between steps for the Movement and Video Quality sliders based on the type of content users chose to watch. For example, if users chose to watch a lecture, the audio slider might be biased to the high end and the movement slider might be biased to the low end. However, if the user chose to watch a sports event the movement slider might be biased toward the high end and the audio slider might be biased toward the low end. Research should be conducted on the merit of having one video quality slider (by combining the Movement and Video Detail parameters) and one audio quality slider. Study is required to determine the appropriate number of steps for the sliders.

The issue of adjusting video resolution was not addressed in the current study. A separate study is required to design an interface for adjusting video resolution. A study should be conducted where users are required to adjust video resolution either by stretching the video windows or by choosing between fixed resolutions in a pop-up list.

Many users were unable to start the presentation without a prompt. One reason for this was the users' inability to distinguish between the QoS preview and the real presentation. A second reason was the unclear functionality of the button named, "Use This Setting." Therefore, further research is required to identify a more appropriate name for this button linking it clearly to its' functionality and also to determine if a separate button should be installed for each screen.

Adjusting audio is a complex issue. Users were initially unable to switch between audio tracks using the arrow buttons below the videos. With experience, however, users were able to switch the audio between the two videos. Further research should be conducted on the merit of developing a separate interface for negotiating audio quality independently from video quality because determining which of the two videos had sound was confusing for users. Users did not like the appearance of the arrow buttons and suggested that more conventional buttons be used.

Users did not use the presets in the second iteration of the study until they were instructed to use them. Instead they preferred to use slider bars. It would be speculation to suggest that users were able or unable to use the presets as the use of the presets was optional in the usability test for negotiation of the

desired QoS. Another usability test should be conducted where users are required to use the presets to adjust quality before the purpose of the presets is explained. Other research is required to determine if presets are better presented in the form of radio buttons or pull-down lists. Further studies are required to determine if one or two preset lists would be more appropriate. Additionally, studies are required to investigate the validity of the video content categories originally identified by Apteker. Research is required to identify appropriate labels for the video categories. Research is required to understand how the categories will be used. Other categories not considered in this study should also be investigated. Some users identified categories including "Cartoons" and "Symphony Orchestra" during the current study. For example, a study should be done to determine if users are more or less sensitive to changes in video detail and frame rate for cartoons than other types of video.

Users were able to add personal presets using the interface provided. Research is required into the design and practical use of personal presets to understand how the personal presets will be used. For example, recognizing many personal presets with arbitrary names and remembering what QoS each may represent may be a challenging task for users. Research is required to identify a way to help users organize personal presets because users may not be willing or able to create unique and recognizable names for personal presets.

Users also indicated that the names and settings for the QoS presets required improvement. They also indicated they wished to have the capability to edit the name settings of presets. Some users indicated an additional requirement for an audio-only preset option. Further research is required to identify more appropriate preset option names and settings, as well as a feature for editing presets.

Further studies are required to examine the issues of cost. A study should be conducted which requires the user to make QoS decisions while considering the cost factor. Users could be required to answer specific questions about a presentation. Users could be provided limited resources to watch a video for a given period of time. Cost of service could be determined by the quality and duration of the presentation. Users could be rewarded at the end of the study based on their number of correct answers and the amount of resources they saved. Research is required to determine how the preset labels and preset categories should be represented in terms of cost.

Further studies should be conducted to determine if users are able to adjust QoS to compensate for

slow CPU speed. Movement and Video Detail are determined by the quality of the source video and the playback performance of the end-system. In the current study, the best possible system was used for video playback. A study should be conducted where users are required to adjust video quality using a PC with a slow CPU (e.g. Pentium 166) because many students use older machines for watching video.

Several usability testing techniques in this study deviated from the norm. One limitation included the exclusion of comments and rating scales used to collect user feedback. User comments and ratings were not analyzed for the first round of testing because subjects responded positively to ease of use and user satisfaction scales despite being unable to complete tasks. In the first round of testing, the weights associated with the 7-point scales were in the wrong direction. For example, low numerical values were used to represent positive responses and high numeric values were used to represent negative responses. While users were required to provide a verbal explanation for each of their numeric ratings, it is possible that the weights used for the scales contributed to the clustering of scores on the high end of the scales.

