Marquee: A Tool For Real-Time Video Logging

Karon Weber and Alex Poon

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
Weber@parc.xerox.com 415-812-4344

ABSTRACT

We describe Marquee, a pen-based video logging tool which enables users to correlate their personal notes and keywords with a videotape during recording. We present our observations about coordinating the task of logging in real time and describe the three phase, user-centered approach we took in designing the tool. Our early work explored the functionalities needed by users to successfully create a log. In the second phase we focused on testing our intuitions about logging by conducting user studies with paper mock-ups. In the final phase, we implemented a working prototype system and placed it in a setting to see if it supported people logging in real time.

KEYWORDS

video indexing, video annotation, gestural interfaces, pen-based computing, user interfaces, user studies, multimedia

INTRODUCTION

Advancements in video recording technologies have simplified the capture process to the point where the creation of large collections of video is commonplace. To create access points into the recorded footage, users generate a list of location pointers, an *index*, and/or apply critical or explanatory descriptors, *annotations*. The resulting collections of notes, a *log*, supports the users accessing and retrieving footage from these archives. For people to utilize the video records they collect, they need methods and tools to create logs easily that enable access and retrieval from their video archives.

Recently, there have been a number of innovative approaches to making logging video an easier task. Many of these efforts have focused on providing new methods for describing the content and context of the footage [1,2,4,7,9,13,15,17]. In these cases, the indexer manually provides personal interpretations of the material by applying descriptions of the footage correlated to time stamps embedded on the videotape. The granularity and form of these descriptions varies according to the system used, the logger's community of practice and the potential applications of the material. Although these systems allow users to create detailed descriptions of the video, they require them to perform *post-capture reviews* of the tape. Post-processing footage without the aid of a general index or map into the material requires the user to spend an extensive amount of time logging the footage.

An alternative approach to manual indexing is being explored by those interested in processing the video and audio signals automatically [10,11,12,14,16,18]. These methods detect changes in the signals and correlate the points of change with the time stamps on the video. The goal of these methods is to process the footage and to generate an index automatically. Although potentially effective, these methods have not shown that they can extract the same degree of detail as do human loggers. For example, an index created on the basis of speaker segmentation of the material would tell us who was speaking but not the substance of the talk.

We believe that users need a method for generating a rich set of annotations during video recording, that is, in *real time*. In this paper we describe *Marquee*, a tool to provide users with a means for correlating their personal notes with the recorded footage. It presents our observations about coordinating the task of logging in real time and describes the method we used in designing the tool and the resulting prototype. Finally, we suggest further directions for investigation.

SUPPORTING THE TASK OF LOGGING

Our prior investigations into annotation and indexing [17] have demonstrated that describing the contents of a videotape is a personal task and varies in form as well as in the level of detail. In addition, users required a non-restrictive interface to support their wide variations in notation styles. Some users, such as anthropologists and linguists [13], desire a very fine grained transcript of spoken words as well as descriptions of visual and auditory actions. For other users, such as home video enthusiasts, a simple outline or

CHI94-4/94 Boston, Massachusetts USA

sparse notes are enough to support their retrieval needs. In designing our system, we wanted to support the varied needs of loggers as well as take advantage of their current real-time note-taking practices. These observations also influenced our choice of an input device for our system.

Real Time Note Taking

Many people currently take handwritten notes even if an event is being recorded. For example, at PARC, many lectures and meetings are regularly recorded on videotape, but it is not unusual for the participants in these settings to take paper notes as well. This observation led us to design a prototype system that enables users to take notes in real-time and have those notes provide them access into the videotape. Real-time annotation provides a skeletal structure for retrieval; a first pass of an iterative process of logging and review [8]. This is especially true for users who desire a more complete description of the recorded material, as the first-pass log affords them an overview of the footage upon which they can elaborate.

Input Devices

We had two options for the tool's input device—a typing keyboard or a wireless electronic pen. Using a keyboard would potentially allow for rapid text entry and searchable notes; however, an electronic tablet and stylus supports the current work practices of users, changing only the surface on which they are taking notes and not their style of note taking. Marking instead of typing also allows users to add notations that are not words, such as symbols and sketches.

