Evaluating the Contributions of Video Representation for a Life Oral History Collection*

Michael G. Christel
CS Department and HCI Institute
Carnegie Mellon University
Pittsburgh, PA 15213
1-412-268-7799

christel@cs.cmu.edu

Michael H. Frisch
Departments of History and American Studies
University at Buffalo, SUNY
Buffalo, NY 14260
1-716-639-1047

mfrisch@buffalo.edu

ABSTRACT

A digital video library of over 900 hours of video and 18000 stories from The HistoryMakers is used to investigate the role of motion video for users of recorded life oral histories. Stories in the library are presented in one of two ways in two within-subjects experiments: either as audio accompanied by a single still photographic image per story, or as the same audio within a motion video of the interviewee speaking. Twenty-four participants given a treasure-hunt fact-finding task, i.e., very directed search, showed no significant preference for either the still or video treatment, and no difference in task performance. Fourteen participants in a second study worked on an exploratory task in the same within-subjects experimental framework, and showed a significant preference for video. For exploratory work, video has a positive effect on user satisfaction. Implications for use of video in collecting and accessing recorded life oral histories, in student assignments and more generally, are discussed, along with reflections on long term user studies to complement the ones presented here.

Categories and Subject Descriptors

H.3.7: Digital Libraries – *user issues*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *evaluation, video*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Experimentation, Human Factors

Keywords

Oral histories, video retrieval, exploratory search

1. INTRODUCTION

The rapid decline of costs for digital storage and network bandwidth make storing and sharing digital video libraries much more viable than a decade ago. The terabyte of storage for 1000 hours of video for an early 1990s Informedia video library cost $1 million, with that terabyte now under $1000. The three orders of magnitude savings enabled the explosive growth of video on the internet, as evidenced by user-generated video hosting sites like YouTube. In November 2007, more than 75% of U.S. internet

users watched a streaming or progressive video download. Americans viewed nearly 9.5 billion online videos that month, 2.9 billion of which occurred at YouTube [12]. The dramatic reduction in cost and growth in network capacity allow cultural assets like oral histories to be processed, indexed, and shared in new ways beyond text-only representations. Audio interview recordings can be digitized, and for videotaped oral history collections the video can be digitized as well. This paper explores the role of video in oral history collections: should it be captured and represented, and if so, what benefits might encountering interviews in that format offer users?

Oral history has long been recognized as offering a valuable method for recording and studying memories, with extensive life histories representing a particularly valuable dimension. Modern developments in communication have made it close to indispensable for historians because of limitations in traditional written sources. R. Young discusses oral history in the chemical industry [25], as excerpted in an ICOM newsletter [11]:

> In their attempts to reconstruct the past, historians have traditionally relied on personal correspondence, diaries, journals, newspapers, and government documents as primary sources. …Technological innovations since the late 19th century changed the nature of communication and the content of written records. As a result, as historians find traditional written sources insufficient for their research, they turn increasingly to oral history. ...Scientific, technical, and business careers are most often measured in terms of the ...experimental results and technological innovations preserved for posterity in journals and books. But these documents record only the public face. The rich history of the everyday life of the chemical sciences and technologies – the social networks, patterns of patronage, and the "messy vitality" of the laboratory, library, and production plant – is thus rarely fleshed out in the documentary record.

But the significance of oral history extends far beyond research as such. The Sound Archive of The British Library summarizes its importance on an "Oral History" collection web page [17]:

> In recent years oral history has emerged as a powerful means of recording and preserving the unique memories and life experiences of people whose stories might otherwise have been lost…. Oral history has a firm place in schools as an interactive and shared experience offering children a rare chance to question history face-to-face while bringing generations together.

Section 2 discusses oral history representation and overviews The HistoryMakers archive. Sections 3 and 4 present an empirical study using this archive with a directed search task and report results. Section 5 describes a second study

investigating the same subject pool and archive but with an exploratory task, with Section 6 reporting results. Concluding sections discuss the findings of the two experiments with respect to video representation in oral history collections.

One main innovation of this work is spotlighting the pitfalls of digital library search experiments that exclusively study only lookup. Lookup tasks are the emphasis of many NIST TREC benchmarking activities, including TRECVID [16], but search includes learning and investigating activities as well. This paper provides empirical evidence that examining only lookup tasks presents an incomplete picture into interface utility for library search. If only lookup were studied, the paper would end prematurely after Section 4, but by extending the empirical investigations to cover exploratory tasks as well, interesting insights are reached with respect to the subtle interplay between task and multimedia stimuli. This work provides a more thorough assessment of video representation for oral histories, while underscoring the importance of investigating the full range of library search activities.

2. BACKGROUND AND RELATED WORK

2.1 Evaluating the Effectiveness of Oral Histories over Text Transcripts

In general, interviews provide autobiographical accounts that supplement material already available in the written record. They are a primary source for interpretive works such as published books, magazine articles, Web modules, and educational materials [11, 19]. Until quite recently, it has generally been the case that any serious work with oral histories required that recordings be transcribed to produce usable text versions of the content. Many libraries and archives require transcription as a basis for a collection, and oral history has often been defined formally as a transcript of a carefully edited interview with a person of historical significance [1, 19].

This preference is not based on the claim that a transcript is in any sense a truer or better representation that the actual recording. Indeed, the limitations of transcriptions are widely recognized— generally lost or exceedingly difficult to represent are all the dimensions of voice—accent, intonation, emphasis, emotion etc, not to mention the facial expression, gestures, and body language that are visible in video recordings of interviews [4, 6, 22]. Nevertheless, transcription has seemed inevitable—even "natural"—because in the past it was impossible to work easily with large and intractable collections of recorded data, especially in analog form. Analog recordings, most often tape-based, are extremely tedious to work with in terms of locating and reserving the proper playback equipment, setting up the tape, and searching for relevant material [1, 5].

Digital technologies have dramatically altered this situation in recent years, opening up many possibilities for exploring and searching audio and video data directly. In effect, these technologies are restoring as the primary source for oral history: the actual oral and in some cases the visual dimension of the interview recordings [4].