User ratings were not analyzed for the second round of testing because a confusing scale was used for ease of use and user satisfaction ratings. A 100-point scale designed by Ellermeier and Westphal (1991) was used in place of the 7-point scale to provide a way for subjects to give unique ratings for ease of use and user satisfaction. The weights were correctly assigned to address the scaling problem identified in the first round of testing. However, this scale was confusing because it was designed as a unipolar scale which did not include a neutral rating at the midpoint, yet had a label, "neutral." Further, the 100-point scale was divided into category ranges of 10 points (e.g. very easy 1-10, easy 11-20, ... very difficult 91-100). Users did not know how to use the scale. Rather than providing discrete ratings on a 100-point scale, subjects treated the scale as a 10-point scale. Despite being instructed to provide scores within the 10-point category ranges in the scale, users only provided the category ratings (e.g. 10/100, 20/100, ..., 90/100). It appeared that users found the scale difficult to use. It is likely that users did not need such a detailed scale for rating ease of use and user satisfaction. The only advantage of this scale was that it forced users to choose positive or negative ratings. However, a more appropriate scale with more choices than seven and no neutral option would have been an eight- or 10- point scale.

An additional methodological consideration involved users participating in both rounds of testing. Five subjects who participated in the first round of testing participated in the second round of testing. The

reason for including subjects in both tests was to enable users to make comments about changes observed between the first and second user interface designs. However, there is a possibility that these subjects artificially inflated the usability scores for the second iteration. This does not appear to be the case because subjects who participated in both studies performed worse than subjects who only performed in the second round of testing. The poorer performance may be attributable to confusion with the change in conceptual model of how the two videos were used, or it might be explained by individual differences.

This was an exploratory study in an applied area of user interface research. Before the findings of this study can be applied in a broader setting, primary research should be conducted to confirm the results. For example, it was proposed by Apteker et al. (1995) that faster video requires lower frame rates and slower video requires higher frame rates. While this seems counterintuitive, users in this study did appear to agree with the frame rate settings derived from Apteker. Detailed studies are required using a wide variety samples for each content type to identify if users perceive differences in quality at different frame rates. Issues such as the appropriate affordances to employ for selecting presets, adding presets, and adjusting quality require further research. The slider bars were chosen because they are commonly used in windows applications for adjusting the levels of parameters. However, the pull-down menus used for presets were chosen to conserve screen space, and might not be appropriate. Perhaps radio buttons are a more intuitive means to represent the content type presets. Further, the use of two videos to compare video quality is a novel and unusual approach that needs further study. Research should look at other means of comparing videos such as providing a split screen representing two video quality levels in a single video window. Making judgments of audio, which is serial in nature, by associating audio with two videos was confusing. It was difficult to determine which of the two videos the single audio stream was associated with. In this ambiguous situation, it would be interesting to investigate the effect of visual capture on users ability to localize the source of audio. Visual capture is defined as the tendency to associate sound with the video that has most motion. A task, unrelated to the major objectives of the thesis, was included at the end of the second iteration where users were required to look at the videos and indicate which of the two videos they thought the audio to be coming from. One video was set to 30 frames per second and the other video was set to one frame per second. Users appeared to be more likely to perceive audio to be associated with the video with a higher frame rate. This informal finding should be investigated further.