Research in using pen based computing [3,5,19] supports our choice of a pen interface. However, we were hesitant to employ this technology without having reliable handwriting recognition for later searching. At the same time, we chose not to constrain our users by requiring them to print characters, or work within gridded writing areas. To avoid the limitations of current character recognition, we decided to support retrieval through the manual application of machine-readable *keywords*, a concept which we will describe later.

DESIGN METHOD

We undertook the task of designing the logging tool in three stages. The early work focused on understanding the design space. By constructing Hypercard envisionments of the system, we explored the types of functionalities users need to successfully create an index. The second phase concentrated on testing our intuitions about the logging task by conducting user studies with paper mock-ups. Finally, we implemented a working prototype and placed it in a setting to see if it would support people creating logs in real time.

Envisionments

Our first step was to construct several different Hypercard prototypes which explored the issues of logging video in real time. We decided to support two forms of notation: handwritten notes and keywords. The handwritten notes provided personal interpretations of the recorded event, while keywords allowed for structuring and/or categorization of the footage. Therefore, each prototype included a note-taking area on which users could take handwritten notes as well as a set of buttons that represented keywords. Users could toggle on and off the buttons in real-time when particular events in the video began and ended (Figure 1).

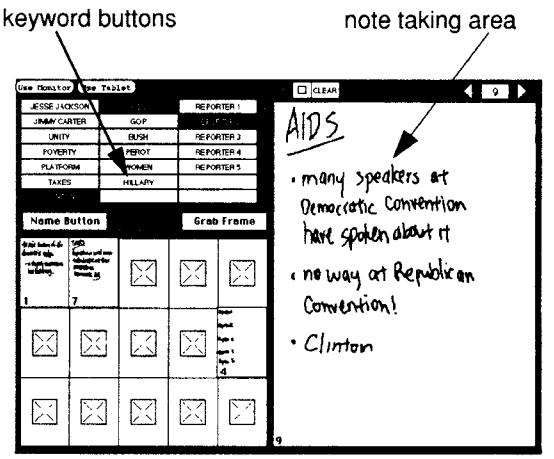

Figure 1: An Early Hypercard Prototype

In designing these early prototypes, we made two important assumptions about the interface. We required the user to take notes on small, independent sheets and assumed that the user could navigate among those individual sheets. The second assumption was that the user could coordinate the task of writing notes with the task of toggling keyword buttons, both in real time.

Paper Mock-Ups

We conducted a series of short studies to see how people would coordinate taking notes with the constraints of processing information in real time. Exploring the design issues by using paper mock-ups facilitated quick feedback on our ideas. In addition, they helped us to focus on the issues of supporting real-time logging instead of system implementation details, therefore allowing us to drive the design of the tool's interface from the user studies.

Taking Notes in Real Time. The first user trial simply asked "what would people do if asked to log segments of video in real time?" We gave the subjects a pen, a stack of 3x5 cards (to represent the tablet size), and a video monitor that displayed video segments with visible timecode. Our subjects were volunteers from the PARC community who had a wide range of expertise in logging video. Some had never logged a videotape, while others were members of the video services department. At the start of each session, we read short descriptions about the materials they were going to watch and provided them with purposes for logging the videotapes. For example, we showed them a political round table discussion and asked them to log the tape so that their boss could