These developments permit us to pose new questions about the differential effects of encountering oral history as text or rather as audio-visual documents. Major projects, in developing new technologies for processing and interacting with audio and video documentation, have begun engaging these questions, focusing particularly on access to the primary media documentation, in contrast to transcript representations [8, 15, 18, 21]. How does the medium of presentation affect what users find in oral histories, their reflections on the collection, and their overall experience in accessing and interacting with the oral history library?

Gustman et al. discuss the many challenges to processing and accessing a very large archive of videotaped oral histories from thousands of survivors, liberators, rescuers, and witnesses of the Nazi Holocaust [7]. They note the presumed value of video representation over text transcript representation for Survivors of the Shoah Visual History Foundation (VHF):

> Text summaries do not convey the emotional content of the testimonies to the user. For students, the main end-users for the VHF archive, the text summaries cannot substitute for the full video because it is the emotional content of the testimonies that vividly shows the horrors of intolerance and thus gets messages of tolerance across to students. For the same reason it is questionable whether [text] summaries provide a sufficient basis for selection of testimonies for use in the classroom.

Such claims seem reasonable, but in fact they have not been tested carefully. They are generally not supported with published experiments exploring user response, thus, leaving open the value of videotaped representations for oral history consumers [6, 18].

This paper directly addresses this question through two experiments with users of a large corpus of indexed video life oral history interviews. We ask what the effect is, for users of this collection, in having access to the video dimension of these recordings, not just to the audio and the transcript.

2.2 The HistoryMakers Archive

The HistoryMakers, established in 1999, is a non-profit institution headquartered in Chicago whose purpose is to record, preserve and disseminate the content of video oral history interviews highlighting the accomplishments of individual African Americans and African-American-led groups and movements [9]. Its aim is to provide a unique scholarly and educational resource for exploring African American history and culture. It is unique among collections of African American heritage because of its large and varied scope, with interviewees from across the United States, from a variety of fields, and with memories stretching from the 1890s to the present. Rather than focus on one particular part of a person's life, such as their career or participation in the civil rights movement, the interviews are life oral histories covering the person's entire span of memories as well as their own family's oral history passed down from times of slavery, Reconstruction, and the late 19th-early 20th century.

Collaboration between The HistoryMakers and the Carnegie Mellon University (CMU) Informedia research group made use of speech alignment, image processing, and language understanding technologies to promote multiple levels of access and fuel the viewing of the actual video recordings in a large oral history corpus [3]. Specifically, The HistoryMakers provided 913 hours of videotaped interviews with 400 prominent African American individuals conducted primarily from 2001-2005. Archivists and staff hired by The HistoryMakers also provided human-generated transcripts, story boundaries, and reviewed and corrected the key frame per story (a representative image for the story extracted from the video). The resulting HistoryMakers Digital Library, augmented with Informedia processing to provide additional

metadata such as time-aligned transcripts at the spoken word level of granularity [3], has 18254 interview story segments. The HistoryMakers Digital Library is a sufficiently sized corpus on which to conduct research into representations and user reactions to digital delivery of oral history libraries.

2.3 Summer 2007 HistoryMakers Workshop

In mid-2007, the HistoryMakers Digital Library became ready for user testing. A consortium of initial institutional testers was developed, with the corpus and database installed at the History Makers office in Chicago, the University of Illinois, the Schomburg Center for Research in Black Culture of the New York Public Library, and, in Buffalo, both the State University of New York at Buffalo (UB), and Randforce Associates, in the UB Technology Incubator.

A workshop was organized at UB in the summer of 2007, bringing together corpus administrators and staff from The HistoryMakers and CMU Informedia; the Schomburg Center; History and African American Studies faculty from the University at Buffalo and the University of Illinois (with participants from the Chicago, Springfield, and Urbana-Champaign campuses); and Randforce Associates from the UB Technology Incubator.

At this workshop, participants were trained in the uses of the digital video library interface, and designed a range of explorations and experiments to be developed at each site, some by expert users and others, in classroom settings, involving students first encountering the corpus and database. These discussions refined a set of survey instruments to generate user demographics and satisfaction data, as well as a set of exploratory historical focus topics to be examined in comparable investigations by different researchers at the several sites. For novice student users encountering the system, a set of inquiries was developed, with particular attention to exploring the similarities and difference between narrower specific search experiences, and broader thematic exploration.

These latter rubrics were used to design the experiments detailed in follow-up sections below, where the specific inquiry is termed "treasure hunt," and the broader approach termed "open ended exploration." Our research investigates these questions:

- What if the oral history narrative is visually rendered without motion video, but with a still photographic image taken of the speaker? Will the corpus retain its value across representative tasks if stories are presented as still imagery with audio versus full motion video?

- Does the impact of video change based on task?

- For faculty assigning students to explore oral history collections, what are the implications of having access to the video representation as well as audio and text?

3. USER STUDY: NOVICES WITH TREASURE HUNT TASK

3.1 Participants

Study participants were recruited through the "Experiment Scheduling Site" web page provided by the Center for Behavioral Decision Research at Carnegie Mellon University. This page attracts subjects from the Pittsburgh community within walking distance of the University of Pittsburgh and Carnegie Mellon University, predominantly but not exclusively college students. The 24 subjects who participated in this study had no prior experience with the interface or data under study and no connection with the research group.

They began the study by completing a demographics questionnaire designed as an artifact of the UB Workshop (Section 2.3). The subjects were 8 female and 16 male with a mean age of 23 (2 subjects less than 20, 4 in 30-35 age range). All subjects had at least a high school education, with 17 having at least some college education and 9 having some graduate school education. The subjects' reported race/ethnicities using U.S. Census 2000 Form wording were 10 "Asian," 9 "White," 2 mixed or other, 1 "Black, African American, or Negro," 1 "Hispanic or Latino," 1 "Pacific Islander."