References

- Antoniades, M., Hunter, J., & Witana, V. (1998). A Review of Video Streaming over the Internet. [On-line]. Available: <http://www.dstc.edu.au/RDU/staff/jane-hunter/video-streaming.html>
- Antoniades, M., Richards, A., Rogers, G., & Witana, V. (1998). Mapping user level QoS from a single parameter. In Second IFIP/IEEE International Conference on Management of Multimedia Networks and Services. Versailles, November 1998. [On-line]. Available: <ftp://ftp.tip.csiro.au/pub/people/rliu/marshnet.ps.gz>
- Antoniades, M., Richards, A., Rogers, G., & Witana, V. (1998). SuperNOVA Prototypes [On-line]. Available: <http://www.dstc.edu.au/cgi-bin/redirect/rd.cgi?http://archive.dstc.edu.au/FastWeb/>
- Apteker, R. T., Fisher, J. A., Kisimov, V. S., & Neishlos, H. (1995). Video Acceptability and Frame Rate. IEEE Multimedia, Vol. 2, No. 3, 32 – 40.
- Aurrecoechea, C., Campbell, A.T. & Hauw, L. (1998). A Survey of QoS Architectures. ACM/Springer Verlag Multimedia Systems Journal, Special Issue on QoS Architecture, Vol. 6, No. 3, 138-151.
- Balter, P., Berts, J., & Persson, A. (1998). Objective and subjective quality assessment of compressed digital video sequences. Unpublished Master's thesis, Chalmers University of Technology, Goteborg, Sweden [On-line]. Available: <http://www.etek.chalmers.se/~e4joni/xjobb.html>
- Bochmann, G. von, Dssouli, R., & Hafid, A. (1998). Distributed Multimedia Applications and Quality of Service: A Review. Electronic Journal on Network and Distributed Processing, No. 6, 1-50.
- Bochmann, G. von, Dssouli, R., & Hafid, A. (1998). A Quality of Service Negotiation Approach with Future Reservations: A Detailed Study. Computer Networks and ISDN Systems, Vol. 30, 8.
- Bochmann, G. von, Fischer, S., & Salem, M.-V. O. M. (1997). Application Design for Cooperative QoS Management. [On-line]. Available: <http://beethoven.site.uottawa.ca/DSRG/SelectedPublications.htm>
- Brown D. & Willis D. (1998). Videoconferencing on Frame Relay Networks - How Much Bandwidth to Buy? Network Computing [On-line]. Available: <http://www.nwc.com/917/917f1side1.html>
- Ellermeier, W. and Westphal, W. (1991). On the "absoluteness" of category and magnitude scales of pain. Perception and Psychophysics, Vol. 49, No. 2, 159-166.
- Fry, M., Ray, P., Seneviratne, A., & Witana, V. (1997). Managing QoS in Multimedia Services. Journal of Network and Systems Management, Vol. 5, No. 3.
- Huard, J.-F. & Lazar, A. A. (1997). On End-to-End QoS Mapping. [On-line]. Available: <http://comet.ctr.columbia.edu/~jfhuard/pub/iwqos97.ps.gz>
- Michelitosch, D., Ott, M., Reininger, D., & Welling G. (1997). An Archetecture for Adaptive QoS and its Application to Multimedia Systems Design. Computer Communications Journal, Special Issue on Quality of Service in Distributed Systems, Guest Editors, A.T. Campbell and S. Keshav (ENSIM Inc.), Vol 21 No. 4, April 1997 [On-line]. Available: <http://comet.columbia.edu/~campbell/special.html>
- Richards, A. & Witana, V. (1997). A QoS Framework for Heterogeneous Environments. Available: <http://www.dstc.edu.au/cgi-bin/redirect/rd.cgi?http://archive.dstc.edu.au/FastWeb/>
- Watson, A., & Sasse, M. A. (1998). Evaluating Audio and Video Quality in Low-Cost Multimedia Conferencing Systems. [On-line]. Available: <http://www.cs.ucl.ac.uk/staff/A.Watson/iwc.html>

Watson, A., & Sasse, M. A. (1998). Measuring Perceived Quality of Speech and Video in Multimedia Conferencing Applications. ACM Multimedia '98, 98, 55 – 60.

Appendix A – Scenarios Iteration One

DISCOVERY PHASE:

Please examine the main window and describe what each item appears to be. Please begin at the top of (point at top left) the screen and go down the screen from left to right describing what you think each item is. Please explain would happen if you click each of the items that appear to be buttons.

Follow up questions to Discovery Phase.

1. Do you like the way this screen looks?
2. What does the video on the left represent? And the one on the right?
3. What does "Proposed Setting" mean?
4. What does "Current Setting" mean?
5. How does "Undo" differ from "Cancel"?
6. How does "OK" differ from "Accept"?
7. Can you play these videos? If so, how?

UNSTRUCTURED TASKS:

Task 1. You are interested in watching an action movie being broadcast over the Internet. You have selected to view the action movie at the film.com web site. You have seen this movie a few times before, so you want to minimize the price. How would you use this application to ensure you will get at least the lowest acceptable quality when you watch the action movie? Now, actually use the interface to watch the action movie with the lowest acceptable level of quality.

Follow up questions to Task 1.

1. Please rate the ease of use of this application.

[very easy, easy, neutral, difficult, very difficult]

2. Please indicate your satisfaction with user interface for the task you just performed.

[very satisfied, satisfied, neutral, unsatisfied, very unsatisfied]

Task 2. Explain what you would do to slightly improve or decrease the quality for this presentation. Now, use the user the interface to slightly improve or decrease the quality for this presentation

Follow up questions to Task 2:

1. Can you describe the difference your changes made?
2. How can you tell if there is difference?
3. What do the graphics mean on the slider bars and below the slider bars?
4. Please indicate the ease of judging video quality using the user interface

[very easy, easy, neutral, difficult, very difficult]

5. What might make it easier for you to tell judge the audio and video quality?

DIRECTED TASKS:

Task 3. "Use the PRESETS to get the desired settings for EACH the following video content types."