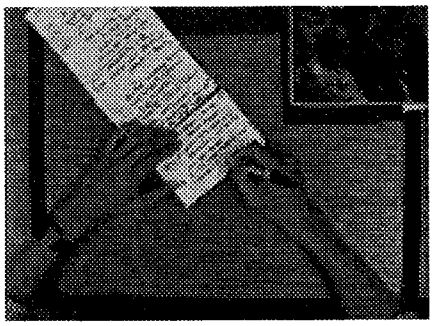

Figure 2: Subjects Used the Index Cards as a Long Scroll

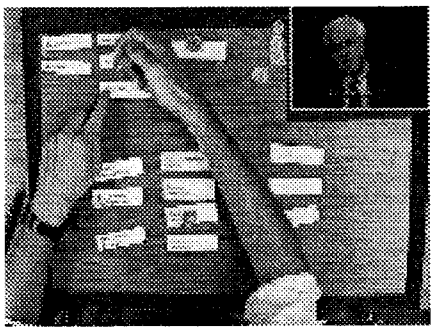

Figure 3: Subjects Used Coins and Labels to Toggle Keywords

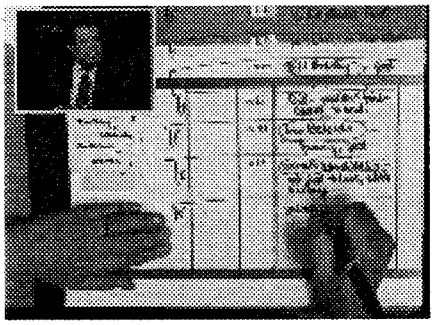

Figure 4: Subjects Took Notes and Used Keywords with Paper Mock-Up of Marquee

later retrieve portions of the tape and find each guest's political position.

We observed that all of the subjects could write notes in real time. They all utilized the visible timecode on the monitor and marked times when something happened which interested them. We also observed that none of the loggers had problems managing the set of index cards; however, all of the subjects arranged their note cards in a linear pile as if creating a long scroll (Figure 2). From this exercise we confirmed that logging is perceived as a linear task and determined that we needed to support users seeing a range of their notes in context, thus "preserving the linear integrity" [1] of the material.

Applying Key Words In Real Time. Our next set of observations revolved around the notion of applying keywords to recorded material in real time. We wanted to know if it was feasible to have people coordinate the task of note taking with the task of categorizing the information. Our initial intuition was to explore buttons as a mechanism for turning on and off keywords. We proposed that users would have a suite of buttons at their disposal and could toggle them on and off. The user could create and apply keywords while also continuing to write notes. These buttons would automatically be attached to the timecode on the videotape and, therefore, create an index for the tape.

To test this idea, we provided our subjects with a set of coins and paper button labels. To turn on a keyword, the logger simply placed a coin on a label and removed it when it was to be turned off (Figure 3). Again, we asked them to watch several types of video material, gave them a purpose for logging, and asked them to utilize the 3x5 cards to take handwritten notes. This was a very difficult task for most of our subjects. They all complained that they could not keep up with both taking notes and turning on and off the buttons. Most of them concentrated on either the keywording or note taking but not both. We observed that the coordination of the two tasks as well as digesting the material was a frustrating exercise. We had made the job of indexing more difficult rather than simpler.

As a result of our studies, we recognized that we had to rethink the application and management of keywords. Our subjects needed a technique which would allow them to apply a keyword without the pressures of real time. In addition, they needed a way to see and navigate through their collections of notes; single pages did not support them. On the basis of these observations, we rethought the interface design, incorporated solutions to these problems, and then tested the redesign with another group of study participants.

Paper Marquee

We built a paper mock up of the redesigned system (Figure 4) using a roll of drafting vellum set up so that the writing area would scroll up and down. To support the attachment of a timecode from the tape to the log, users created a *timezone*, a large writing area correlated to a single time on the tape. The horizontal surface of the vellum was divided into three columns: a *keyword palette*, a *keyword striping area*, and the *timezone note-taking area*. The keyword palette is an area in which keywords are created, displayed, and organized. The striping area allows for the attachment of keywords to the timezones.

To operate the paper system, the subjects performed a series of pen-based commands which were supported by manual execution from the design team. For example, we provided a pad of small, yellow labels on which subjects were able to write the keyword terms and their identification numbers. If they created a keyword from their hand written notes (circled the strokes), the designers would write the keyword on a yellow label and then attach it to the keyword palette. Timezone creation occurred when the subject drew a horizontal line across the vellum, and the timecode was manually entered by one of the designers.