The participants were generally not familiar with African American oral histories. On a 5-point scale responding to "How familiar are you with African American oral histories and/or The HistoryMakers?" (1=*Not at all*, 5=*Very much*), most indicated little familiarity (distribution for 1-5 were 11-7-6-0-0). The participants were experienced web searchers but inexperienced digital video searchers. For "Do you search any web/online information systems?" (1=*Not at all*, 5=*Very frequently (several times daily)*), the answer distribution was 1-1-3-8-11 while for "Do you use any digital video retrieval system (video stored and searchable on a computer)?" with the same scale, the distribution was 8-9-5-2-0. The latter two questions have been asked in numerous digital video library studies conducted by the lead author, e.g., see [2] with similar participant groups, and since 2002 the web search experience has grown from modest to frequent and video search experience from not at all to modest. It is anticipated that as video search sites like YouTube maintain popularity that encourages numerous imitating sites [12], the college student of the future will have even greater experience in online video searching. Each subject spent about 60 minutes in the study and received $10 for participation, with cash prizes to the top 3 scores in the treasure hunt task to motivate performance.

3.2 Video Retrieval Systems under Study

The video library interface used for the study was a streamlined version of the one used with NIST TRECVID 2006 search tasks [16] that achieved the best-ranked interactive search performance at that time. The HistoryMakers Digital Library interface was streamlined in its emphasis on text search, and removal of shot-based visual concept search mechanisms that are useful for TRECVID topics relevant for broadcast news (e.g., find all buildings and roads), but do not apply for an oral history corpus where most shots are head-and-shoulder shots of the person being interviewed.

The two systems for our investigation differed only in the behavior of the story playback window, shown in Figure 1. The two treatments are as follows:

- *Video*: whenever the user selects "Show story" the story is played with full motion video synchronized to the audio; e.g., in Figure 1, Ernie Banks would be seen talking.

- *Still*: whenever the user selects "Show story" the story is played with the same audio as in the *Video* treatment, but the image stays fixed, showing just the single representative image for the story. Representative images, the same images used as thumbnails elsewhere in the interface but at lower resolution when represented as thumbnails, went through a human inspection process by The HistoryMakers archive developers to verify good

representative qualities like the speaker's eyes being open. The representative image for the story "Ernie Banks describes his personality…" is shown in Figure 1.

Figure 1. Story playback window: for "Video" treatment, story plays back as video; for "Still", the image stays fixed as shown (same audio for both).

With both systems under investigation, the user could issue text queries, image queries, and browse the story sets returned from queries. All other aspects of the interfaces are identical, except for what appears in the upper area for the story playback window: one treatment has video, the other just a still image.

3.3 Procedure

Participants worked individually with an Intel® Pentium® 4 class machine, a 1280 x 1024 pixel 18-inch color monitor, and headphones in a Carnegie Mellon computer lab. Participants' keystrokes and mouse actions were logged within the retrieval system during the session. They first signed a consent form and filled out an online demographic questionnaire.

Participants were given a one page instruction sheet stating the following:

Treasure Hunt, SYSTEM 1

One story best answers or discusses each of the 12 topics below. There is no order to these 12 – find as many of the stories as you can in your 20 minutes, and capture them into the Collector pane shown at the right. Capture a story quickly by clicking the story playback window and typing *F1* while over the playback area, or by typing *F1* while over the thumbnail image. You do not need to use any other search tool besides text search, and remember that:

coffee cup   as a query will return all stories mentioning coffee or cup or both

(coffee cup)   as a query will return only stories mentioning both coffee and cup

"coffee cup"   as a query will return only stories mentioning the phrase "coffee cup"

Pilot testing before the study, along with feedback from the UB workshop, showed that the text search syntax for and-query and adjacency-query was atypical and often forgotten, so the example was given in paper form. Additional feedback from student users in late 2007 has led to the decision to replace the underlying text search service for The HistoryMakers system with the Lemur Toolkit [13], a more mature and capable search engine, but this replacement was not ready in time for the studies reported here.

The one-page instruction sheet for System 2 starts the same way, but with "SYSTEM 2" in the title. The bottom of each sheet lists the twelve topics for discovery, given in Tables 1 and 2.

Table 1. Topics for treasure hunt study, SYSTEM 1.

Detroit Edison law case
Photo of a musician quartet in Singapore
Strong texture of Japanese hair
African language and offering opinion that U.S. schools should teach more than one language
Malaria caught by a family member in the American South
Dance of God of Life and Death in Puerto Rico
Making All State in a sport the year after breaking leg
Escape of a mother and daughter from Nigeria via neighboring country Benin
DuBois and Kent State in reflecting on all Kent State had to offer for studying philosophy and religion
Photo shaking hands with Richard M. Nixon
Holds a nickname "Pit Bull"
Low payment for earning a patent

The systems were counter-balanced in this within-subjects design: half the subjects experienced *Video* first as SYSTEM 1, and then *Still* as SYSTEM 2, while the other half saw *Still* first and then *Video*. The user spent exactly 20 minutes with each system addressing the listed twelve topics for the treasure hunt, followed by a questionnaire. The post-system questionnaire consisted of a text comment input area for typing open-ended comments, and six five-point scales authored by the participants of the earlier UB workshop:

- Unsatisfying – Satisfying
- Not Historically Meaningful – Historically Meaningful
- Not Inspiring – Inspiring
- Not Entertaining – Entertaining
- Inaccurate – Accurate
- Too little content – Too much content

A running timer showed the countdown to 20 minutes for the use of each system. A concluding questionnaire provided a final open-ended text commenting area, and asked the user 3 questions, which the answer picked from "SYSTEM 1", "SYSTEM 2", or "No difference":

- Which of the two systems did you find easier to learn?

- Which of the two systems did you find easier to use?

- Which of the two systems did you like best overall?

With time for questionnaires, each participant took about an hour for the study.

Table 2. Topics for treasure hunt study, SYSTEM 2.