Follow up questions to Task 3.

1. Why did you choose this setting?
2. Is this category useful?
3. Are there any categories missing?
4. Are these the right categories?
5. Are the category names right?
6. Are the category examples OK?

EVALUATE PRESETS

Task 4. Choose an appropriate preset for each video content example. Then use the slider bars to adjust the quality to where you think it should be. Even if the preset settings are OK, use the sliders to adjust them further.

Task 5. Use the PRESETS to compare the "Talk Show" setting to the "Arts, Crafts, and Games" setting.

Task 6. Use the AUDIO buttons to compare the sound quality of the "Talk Show" setting to the "Arts, Crafts, and Games" setting.

Task 7. Use the Movement Slider bars to increase the video quality for the proposed setting.

Task 8. Add a preset.

Task 9. Apply your changes.

POSTTEST QUESTIONS

1. Would you prefer if the video service provider prepared a recommended setting for the presentation?
2. Which do you prefer to use, Presets or Sliders?
3. Are personal presets necessary or desirable?
4. Is it reasonable to use presets and sliders?
5. Would sliders alone be acceptable?
6. Are the captions for the videos (current/proposed) adequate? Do you have better suggestions?

7. Is the number of steps for the slider bars (four) OK?
8. Is it OK to have discrete settings or should these be continuous?
9. Are the video comparisons helpful?
10. Did you look at both videos?
11. Would a single video be OK during the previews?

[absolutely one, one, somewhat one, neutral, somewhat two, two, absolutely two]

12. Would you want to be able to change the video quality on the fly like this? Would you use it?
13. Is it too complex to change/adjust?
14. Is this preview useful or would you rather make changes while watching the actual presentation?
15. Would you want to adjust your settings during a presentation?
16. Would you like to use the slider bars during the presentation? Presets? Presets and Slider bars?
17. Would you want to see two videos if you made changes during the presentation like you have now?
18. Would a single video window be enough for making changes during the presentation?

[absolutely one, one, somewhat one, neutral, somewhat two, two, absolutely two]

19. Does it make sense for the "Accept" button to be on the picture? Do you like this on the picture?
20. How do the speaker buttons work? Are these buttons large enough?
21. Why is the "Undo" button on the picture? Do you like this on the picture?

Appendix B – Informed Consent Form

Informed Consent

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent must provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Present Study: An Effective User Interface for Negotiating Quality of Service for Video

Research Personnel: The following people are involved in this research project and may be contacted at any time. Jeff Bos (Principal Investigator, 520-6628), Ronald Boring (Research Assistant, 520-6628), and Dr. R. Dillon (Faculty Advisor, 520-6629). Should you have any ethical concerns about this study please contact Dr. M. Gick (Chair, Dept. of Psychology Ethics Committee, 520-2600, ext. 2664) or Dr. Kim Matheson (Chair, Department of Psychology, 520-2600, ext. 2648).

Purpose: The purpose of this study is to design and evaluate a user interface for adjusting the quality of audio and video. Setting the quality of audio and video quality is a complex task. We are interested in designing a user interface that is easy to use.

Task Requirements: You will adjust the quality of 5 different video clips using a computer prototype. You will be asked to speak out loud when you use the computer prototype. Your comments, the errors you make, and the ratings you assign for satisfaction will be recorded.

Duration and Locale: The experimental session will last approximately 1 hour. Testing will take place in Room 314 of the Social Sciences Research Building at Carleton University.

Anonymity/Confidentiality: The data collected in this experiment is confidential. The audio portion of the sessions will be recorded on videotape. All data is collected such that your name is not associated with the data. The coded data is made available only to the researchers associated with this project.

Right to Withdraw: You have the right to withdraw from the experiment at any time without academic penalty.

I have read the above description of the Video Quality study and I understand the conditions of my participation. My signature indicates that I agree to participate in the experiment.