We asked our study participants to watch a tape and log with the purpose of documenting the opinions of five journalists who were discussing the third night of the 1992 Democratic convention.

We found that people were easily able to synchronize the task of note taking with the task of applying structure to their data. One participant created 38 timezones and applied 15 keywords during the twenty-six minute discussion (Figure 5). He shifted between creating timezones, writing notes, and applying keywords by consistently creating the timezone first and then adding the notes. The application of keywords

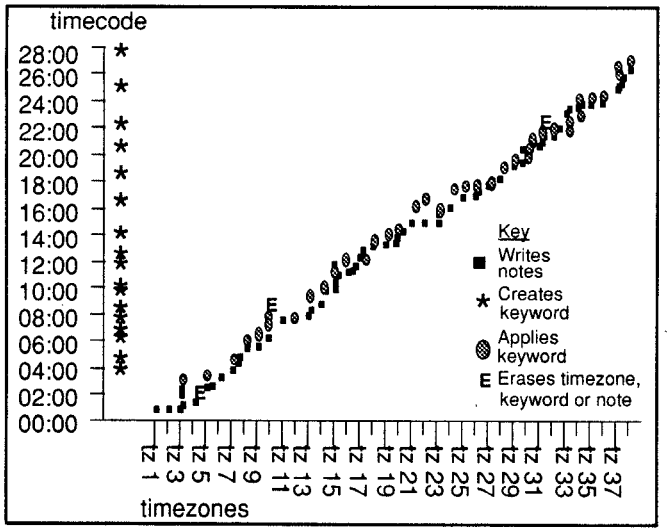

Figure 5: Usage Pattern of One Study Participant Logging with the Paper Mock-Up

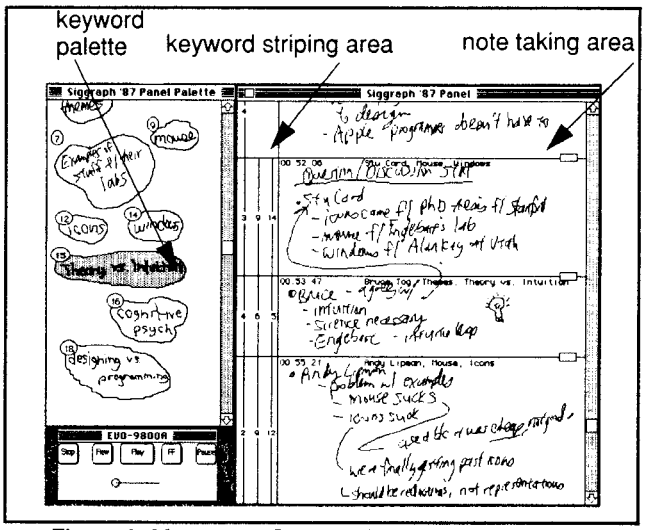

Figure 6: Marquee's Current User Interface

usually occurred after he had written his notes, although several times he returned to a timezone to add keywords. In addition, this subject also erased a timezone, a keyword and several words from his log.

From conducting these user studies we noticed that users could easily switch between taking handwritten notes, creating and applying keywords, and creating timezones. The paper mock-up of our system successfully supported users creating a log in real time.

MARQUEE IMPLEMENTATION

The design envisionments and paper mock-up studies led to Marquee, a prototype real-time logging tool. Marquee runs on a Wacom flat tablet display which is attached to a Macintosh computer. The computer is also connected via the serial port to a Hi8 recording device. To log sequences of events using handwritten notes, keywords, and pen-based gestures, users manipulate an electronic stylus directly on the tablet as if it were a piece of paper (Figure 6).

Real-Time Logging with Marquee

The Marquee log is constructed as the user takes notes, dividing the scrolling note-taking area into a series of timezones. A timezone is created by drawing a line across the horizontal plane of the tablet (figure 7a). This gesture results in the rendering of a computer-generated line and the display of the current time stamp from the videotape. A user may then make notes appropriate to the time of the event. All of their notes are attached to the timezone in which they are written.