Kerosene ball for baseball
Photo of uniformed military personnel with Secretary of Defense Melvin Laird
Dignity from James Edwards to Sidney Poitier to Denzel Washington
Remembering scents of childhood holidays including turkey, sweet potato pies, and the ocean
Discussing conversation with Bishop Tutu about Nelson Mandela held in California
Meeting Bobby Kennedy in a South American country
Jimmy Carter's family stays in his Colorado home
Citing a "history of the world…is of groups" DuBois quote
Being hired onto a television show staff to make the show "psychologically believable"
Photo showing three with same middle name Proctor
Naming three areas in the American South that were mostly Catholic
Arabic welcome in album along with Louis Armstrong's "Wonderful World"

In order to focus subjects on getting to answers quickly, text search was emphasized as the only search mechanism necessary, and the interface was tuned by labeling the keyboard to more easily support the following information-seeking procedure:

a) Issue text search.

b) Browse stories returned for text search, identifying which if any to play.

c) Play a story.

d) If story answers a topic, add it to the shown answer set; if not, iterate back to steps 1, 2, or 3; continue until time runs out or all given topics are answered.

Figure 2 illustrates this process. The user issues a query anticipated to address one of the given topics, e.g., "photo of a musician quartet in Singapore." A one-word query *music* is typed into the text input box at the upper left and the user sees too many matching stories in the created tab. He or she issues a new search on Singapore that returns ten stories in another tab. Stories are

represented as a thumbnail image along with a score thermometer, a relevance bar to the left of each thumbnail, with stories listed in relevance order. Relevance is determined by the underlying text search engine based on the common scoring metric of term frequency inverse document frequency.
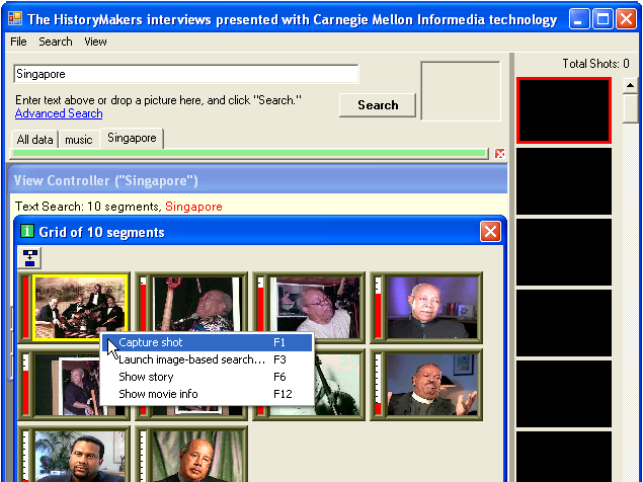

Figure 2. Oral history interface: issue text query (upper left), see results (lower left), interact with thumbnails to show stories, and capture answers that get posted to the Collector pane, initially shown as all-empty slots on the right.

A quick scan of the thumbnails locates an answer: 4 musicians shown, which the user could confirm took place in Singapore by showing and playing the story. Subjects used keyboards with 4 function keys labeled: *F1* to add the story to the answer set, *F3* to issue a color search, *F6* to show and play a story, and *F12* to show in a text explorer view the interviewee name and information about the interview date and location. These 4 operations are also selectable by right-clicking the mouse to bring up the context-sensitive menu as shown in Figure 2.
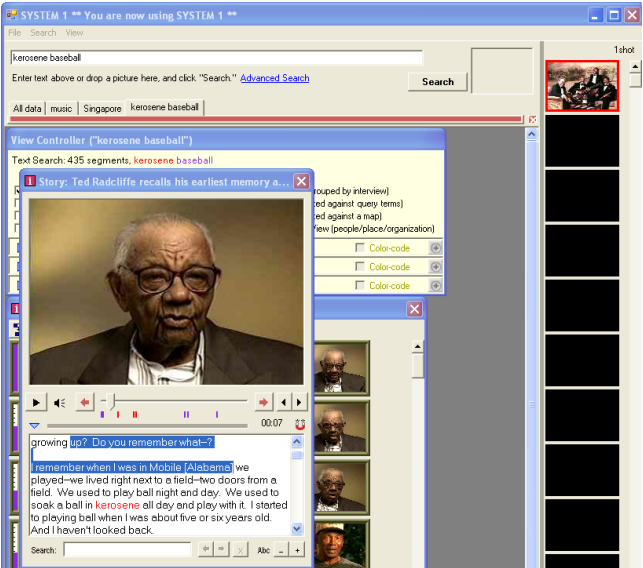

Figure 3. Clicking thumbnail or pressing F6 key leads to story playback window, which confirms story as answer to another topic. The answer is posted to rightmost pane via the F1 key from either the thumbnail or the story player.

Figure 3 illustrates the interaction when looking for the "kerosene ball for baseball" story, after posting the answer about the quartet in Singapore (shown in right pane). The appearance of the speaker in the story playback window depends on the system in use at the time, as discussed in Section 3.2. Through thumbnail and story review, subjects answer as many of the 12 topics as possible in their 20 minutes with SYSTEM 1, answer a questionnaire, and then repeat the process for SYSTEM 2.

4. RESULTS FROM TREASURE HUNT

There was no significant difference in effectiveness in the within-subjects study. When using the *Still* treatment, subjects found an average of 5.0 topics out of 12, standard deviation (std) 2.9, minimum of 0 and maximum of 12. When using the *Video* treatment, subjects found an average of 5.6 topics out of 12, std 3.5, minimum of 0 and maximum of 10. Overall, subjects found a minimum of 2 and maximum of 21 of the 24 topics during their two 20 minute timed sessions. 13 subjects found at least half of the total topics, i.e., at least 12 topics, during the timed sessions. Targeting just these subjects in a post-hoc analysis also showed no significant difference between the *Still* and *Video* treatment on the directed search, treasure hunt task. With *Still*, the top-scoring 13 subjects found an average of 7.8 topics of 12 (std 2.7) and with *Video*, they found an average of 6.9 topics (std 2.4). The difference in the story-showing interface (see Figure 1) did not result in any significant differences in topic-finding performance, at the 0.05 or even 0.1 levels of significance.

Results from the 5-point scales completed after each 20-minute timed session are shown in Table 3. Of the 5 scales, only the Inaccurate-Accurate scale approached being a significant difference with $p < 0.1$. Subjects when using *Video* rated the system as being more accurate (3.59) than when using the *Still* interface (3.18), F=3.32 with 1 df, $p = 0.08$.

