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

Mike Christel  
christel@cs.cmu.edu  
School of Computer Science  
Carnegie Mellon University

Michael H. Frisch  
mfrisch@buffalo.edu  
Depts. Of History and American Studies  
University at Buffalo, SUNY

Joint Conference on Digital Libraries  
June 18, 2008
Talk Outline

• Details on *The HistoryMakers, Inc.*
• Introduction to the oral history corpus
• Is video representation useful for such a corpus?
• Two studies detailed and reviewed:
  • Fact-finding “treasure hunt”
  • Exploratory search
• Discussion of Results
• Implications
Reasons to Pay Attention

- Explosive growth of video as a digital asset
- 3 orders of magnitude drop in costs of video storage over past decade
- (Over)emphasis on lookup tasks in evaluating digital libraries
- Implications regarding what students will gain from cultural assets based on what assignments are given

So, something for everyone: media specialists, archivists, information architects, information retrieval experts, evaluators, librarians, professors, humanities experts, computer scientists, ....
The HistoryMakers Oral History Archive

- Funded in part by IMLS Grant LG-03-03-0048-03
- World’s largest African American oral history archive with accomplished African Americans (www.thehistorymakers.com)
- Purpose:
  - To educate and show the breadth and depth of this important American history as told by the first person
  - To highlight the accomplishments of individual African Americans across a variety of disciplines
  - To preserve this material for generations to come
- Test corpus: 913 hours of interviews from 400 individuals, 18,254 interview story segments (mostly 2001-2005)
- Summer 2007 workshop hosted by SUNY at Buffalo with HistoryMakers beta testers motivated and informed this work
I have found out that poor blacks and poor whites, even though they may be poor, do not believe in the same things. Poor whites have...
Video Representation Needed for Oral History?

• Oral history traditional definition: “transcript of a carefully edited interview with a person of historical significance”

• Transcription seemed natural and inevitable due to past tedium in dealing with analog recorded data despite loss of fidelity:
  • Voice: accent, intonation, emphasis, emotion, etc.
  • Image: facial expression, gestures, body language

• Today’s digital technologies allow audio and video data to be searched and explored directly (e.g., Survivors of the Shoah Visual History Foundation).

• What is the effect, for users of a large corpus of life oral history interviews, to have access to the video dimension, not just the audio and the transcripts?
Study 1: Treasure Hunt Fact-Finding Task

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.

Topics

- 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

...
The federal district court was finding Detroit Edison [Company, DTE] guilty of racial discrimination. And I fined them monetarily.
Participants, Treasure Hunt Task

- 24 participants, recruited via web site form
  - Primarily college students
  - 16 male, 8 female
  - Mean age 23 (2 < 20, 4 in 30-35 age range)
  - 10 Asian, 9 White, 2 Mixed/Other, 1 Black, 1 Hispanic, 1 Pacific Islander
  - Not familiar with African American oral histories
  - Experienced web searchers
  - Inexperienced digital video searchers
- Paid participation, with cash prizes to top 3 scorers to motivate performance
Treasure Hunt Procedure

- Use “System 1” for exactly 20 minutes to answer 12 stated topics given on one-pager
- Answer questionnaire, rest as needed, then proceed to “System 2”
- Use “System 2” for exactly 20 minutes to answer 12 different stated topics given on one-pager
- Answer questionnaire for System 2, then final questionnaire

Half of the subjects saw “Still” (no video) as System 1, half saw “Video” as System 1, with System 2 then showing the other treatment.
Treasure Hunt Topics for System 2

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.

Topics

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

...
Results

• No significant difference on effectiveness:

<table>
<thead>
<tr>
<th>Report of Correct Topics (out of 12)</th>
<th>Still</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>5.0</td>
<td>5.6</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

• For 13 high-scorers, no sig. difference on effectiveness:

<table>
<thead>
<tr>
<th>Report of Correct Topics (out of 12)</th>
<th>Still</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>7.8</td>
<td>6.9</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.7</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Results, Post-Treatment Questionnaires

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

1 3 5

Still  Video

Carnegie Mellon
Results, Final Questionnaire

Easier to learn?  Easier to use?  Like best?