Utilizing the timezone technique instead of time stamping each stroke [6] enables the user to augment their log at any point in time. For example, in our user studies we noticed that several of our users scrolled back to previous timezones to add notes. If each stroke were timestamped to the current timecode rather than to the timezone, the user could not correlate their notes to the times of previous events. The timezone technique provides the means to append the notes either later in the recording session or during a post-capture review of the videotape.

Users create keywords by simply circling any portion of their handwritten notes (Figure 7b). The image within the circle is assigned a unique identifier and copied to the keyword palette. Keywords can also be created in the keyword palette by performing the same gesture. To support quick access to a large number of keywords, users can rearrange the organization of their keywords by dragging them to new locations within the keyword palette.

The process of applying keywords to timezones can be thought of as brushing keyword paint along specified regions of the log. The user simply dips the stylus in the keyword palette (taps it) and paints a vertical line along the timezones they wish to apply the keyword (Figure 7c). Immediately after they complete the gesture, a vertical line appears in the column. Attached to the line is the unique identifier of the keyword. This method of applying keywords eliminates the need for the user to toggle keyword buttons as events begin and end. Although each keyword must be applied individually, attachment can be made during lulls in the recording session or during post-process review.

The user can erase notes by performing a horizontal scratch-out gesture. They can scratch out parts of their notes as well as remove the keyword stripes. They can also scratch out a keyword, which in turn removes all references to the keyword. The decision to recognize the horizontal gesture and not a vertical scratch out gesture eliminated the chance that the system would mistake certain marks (for example, cursive Ws and Ms) for the scratch-out gesture.

The physical size of the timezones can be changed by dragging a small tab attached to each timezone. This allows the user to enlarge an existing timezone to add new notes or insert a new timezone. The timecode associated with the new timezones reflects the current time on the videotape.

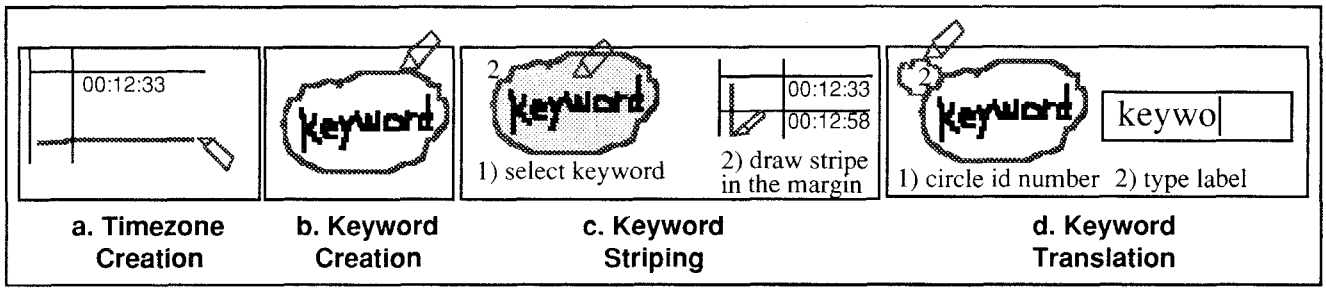

Figure 7: Marquee's Gestural Commands

Before a Marquee log can be stored in a database, users must translate the key words into machine readable text. To accomplish this they simply circle the unique identifier attached to the keyword (Figure 7d). This gesture brings up a small window that contains a bit-mapped representation of the keyword and a text field in which the users can type in the keyword label from a keyboard. The typed words are displayed next to the time stamp in the note taking area. A small circle is placed around the unique identifier in the keyword palette and striping area to indicate that the keyword has been translated to text.

Post-Process Reviewing with Marquee

During the design of Marquee, we focused on creating an interface to support logging, and did not resolve the issues of retrieving video from a Marquee log. Users currently have just two simple methods for accessing segments on a logged videotape. One, users circle a timecode in the Marquee log to cue the videotape to the time stamp or two, users control the VCR directly.