Table 3. Results of post-treatment questionnaires.

5-Point Scale (left = 1, right = 5)	Still Average	Video Average
Unsatisfying – Satisfying	3.50	3.45
Not Historically Meaningful – Historically Meaningful	3.82	3.77
Not Inspiring – Inspiring	3.45	3.41
Not Entertaining – Entertaining	3.27	3.27
Inaccurate – Accurate	3.18	3.59
Too little content – Too much content	3.55	3.32

For the final questionnaire, there was no difference on satisfaction when asking subjects directly which system they favor, SYSTEM 1 or SYSTEM 2. Many times, subjects expressed no opinion, answering the question with "No difference." For which system is easier to learn, subjects answered with 11 no-difference, 7 for the system that was *Video*, and 6 for the system that was *Still*. For which system is easier to use, of the 16 times an opinion was given, subjects rated *Video* as the easier system to use 9 times. For which system the subjects liked best, of the 16 times an opinion was given, subjects rated *Video* as the system liked best 9 times. The only strong pattern in the data was that in the experiment, subjects rated SYSTEM 2 as being easier to learn than SYSTEM 1 for 9 of the 13 stated opinions. As subjects

gained familiarity with text search, thumbnails, keyboard shortcuts, and other interface options they felt more comfortable with the system in general and rated the second system encountered as easier to learn. This preference for SYSTEM 2 did not carry into the easier-to-use or like-best questions though: 8 of 16 opinions on easier-to-use were for SYSTEM 2 and 9 of 16 opinions for which system is liked best. The ending survey did not reveal a preference for either *Still* or *Video* given this directed search task.

The open-ended comments support the questionnaire data. Many users did not know what the difference was between SYSTEM 1 and SYSTEM 2. The second-best performer entered the comment "Was there a difference? I couldn't tell." A few users did spot the difference and commented on it, e.g., the top-scoring performer typed "The video aspect of System 1 brings the viewer into the story more effectively than the pictorial aspect of System 2." However, such recognition and reflection was the exception. A follow-up study should ask subjects to state the difference between SYSTEM 1 and SYSTEM 2. From the collected comments and questionnaire data, we would presume that more than half the subjects would not know the answer. Subjects guessed that there was no difference at all, or incorrectly guessed that the difference was in speed or the text search service or the content.

Automatically collected transaction logs recorded the queries and time spent playing stories using the interface of Figure 1. For the 20-minute session, subjects on average submitted 20.6 text queries with *Still* and 21.5 text queries with *Video*. They took the paper instructions' advice and stayed with text queries as the exclusive directed search mechanism: image search was used only 5 of 1015 searches. The text searches were imprecise, with the maximum number of stories per search, set at 1000, often returned due to the default or'ing of terms and the subjects not focusing the search further through and'ing or adjacency search. Subjects on average with *Still* returned 382 stories per text query, while subjects with *Video* returned on average 432 stories per text query. On average, subjects with *Still* used the story player interface for 536 of their 1200 seconds, or 44.6% of the time. Subjects with *Video* used the story player interface for 562 seconds, or 46.8% of the time. On average, subjects with *Still* submitted 8.0 stories into the Collector for consideration as answers, while with *Video* they submitted 8.4.

None of these measures is significantly different statistically between *Still* and *Video*, i.e., regardless of treatment, subjects used the systems in the same way for the directed search. That way was in spending about half the time thinking of and issuing text queries, at a rate of roughly 1.5 to 2 per topic, a small bit of time posting results, and the remaining 45% or so reviewing stories with the playback interface. The open-ended comments inform us more about that playback interface use, though. Subjects felt the need to work extremely quickly to cover all topics in the treasure hunt, and focused on the text transcript prominently displayed at the bottom of the playback window, and not the visual playback area above the transcript. Some subjects were so focused on the transcript that they failed to notice that there was a difference in the visual playback area presentation between SYSTEM 1 and SYSTEM 2.

The nature of the task, directed search under time constraints, is analogous to directed search on stated topics under ambitious time constraints used in NIST TRECVID video retrieval evaluations

[16]. However, this emphasis on the need for great efficiency at the expense of appreciating the oral history corpus and reflecting on its stories may have suppressed any advantages that *Still* or *Video* have for oral history presentation. For directed search, there is very weak support from this study that *Video* offers a subjective benefit (subjects rate accuracy higher), with performance metrics, log data, and the remainder of the subjective questionnaires returning no significant differences. However, what if the task were more in line with History professor assignments demanding not simple fact-finding but essay-building and more exploratory work? From the UB workshop, this latter type of task was most discussed by the History and African American Studies faculty, and two such tasks for use with The HistoryMakers Digital Library form the basis for the study reported in the next section.

5. USER STUDY: NOVICES WITH EXPORATORY OPEN-ENDED TASK

5.1 Participants

Study participants were recruited with the same web page as discussed in Section 3.1, attracting the same user community. The 14 subjects who participated in this study had no prior experience with the interface or data under study and no connection with the research group (and hence did not take part in the study discussed in Section 3). The subjects were 7 female and 7 male with a mean age of 25.5 (3 subjects less than 20, 2 older than 30). All subjects had at least a high school education, with 13 having at least some college experience and 6 having some graduate school education. The subjects' reported race/ethnicities were 9 White, 3 Asian, 1 Black, and 1 Pacific Islander. The participants were generally not familiar with African American oral histories. On a 5-point scale responding to "How familiar are you with African American oral histories and/or The HistoryMakers?" (1=*not at all*, 5=*Very much*), most indicated little familiarity (distribution for 1-5 were 5-6-2-1-0). The participants were experienced web searchers but inexperienced digital video searchers. For "Do you search any web/online information systems?" (1=*not at all*, 5=*Very frequently (several times daily)*), the answer distribution was 0-0-2-5-7 while for "Do you use any digital video retrieval system (video stored and searchable on a computer)?" with the same scale, the distribution was 5-3-5-0-1. Each subject spent about 60 minutes in the study and received $10 for participation. No additional cash prize was offered to motivate performance.