Participant's Name: _____

Participant's Signature: _____

Researcher's Name: Jeff Bos

Researcher's Signature: _____

Date: _____

Appendix C - Iteration One Usability Test Introduction

USER #_____ TIME AND DATE:_____

INTRODUCTION:

- Programs exist which can be used for listening to music and watching videos stored on other computers over the Internet. Some examples of these programs are RealVideo, Apple QuickTime, and Windows NetShow.
- I will demonstrate a streaming video for you now (**Demonstrate Bloomfield**).
- New technologies have recently been developed called, "Quality of Service" that allow users to set quality they are willing to pay for. For example, higher video quality costs more money and lower quality video costs less money.
- I will demonstrate video at the maximum possible quality (**Demonstrate DVD**).
- This technology would let you watch movies on your computer. You could use this to tune into a live broadcast such as a lecture or a concert.
- This is a usability test of a Windows application for setting the quality of digital video clips. The application is designed to help the user set the quality of streaming video to the **lowest acceptable** level of quality.
- The usability test will be used to evaluate the prototype and is not a test of your abilities. I am interested in identifying problems with this design so I can find out how to make the application easier to use.
- During the usability test I will ask you to perform certain tasks and answer questions about the things you see on the screen. There are no right or wrong answers to the questions I will ask you.
- When you are performing each task, please describe what you are thinking out loud to help me understand what you think is happening. For example, I will show you how I would think out loud when I create a new document in Microsoft Word:

"First I would click on blue 'W' icon to open Word. OK, I will try to figure out how to open a new word file. In Windows programs I usually find my files under the 'file' button in the toolbar so I will click on 'file' and look for 'new.' OK, I see 'new' so I will click on it. Hmm, I did not expect this to happen. OK, I want to create a blank document so I will click OK. Alright, that is what I expected to happen."

- Now I would like you to practice thinking out loud. Use the Windows **Explorer** to access the Altavista web site and search for the price of the "**Canon Elura**" video camera. As you use the computer to complete this task describe what you are thinking during every step of the process.

The user will be given the opportunity to practice thinking aloud.

- My assistant will be writing notes during the session to record your comments and to make a note of the things that cause you problems.
- May record this session on videotape, so I can review what was said in case I have trouble understanding my notes? Only your voice will be recorded during the session. Your image will not be recorded. No one else will hear the tape.
- Do you have any questions about the usability test before we begin?

Appendix D – Rating Scales for Iteration One Usability Test

Please rate the ease of use of this application.

1	2	3	4	5	6	7
very easy	easy	neutral	difficult	very difficult		

Please indicate your satisfaction with user interface for the task you just performed.

1	2	3	4	5	6	7
very satisfied	satisfied	neutral	dissatisfied	very dissatisfied		

Appendix E – Scenarios Iteration Two

DISCOVERY PHASE:

Please examine the main window and describe what each item appears to be. Please begin at the top of (point at top left) the screen and go down the screen from left to right describing what you think each item is. Please explain would happen if you click each of the items that appear to be buttons.

Follow up questions to Discovery Phase.

1. Do you like the way this screen looks?
2. Why are there two buttons both labeled Watch This Video?
3. How does the Cancel button differ from clicking on the X at the top corner of the window?

UNSTRUCTURED TASKS:

Task 1. You are interested in watching an action movie being broadcast over the Internet. You have selected to view the action movie at the film.com web site. You have seen this movie a few times before, so you want to minimize the price. The higher the quality, the more you will have to pay. Look at the user interface on the screen and explain how you would use this program to view a lecture at the lowest acceptable level of quality. Then, actually use the user the interface to view a lecture at the lowest acceptable level of quality.

Follow up questions to Task 1.

1. Are you comparing the videos?
2. When are you comparing the two videos? Are you comparing the videos before or after you decide on your settings?
3. Please rate the ease of use of this application.

[very easy, easy, neutral, difficult, very difficult]

4. Please indicate your satisfaction with user interface for the task you just performed.

[very satisfied, satisfied, neutral, unsatisfied, very unsatisfied]

Task 2. Explain what you would do to slightly improve or decrease the quality for this presentation. Then, use the user the interface to slightly improve or decrease the quality for this presentation.

Follow up questions to Task 2.

1. Can you describe the difference your changes made?
2. How can you tell if there is difference?
3. What do the orange arrow graphics mean below the two videos? Do you like the colour?
4. What might be the purpose of having two videos?

5. Did you compare the two videos?
6. Please indicate the ease of judging video quality using the user interface.

[very easy, easy, neutral, difficult, very difficult]

7. What might make it easier for you to tell judge the audio and video quality?

DIRECTED TASKS:

Task 3. "Use the PRESETS to get the desired settings for EACH the following video content types."

Follow up questions to Task 3.

1. Why did you choose this setting?
2. Is this category useful?
3. Is the category name right?
4. Is the category example OK?
5. Are the presets useful?
6. In general, are there any categories missing?
7. Are these the right categories?

EVALUATE PRESETS

Task 4. Choose an appropriate preset for each video content example. Then use the slider bars to adjust the quality to where you think it should be. Even if the preset settings are OK, use the sliders to adjust them further.