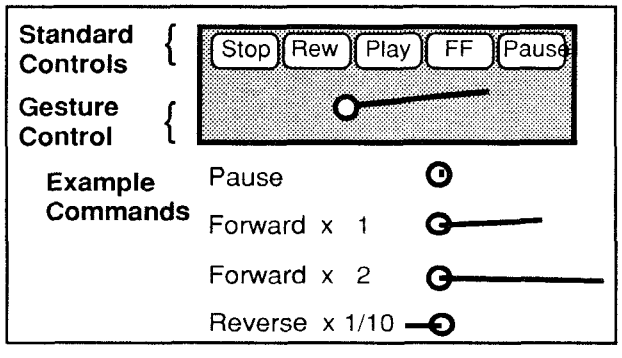

Figure 8: Pen-Based VCR Control

For more direct control of the videotape, we built a pen based interface to drive the VCR (Figure 8). For standard operations, the user has access to five buttons which command the VCR to play at set speeds. Users may also control the VCR without having to change control devices or shift their focus from the video screen. To send a command to the VCR, the user draws a horizontal line within the window. The direction of the line determines the direction that the tape will play, while the length of the line determines the speed at which the VCR will play. Shuttling back and forth to pinpoint an exact location only requires the user to shift the pen from left to right. A tap of the stylus within the window pauses the tape. The advantage of this video control interface is that it allows users to control the VCR with the same instrument with which they take notes, and to keep their eyes focused on the video screen rather than on the control interface.

MARQUEE IN USE

Our goal has been to explore methods for creating annotations in real time. After developing the prototype we conducted informal evaluations of the tool to make initial assessments about the usability of the note-taking interface and to begin to develop an understanding about the types of notes people make while using a system in real time.

In one study, we offered Marquee to a large group of researchers at PARC who were planning to hold a three-hour brainstorming meeting. The day before the meeting, we held a training session for the discussion participants, in which we demonstrated how to use the tool, and suggested approaches to logging. This training provided an opportunity for the participants to practice making notes and drawing the commands.

Three of the 15 participants took turns creating a log into the videotape in real time. The tool easily accommodated their diverse notation styles, and they were able to coordinate the application of keywords with hand written notes. Using Marquee did not intrude upon the brainstorming process, and none of the meeting participants mentioned that they were bothered by the presence of the Marquee logger.

Observations About the Log

Three people used Marquee during the meeting: one was a technical writer who had walked in cold to the meeting and was asked to begin documenting the activity; the second was a researcher who was familiar with the project but was only an observer of the meeting; and the third had attended the training and was a meeting participant. We observed that each of the note takers had a slightly different style. The first logger focused on capturing the technical content of the discussion and did not attach ownership to the ideas being discussed. During his turn (1 hour, 24 minutes), he created 38 timezones and wrote 7 keywords but did not apply them to the log. In addition, this user had illegible handwriting which made it difficult to read his part of the log.

The second note-taker began many of her 36 timezones with the name of the speaker. She summarized the contents of the discussion with short phrases. She also created 9 keywords and applied them 39 times as she took notes. Unlike the other two note takers, she made occasional references to other actions occurring during the meeting, such as noting when the group changed pages on the electronic white board. She logged for 53 minutes.

The third performed the task of note taking in much the same way as the second, but did not include the identification of speakers. Her notes focused on the main ideas of the discussion, and in some places they resembled an outline. Many of her 39 timezones were divided into numbered and indented sequences, giving her part of the log a visual framework. She created 9 keywords and applied them 50 times during her 1 hour and 11 minute session.

Using the Log for Retrieval

At this time, two people have reviewed parts of the recorded brainstorming meeting. In both review sessions the user had a 22 page hard copy version of the log as well as the electronic version. Both retrievers, had been present during the recording; however, neither of them had operated Marquee during the meeting. One was a primary participant in the meeting while the other was an observer.