5.2 Video Retrieval Systems under Study

The systems under study are exactly the same as used in the study reported in Section 3. There are two systems, *Video* and *Still,* with their only difference being in the visual display of the top portion of the story playback window, exactly as presented in Section 3.2. The systems were counter-balanced in this within-subjects design: half the subjects experienced *Video* first as SYSTEM 1, and then *Still* as SYSTEM 2, while the other half saw *Still* first and then *Video*.

5.3 Procedure

Participants worked individually with an Intel® Pentium® 4 class machine, a 1280 x 1024 pixel 18-inch color monitor, and headphones in a Carnegie Mellon computer lab. Participants' keystrokes and mouse actions were logged within the retrieval system during the session. They first signed a consent form and filled out an online demographic questionnaire. They were given a one page instruction sheet stating the following:

Task Overview

Many stories are directly or peripherally relevant to each of the topics below. Assume you have to create a report addressing the stated topic. What stories might you reference in your report? Find relevant stories for the given topic in your 20 minutes, and capture them into the Collector pane shown at the right. Capture a story quickly by clicking the story playback window and typing F1 while over the playback area, or by typing F1 while over the thumbnail image. You do not need to use any other search tool besides text search, and remember that… (then, same coffee cup example as reported in Section 3.3, so text search emphasized as with the study reported in Section 3).

The topics were listed on the one-page instructions after the brief text preface "Capture stories that you find interesting and worth reporting for the general topic:" with the SYSTEM 1 topic being:

Identify characteristics that resulted in the leadership effectiveness of the interviewee (the speaker in the interview). What traits helped to make this person an effective leader?

The SYSTEM 2 topic was:

Discuss the civic, social, or political organizations that the interviewee (the speaker in the interview) founded. How did his or her ideology change over time?

6. EXPLORATORY TASK RESULTS

Accuracy and correctness in performance were not evaluated, in part because of the difficulty in evaluating the Collector contents (e.g., is one extremely relevant story better than three somewhat relevant ones or two contrasting ones?), and in part because the task was set up as but one step along the way toward producing a report, but the report was not generated within the 20 minute limit. Changing the task to produce an artifact like a report that would be easier to grade is a consideration for future longitudinal studies, but here the goal was to focus on *Still* vs. *Video* and keep the time-on-task at 20 minutes, allowing a discussion across both the user study of Section 3 and Section 5 with their 20-minute sessions.

Results from the 5-point scales completed after each 20-minute timed session are shown in Table 4. Of the 5 scales, only the Inaccurate-Accurate scale showed a significant difference. Subjects when using *Video* rated the system as being more accurate (3.36) than when using the *Still* interface (2.71), F=4.94 with 1 df, $p < 0.05$.

For the final questionnaire, there was an overwhelming difference on satisfaction when asking subjects directly which system they favor, SYSTEM 1 or SYSTEM 2. For which system is easier to learn, subjects answered with 9 no-difference, 4 for the system that was *Video*, and 1 for the system that was *Still*. For which system is easier to use, of the 5 times an opinion was given, subjects rated *Video* as the easier system to use 4 times (same pattern as with "easier to learn" question, but distributed across a different set of 5 subjects). For which system the subjects liked best, of the 8 times an opinion was given, subjects rated *Video* as the system liked best 7 times. Across all questions, the preference

was clearly for *Video* over *Still*. Even the pattern seen with the results in Section 4, that subjects will rate SYSTEM 2, the second encountered, as easier to learn than SYSTEM 1, did not hold: of those expressing opinions, 3 rated SYSTEM 1 as easier to learn, 2 picked SYSTEM 2. It was not the timing of systems that mattered, it was their nature, i.e., whether *Video* or *Still* was presented. Clearly, the subjects strongly preferred *Video* given this exploratory task.

Table 4. Results of post-treatment questionnaires.

5-Point Scale (left = 1, right = 5)	Still Average	Video Average
Unsatisfying – Satisfying	3.00	3.14
Not Historically Meaningful – Historically Meaningful	3.64	3.5
Not Inspiring – Inspiring	3.57	3.64
Not Entertaining – Entertaining	3.00	3.14
Inaccurate – Accurate	2.71	3.36
Too little content – Too much content	3.36	3.5

The 23 contributed open-ended comments support the questionnaire data. While 14 of these comments dealt with text search or the given topic, 9 discussed the SYSTEM 1 and SYSTEM 2 differences. Three expressed frustration with the still-image and audio presentation, e.g., "it was a little awkward just looking at their pictures and only hearing them speak." Four praised the video presentation, e.g., "it's more engaging to see the person speak than to just hear them." One indicated that the user did not care between still-image and video, and one "couldn't tell any difference between the systems." It is much more likely that with the exploratory task, subjects recognized the difference between systems, whereas with the focused directed search task they did not. If subjects had to state the difference between SYSTEM 1 and SYSTEM 2, at least 8 and perhaps as many as 13 would have been able to identify the difference.

Automatically collected transaction logs show that for the 20-minute session, subjects on average submitted 4.6 text queries with *Still* and 4.8 text queries with *Video*. They took the paper instructions' advice and stayed with text queries as the exclusive directed search mechanism: image search was used as 7 of the 140 performed searches. The text searches were broad, with the maximum number of stories per search, set at 1000, often returned due to the default or'ing of terms and the subjects not focusing the search further through and'ing or adjacency search. Subjects on average with *Still* returned 552 stories per text query, while subjects with *Video* returned on average 497 stories per text query. On average, subjects with *Still* used the story player interface for 799 of their 1200 seconds, or 66.6% of the time. Subjects with *Video* used the story player interface for 814 seconds, or 67.8% of the time. On average, subjects with *Still* submitted 11.2 stories into the Collector, while with *Video* they submitted 10.6.

None of these measures is significantly different statistically between *Still* and *Video*, i.e., regardless of treatment, subjects used the systems in the same way for the exploratory search – spending much less time thinking of and issuing text queries, at a rate of roughly one every four minutes. They spent a small bit of time posting results, and the majority of time, two-thirds of the session, reviewing stories with the playback interface shown in Figure 1.