Task 5. Use the PRESETS to compare the "Talk Show" setting to the "Arts, Crafts, and Games" setting.

Task 6. Use the AUDIO buttons to compare the sound quality of the "Talk Show" setting to the "Arts, Crafts, and Games" setting.

Task 7. Use the slider bars to increase the video quality for the left video.

Task 8. Create your own personal preset named "My preset."

Task 9. View the presentation you are paying to see using the settings you have recently selected.

POSTTEST QUESTIONS

1. Which do you prefer to use, Presets or Sliders?
2. Are personal presets necessary or desirable?
3. Is it reasonable to use presets and sliders?

4. Would sliders alone be acceptable?
5. Are the captions for the videos (Trial Setting A/Trial Setting B) adequate? Do you have better suggestions?
6. Is the number of steps for the slider bars (four) OK?
7. Is it OK to have discrete settings or should these be continuous?
8. Are the video comparisons helpful?
9. Did you look at both videos?
10. Would a single video be OK during the previews?

[absolutely one, one, somewhat one, neutral, somewhat two, two, absolutely two]

11. Would you want to be able to change the video quality on the fly like this? Would you use it?
12. Is it too complex to change/adjust?
13. Is this preview useful or would you rather make changes while watching the actual presentation?
14. Would you want to adjust your settings during a presentation?
15. Would you like to use the slider bars during the presentation? Presets? Presets and Slider bars?
16. Would you want to see two videos if you made changes during the presentation like you have now?
17. Would a single video window be enough for making changes during the presentation?

[absolutely one, one, somewhat one, neutral, somewhat two, two, absolutely two]

18. Does it make sense for the "Watch this Video" button to be on the picture? Do you like this on the picture?
19. How do the audio buttons work? Are these buttons large enough?

Appendix F - Iteration Two Usability Test Introduction

USER # _____ TIME AND DATE: _____

INTRODUCTION:

- The purpose of this study is to look for usability problems. In order to help us with this study we will ask you to be critical in your comments, so we can identify problem areas that make this application difficult to use.
- Programs exist which can be used for listening to music and watching videos stored on other computers over the Internet. Some examples of these programs are RealVideo, Apple QuickTime, and Windows NetShow.
- I will demonstrate a streaming video for you now (**Demonstrate Bloomfield**).
- New technologies have recently been developed called, "Quality of Service" that allow users to set quality they are willing to pay for. For example, higher video quality costs more money and lower quality video costs less money.
- I will demonstrate video at the maximum possible quality (**Demonstrate DVD**).
- QoS technology would let you watch movies on your computer. You could use this to tune into a live broadcast such as a lecture or a concert.
- The prototype is designed to help the user set the quality of video clips to the **lowest acceptable** level of quality. The usability test will be used to evaluate the prototype and is not a test of your abilities. I am interested in identifying problems with this design so I can find out how to make the application easier to use.
- During the usability test I will ask you to perform certain tasks and answer questions about the things you see on the screen. There are no right or wrong answers to the questions I will ask you.
- When you are performing each task, please describe what you are thinking out loud to help me understand what you think is happening. For example, I will show you how I would think out loud when I create a new document in Microsoft Word:

"First I would click on blue 'W' icon to open Word. OK, I will try to write a FAX. In Windows programs I usually find my files under the 'file' button in the toolbar so I will click on 'file' and look for 'new.' OK, I see 'new' so I will click on it. Hmm, I did not expect this to happen. OK, I want to create a blank document so I will click OK. Alright, that is what I expected to happen."

1. Now I would like you to practice thinking out loud. Use the Windows **Explorer** to access the Altavista web site and search for the price of the "**Canon Elura**" video camera. As you use the computer to complete this task describe what you are thinking during every step of the process.
- My assistant and I will be writing notes during the session to record your comments and to make a note of the things that cause you problems. This will help me to improve the user interface.
 - May I record this session on videotape, so I can review what was said in case I have trouble understanding my notes? Only your voice will be recorded during the session. Your image will not be recorded and no one else will hear the tape.
 - Do you have any questions about the usability test before we begin?

Please look for things that need to be improved. We want you to identify problems we can fix. Don't just say it's good if there is something you don't like.

Appendix G – Rating Scales for Iteration Two Usability Test

Please rate the ease of use of this application.

1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
very easy		easy		neutral		difficult		very difficult	

Please indicate your satisfaction with user interface for the task you just performed.

1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
very satisfied		satisfied		neutral		dissatisfied		very dissatisfied	