The first retriever used the Marquee log to browse the contents of the videotape. He spent most of his reviewing session browsing from point to point on the tape by circling timecodes in the log. He found that having the hardcopy log supported his browsing by allowing him to look ahead to timezones in which keywords had been attached. He also augmented the log during his review, adding new notations to some of the timezones.

The second retriever stated that he reviewed the tape specifically "looking for notes about action items that might have been brought up." He told us that he did not find the notes in the log useful to him, as they did not describe events in which he was interested. He followed up this statement by speculating that "If I were trying to teach someone to use Marquee for the first time in a setting like a meeting...I would have used as examples for the keywords, process, and activity things like 'new topic, action item." This retriever did eventually find the parts of the meeting he was interested in reviewing and augmented the log with his own annotations to reflect the process notes he desired.

Discussion

The use of Marquee by the brainstorming meeting attendees demonstrates that logging can be accomplished in real time and that a Marquee log can indeed assist people in accessing information recorded on a videotape. Our observations raise issues for future design and study.

Participating and Taking Notes. We observed that although Marquee could be used in real time, none of the note takers spoke during the meeting. They became scribes of the activity rather than participating note takers. However, none of the loggers were principle members of the project team, and we were told later by several of the principles that they were hesitant to use the system because they felt it might impede their participation.

We need to explore further whether participants can take notes with Marquee while still participating, or if effective note taking requires a scribe. We believe that the ability to participate while using Marquee will be dependent on the desired detail of the log and on the user's understanding of how to log for the purposes of retrieval.

Individual Note-Taking Styles. We found that Marquee did support the individual styles of the loggers; however, the individual styles did not necessarily support other users of the log. We have not yet investigated how to support multiple users creating shared logs nor multiple users creating correlated indices. The fact that both of our retrievers desired different kinds of notes suggests that there be continued study into effective annotation to support retrieval.

Changing the Behavior of Note Taking. Marquee influences the approach that users take towards real-time documentation because Marquee is intended to support retrieval. In fact, one of the retrievers stated that, "[If I were using Marquee] I would not try to capture [all ideas] in a mad scramble. I would put enough down to be able to find it on the tape....In other words, the whole note taking would change if we had a tape to work from."

In designing Marquee we carefully chose to support user's prior note taking methods; however, our initial observations have revealed that these methods do not necessarily support retrieval. As people begin to utilize tools like Marquee, their methods for indexing and annotating will change to accommodate navigation into recorded material. We foresee further research into understanding the differences between writing paper notes for recall and note taking to support the replay of videotaped events.

The Need for Overviews. During the two retrieval sessions, users utilized the hardcopy logs as overviews in order to locate potential points of interest. This observation has led us to explore ways of automatically generating overviews for the contents of a log, so that users can quickly gain a sense of the material on a tape. For instance, we could use the keyword labels applied to timezones to create a table of contents for a recording.

Logging as an Iterative Process. Both of our retrievers added notes to the log when reviewing the meeting tape. We believe that logging is an iterative process and that allowing at least the creation of a skeleton in real time will benefit the retriever. We found that the skeleton was useful, and the system supported users adding notes into the log during recording and during post-capture review. Additional work needs to be done to investigate how users will want to embellish existing annotations with new ones. For example, we envi-

sion a system that tracks changes made by each individual user, resulting in a way of selectively viewing a particular user's interpretations.

SUMMARY

In designing and building Marquee we believed that users needed a system for generating a rich set of annotations during the recording of a videotape. We took a user-centered approach, first creating simple Hypercard prototypes to explore different interface designs, then using paper mockups to test our early assumptions, and finally conducting informal evaluations of the working system itself. The result of our investigations is a pen-based system which provides users with a means for correlating their personal notes and keywords with recorded footage. Our observations point to further work in understanding the task of logging in real time. They suggest that Marquee is a powerful tool for creating a log in real time, which when combined with other indexing methods, provides a robust retrieval tool.

ACKNOWLEDGMENTS

The authors wish to acknowledge Frank Halasz, Tom Moran and Sara Bly for their continued support and contributions to the project. We particularly thank all the people who participated in our user studies and system trials. Thanks to Francoise Brun-Cottan, Steve Harrison, Gordon Kurtenbach, Cathy Marshall and Patricia Sheehan for their critical reading of this paper.