The nature of the task, exploratory search under relaxed time and performance constraints, is in agreement with how a number of African American Studies and History professors use oral histories in their classes: as assignments leaving room for student exploration instead of more directed lookups of facts or episodes. For such broad thematic exploration, represented as the topics used with SYSTEM 1 and SYSTEM 2, the subjects played many more stories, and noticed and reacted to the visual representation, i.e., the interface portion above the transcript in Figure 1.

7. DISCUSSION

Dependent on task, the visuals in the story playback area are either unacknowledged with no effects, or strongly preferred as motion video rather than still image form. The two studied tasks were the treasure hunt directed search and an open-ended exploratory search. An exploratory search "may be characterized by the presence of some search technology and information objects that are inherently meaningful to users …often motivated by a complex information problem, and a poor understanding of terminology and information space structure" [23]. Marchionini breaks down three kinds of search activities: lookup, learn, and investigate, noting exploratory search as especially pertinent to the learn and investigate activities [14]. The first study, the treasure hunt task, was solely in the "lookup" category, as are the tasks in many traditional information retrieval forums studying fact retrieval or question answering. The second study required learning and investigating activities, which Marchionini further breaks down as follows [14]:

- Learn: knowledge acquisition, comprehension/interpretation, comparison, aggregation/integration, socialization

- Investigate: accretion, analysis, exclusion/negation, synthesis, evaluation, discovery, planning/forecasting, transformation

Not surprisingly, the teaching faculty and library science professionals at the UB Workshop were quite interested in the use of The HistoryMakers Digital Library for more than just traditional fact retrieval and lookup, but also for these learning and investigating activities. Hence, two studies were conducted, and user actions differed based on the task. For treasure hunt lookup, significantly more text searches were issued in the 20-minute sessions than with exploratory search, regardless of interface treatment. With exploratory search, significantly more time is spent playing stories regardless of format (still or video) versus the treasure hunt task.

In the lookup task, subjects were driven to find facts, being ignorant of the oral history visual presentation differences. A prior eye-tracking empirical study looked at digital video surrogates comprised of text and three thumbnail images to represent each document, and found that participants looked at and fixated on text far more than pictures [10]. They used the text as an anchor from which to make judgments about the list of results. The authors conclude that perhaps the task may have affected performance, but no evidence for such an influence was found [10]. In the pair of studies reported here, we do have a body of evidence suggesting different foci of interest within the story playback interface of Figure 1. For fact lookup, subjects

focused on the lower half of the interface, the text transcript, while for the exploratory task, they acknowledged and expressed a strong preference for the upper half being motion video. Given a targeted search task, text representation dominates user attention. For exploratory search tasks, participants learning and investigating through a greater frequency of story playback express a strong preference for motion video representations of the oral history interviews.

For the first impressions study here, the 20-minute time limit was restrictive, especially for the exploratory task. So that users would not be overwhelmed with interface choices, text search was emphasized, but the resulting subjective scores in Table 4 are low, with the low-rated satisfaction supplemented by open-ended text comments asking for more ways to browse the oral histories and more access strategies than just text search. In fact, the full HistoryMakers Digital Library through Informedia processing does offer many additional exploratory interfaces such as map and timeline views, visualization-by-example VIBE plots, query previews based on interviewee and subject headings, as discussed elsewhere [3]. As noted above, these two studies were conducted with a streamlined interface to focus on the *Still* versus *Video* question for first-time users of the corpus, but even the short term study here suggests that for exploratory search, users want more options. They made use of image search for 5% of their issued searches, for example, as opposed to image search being used for only 0.5% of the searches in the treasure hunt task.

Marchionini notes that searches supporting investigation involve multiple iterations that may take place over long periods of time [14]. Shneiderman and Plaisant acknowledge that evaluating tools for exploratory, creative work is difficult, and call out the need for longitudinal research with "Multi-dimensional In-depth Long-term Case-studies (MILC)" [20]. Ideally, MILC research could be conducted with the campuses, faculty, and students involved in The HistoryMakers testing, i.e., those participants from the UB Workshop, to see changing patterns of use and utility as the faculty and students gain familiarity and experience with the system. Such research is a logical extension to the "first impressions" research reported here. In the term "Multi-dimensional In-depth Long-term Case studies" the multi-dimensional aspect refers to using observations, interviews, surveys, as well as automated logging to assess user performance and interface efficacy and utility. The in-depth aspect is the intense engagement of the researchers with the real users to the point of becoming a partner or assistant. Long-term refers to longitudinal studies that begin with training in use of a specific tool through proficient usage that leads to strategy changes for the expert users. Case studies refer to the detailed reporting about a small number of individuals working on their own problems, in their normal environment.

Somewhat surprisingly, both studies confirm that when the stories are presented as motion video, the subjects rate the system as being more accurate. This conclusion merits further investigation as here we only research impressions after a 20 minute session. If subjects work with a system for days or a whole semester, will they still rate interfaces with video representation as more accurate? And by "accuracy," are users referring to the historical content as such, or rather to the way video provides a more accurate sense of the personality and reliability of the observed interviewee? Longer term studies informed by these and similar

questions are being designed as ways to build from the results reported here.

In discussing the Australia National Library oral history program, Barry York comments that oral history is important in providing a "richer texture to the fabric of our history and heritage" [24]. Faculties introducing their students to The HistoryMakers Digital Library are advised against giving only lookup tasks, if a goal of the assignment is to appreciate the "texture" of the oral history corpus. With a fact-finding assignment, students will likely be extremely focused on text transcripts and not acknowledge or be moved by the oral or visual contents of the interviews. The same advice holds for other oral history corpora like VHF [7]: if students are given a fact-finding task against Shoah materials and they have text transcripts to work with, they will fixate on them.

With exploratory tasks, however, the students make use of and appreciate the motion video rendering. Given the low costs of storing and transferring video, and the increasing awareness of and experience with digital video by students as noted above, oral history archivists are encouraged to digitize and make available not only the oral representation but also the motion video representation when source videotapes are in hand. For oral history collections now being gathered, the corpus administrators are advised on the benefits of capturing not only audio recordings but motion video as well. These benefits will not show themselves in first time use of the corpus when restricted to fact retrieval. However, with exploratory search, video is much preferred for rendering the interviews and leads to an increased perception of accuracy. Exploratory search tools help information seekers get beyond finding to understanding and use of information resources [14], and as oral history archivists are likely to be quite motivated to have their materials better appreciated, understood and used, video plays a key representational role.

8. CONCLUSION
This study suggests that for exploratory inquiry especially, access to the video dimension of interview data produces enhanced user connection to and interest in first-time encounters with oral histories. The empirical and experimental findings reported here will be followed up in subsequent studies. But considering that video data is coming into wider use in oral history more generally, the experiments reported here have useful implications for designing the interface between data and user provided in various systems. The two experiments suggest the importance of providing structural and interface design support for diverse uses: fast, efficient means to get right at information, but also ways to appreciate fuller context and an easy way to move between detailed and exploratory modes of engaging oral histories documentation. Story player features such as the capacity to chain through stories, to jump to specific points, to scan transcript text for facts, yet also to play full linear story as motion video, serve a broad range of lookup, learning, and investigating search activities. Longitudinal work with consumers of oral histories, e.g., through MILC research with UB Workshop participants, can identify user and contextual profiles for streamlining the interface to better support a varied range of tasks.

Text will always have a role in archival work with oral histories. But clearly, audio and especially video—the primary sources in which oral history content and meaning are recorded—are moving to the forefront of user methodology. Engaging oral history recorded media directly—not mediated through text transcription—is clearly central to the future of the field. If oral

history recordings are truly to become an accessible and usable resource, the task of research is now to understand that process more deeply, and to design systems that can support it as meaningfully and usefully as possible.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] Baum, W. *Transcribing and Editing Oral History*. Nashville, Amer. Assoc. for State and Local History, 1977.

[2] Christel, M., and Conescu, R. Addressing the Challenge of Visual Information Access from Digital Image and Video Libraries. In *Proc. JCDL* (Denver, CO, 2005), 69-78.

[3] Christel, M., Richardson, J., and Wactlar, H. Facilitating Access to Large Digital Oral History Archives through Informedia Technologies. In *Proc. JCDL* (Chapel Hill, NC, 2006), 194-195.

[4] Frisch, M. Oral History and the Digital Revolution. In Perks, R. and Thomson, A., eds., *The Oral History Reader, Second Edition* (London, Routledge, 2006), 102-114.

[5] Frisch, M. Three Dimensions and More: Oral History Beyond the Paradoxes of Method. In Hesse-Biber, S., and Leavy, P., eds., *The Handbook of Emergent Methods* (New York, Guilford Press, 2008).

[6] Good, F. Voice, Ear and Text: Words, Meaning, and Transcription. In Perks, R. and Thomson, A., eds., *The Oral History Reader, Second Edition* (London, Routledge, 2006), 362-373.

[7] Gustman, S., Soergel, D., Oard, D., Byrne, W., Picheny, M., Ramabhadran, B., and Greenberg, D. Supporting Access to Large Digital Oral History Archives. In *Proc. JCDL* (Portland, OR, 2002), 18-27.

[8] Henson, P.M., and Schorzman, T.A. Videohistory: Focusing on the American Past. *Journal of American History* 78 (1991), 618-627.

[9] HistoryMakers African American oral history archive, *The HistoryMakers*, http://www.thehistorymakers.com/.

[10] Hughes, A., Wilkens, T., Wildemuth, B., & Marchionini, G. Text or pictures? An eyetracking study of how people view digital video surrogates. In *Proc. Conf. Image and Video Retrieval (CIVR)* (Urbana-Champaign, IL, 2003), 271-280.

[11] International Council of Museums Committee for Conservation Working Group "Theory and History of Conservation-Restoration" *newsletter 12* (April 2006), http://icom-cc.icom.museum/Documents/WorkingGroup/TheoryHistory/ICOM-CCWG_on_THC_newsletter12.pdf.

[12] Jaques, R. US surfers flock to web video. *Computing.co.uk*, 21 Jan. 2008, www.computing.co.uk/vnunet/news/2207678/three-quarters-surfers-watch.

[13] Lemur Toolkit for Language Modeling and Information Retrieval, http://www.lemurproject.org/.

[14] Marchionini, G. Exploratory Search: From Finding to Understanding. *CACM* 49(4), 2006, 41-46.

[15] Moore, K. Perversion of the Word: The Role of the Transcript in Oral History. *Words and Silences: Journal of the International Oral History Association* 1 (1997).

[16] NIST, Digital Video Retrieval at. *TREC Video Retrieval Evaluation*, http://www-nlpir.nist.gov/projects/trecvid/.

[17] Perks, R., curator. *Oral History*, Sound Archive, The British Library, //www.bl.uk/collections/sound-archive/history.html.

[18] Progler, J.A. Choices in Editing Oral History: The Distillation of Dr. Hiller. *Oral History Review* 19 (1991), 1-16.

[19] Ritchie, D. *Doing Oral History: A Practical Guide, Second Edition.* New York, Oxford Univ. Press, 2003.

[20] Shneiderman, B., and Plaisant, C. Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. In *Proc. BELIV'06 Workshop, Advanced Visual Interfaces Conf.* (Venice, May 2006), 1-7.

[21] Sipe, D. The Future of Oral History and Moving Images. In Perks, R. and Thomson, A., eds., *The Oral History Reader, Second Edition* (London, Routledge, 2006).

[22] Tedlock, D. *The Spoken Word and the Work of Interpretation*. Philadelphia, Univ. Penn. Publications, 1983.

[23] White, R.W., Kules, B., Drucker, S.M., schraefel, m.c. Supporting Exploratory Search, Introduction. *CACM* 49(4), 2006, 36-39.

[24] York, B. Affirming Humanity: Why Oral History Matters. *National Library of Australia News* (May 1997), 3-5, http://www.nla.gov.au/events/yorknews.html.

[25] Young, R. Oral History at CHF. *Chemical Heritage* 23(2), Summer 2005.