REFERENCES

1. Aguerre Smith, T. and Davenport, G. The Stratification System A Design Environment for Random Access Video. *Proceedings of the 3rd International Workshop on Network and Operating System Support for Digital Audio and Video*, San Diego, California. November 1992.

2. Davis, M.E. Director's Workshop: Semantic Video Logging with Intelligent Icons. AAAI-91 Intelligent Multimedia Interfaces Workshop Position Paper 1991.

3. Gould, J.D. and Salaun, J. Behavioral Experiments on Handmarkings. *CHI+GI 1987 Conference Proceedings*. April 5-9, 1987 Toronto, Canada.

4. Harrison, B. *The VANNA System: Video ANNotation and Analysis*, (a videotape). University of Toronto. 1991.

5. Kurtenbach, G. and Buxton, W. Issues in Combining Marking and Direct Manipulation Techniques. *UIST 1991 Conference Proceedings*. November 11-13, 1991, Hilton Head South Carolina.

6. Lamming, M.G. and Newman, W.N. Activity-based Information Retrieval: Technology Support of Personal Memory. Personal Computers and Intelligent Systems, F.H. Vogt editor. Information Processing 92, Vol. III. Elsevier Science Publishers B.V., North Holland. 1992

7. Mackay, W. and Davenport, G. Virtual Video Editing in Interactive Multimedia Applications. *Communications of the ACM*, Vol. 32, No. 7. 1989

8. Minneman, S. and Harrison, S. Where Were We: making and using near-synchronous, pre-narrative video. *Proceedings First ACM International Conference on Multimedia*. 1-6 August 1993, Anaheim, California.

9. O'Connor, B. Selecting Key Frames of Moving Image Documents: A Digital Environment for Analysis and Navigation. *Microcomputers for Information Management*, Vol. 8, No. 2.June 1991.

10. O'Shaughnessy, D. Speaker Recognition. *IEEE ASSP Magazine*, October 1986 p. 4-17

11. Otsuji, K. and Tonomura, Y. Projection Detecting Filter for Video Cut Detection. *Proceedings First ACM International Conference on Multimedia*. 1-6 August 1993, Anaheim, California.

12. Siu, M.H., Yu, G., Gish, H. An Unsupervised, Sequential Learning Algorithm For The Segmentation of Speech Waveforms With Multiple Speakers. *Proceeding of International Conference on Acoustics, Speech and Signal Processing*. Vol. 2. March 23-26, 1992, San Francisco, California.

13. Trigg, R.H. Computer Support for Transcribing Recorded Activity. *SIGCHI Bulletin*, Vol.21, No. 2, October 1989.

14. Tonomura, Akutsu, S., Otsuji, K., Sadakata, T.VideoMAP and VideospaceIcon: Tools for Anatomizing Video Content. *INTERCHI'93 Conference Proceedings*. 24-29, April 1993.Amsterdam, The Netherlands.

15. Turner, J. Representing and Accessing Information in the Stockshot Database at the National Film Board of Canada. *The Canadian Journal of Information Science*, Vol. 15, No. 4, December 1990.

16. Ueda, H., Miyatake, T., Sumino, S., and Nagasaka, A. Automatic Structure Visualization for Video Editing. *INTERCHI'93 Conference Proceedings*. 24-29, April 1993.Amsterdam, The Netherlands.

17. Weber, K. and Bly, S. A. Annotating Video Records With Multiple Media. Palo Alto Research Center Technical Report, ISTL-CSA-92-12-01, 1992

18. Wilcox, L.D., Smith, I. and Bush, M.A. Wordspotting for Voice Editing and Audio Indexing. *CHI '92 Conference Proceedings*. May, 1992. Monterey, California.

19. Wolf, C. Can People Use Gesture Commands? *SIGCHI Bulletin*, Vol. 18, No. 2, October 1986.