- No diff.
- System 2
- System 1
Results, End Questionnaire Continued

- General ignorance of difference between System 1 and System 2, e.g., “Was there a difference? I couldn’t tell.”
- Subjects guessed no difference at all, or a difference in speed or text search service or content.
Subjects used Systems in Same Manner

(Averages, per 20 minute session)

<table>
<thead>
<tr>
<th></th>
<th>Still</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text queries issued</td>
<td>20.6</td>
<td>21.5</td>
</tr>
<tr>
<td>Returned stories per text query</td>
<td>382</td>
<td>432</td>
</tr>
<tr>
<td>Time spent with story player</td>
<td>44.6%</td>
<td>46.8%</td>
</tr>
<tr>
<td>Submitted answers</td>
<td>8.0</td>
<td>8.4</td>
</tr>
</tbody>
</table>
Discussion, Treasure Hunt Task

- Directed search under time constraints similar to NIST TRECVID video retrieval evaluations
- Performance metrics, log data, subjective questionnaires show no significant differences
- Emphasis on need for great efficiency may have suppressed any differences between Still and Video

- What if task were more in line with work discussed at University at Buffalo HistoryMakers (UB) Workshop: not simple fact-finding, but essay building and more exploratory work?
Study 2: Exploratory Open-Ended Task

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? 

System 1 Topic

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?
Participants, Exploratory Task

- 14 participants, recruited via web site form
  - Primarily college students
  - 7 male, 7 female
  - Mean age 25.5 (3 < 20, 2 older than 30)
  - 9 White, 3 Asian, 1 Black, 1 Pacific Islander
  - Not familiar with African American oral histories
  - Experienced web searchers
  - Inexperienced digital video searchers
- Paid participation, with no additional cash prize as accuracy not measured for open-ended task
Results, Post-Treatment Questionnaires

- Too little content
- Inaccurate
- Not Entertaining
- Not Inspiring
- Not Historically Meaningful
- Unsatisfying

Too much content
Accurate
Entertaining
Inspiring
Historically Meaningful
Satisfying

1 3 5
Still Video
• Overwhelming preference for Video over Still whenever an opinion is given.
• 23 typed comments, 9 of which dealt with System 1 vs. System 2 differences: 3 disliked Still, 4 liked Video, 1 did not care and 1 could not tell the difference
Subjects used Systems in Same Manner

(Averages, per 20 minute session)

<table>
<thead>
<tr>
<th></th>
<th>Still</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text queries issued</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>Returned stories per text query</td>
<td>552</td>
<td>497</td>
</tr>
<tr>
<td>Time spent with story player</td>
<td>66.6%</td>
<td>67.8%</td>
</tr>
<tr>
<td>Submitted answers</td>
<td>11.2</td>
<td>10.6</td>
</tr>
</tbody>
</table>

* No significant difference in these measures

* Regardless of treatment, subjects spent 2/3 of time with player and noticed and reacted to the visual representation
Discussion, Exploratory Search

- G. Marchionini (“Exploratory Search: From Finding to Understanding,” CACM 49, April 2006) breaks down 3 types of search activities:
  - Lookup
  - Learn
  - Investigate
- Study 1 on Treasure Hunt was “lookup” with Study 2’s exploratory search pertinent to “learn” and “investigate”
- Computer scientists and information retrieval specialists emphasize evaluation of lookup activities (NIST TREC)
- UB Workshop library science and humanities participants quite interested in learn/investigate activities
Video Matters (Sometimes)

• For fact-finding tasks, subjects focused on text transcript, oblivious to the story playback area.

• For exploratory search, subjects learning and investigating through much story playback expressed a strong preference for motion video representation of oral history interviews.

• When stories presented as video, subjects rate system as being more “accurate”:
  • Will this trend hold beyond first impressions, e.g., for a semester-long study?
  • Are users referring to historical correctness or higher fidelity display of personality or something else?
Future Work

- Multi-dimensional In-depth Long-term Case-studies (MILC) (Shneiderman and Plaisant, 2006)
  - Investigative task may involve multiple iterations over long periods of time
  - Case studies let users work on their own problems
- Use MILC methodology within context of The HistoryMakers beta testing with engaged groups
- Improve overall satisfaction by providing more time and more tools
Conclusions

- Do not rely solely on lookup tasks to evaluate IR
- Do not rely on lookup to expose students to “texture” of the oral history corpus
- Text metadata (as in transcripts) provides a fast, efficient means to get directly to facts, with video synchronized to text letting user appreciate fuller context and move between detailed and exploratory modes of engaging oral histories.
- Oral history archivists are encouraged to digitize and make available video representations, preferred by users and leading to an increased sense of accuracy.
Many members of the Informedia Project, CMU research community, and *The HistoryMakers* contributed to this work; a partial list appears here:

**Informedia Project Director:** Howard Wactlar

**The HistoryMakers Executive Director:** Julieanna Richardson

**HistoryMakers Beta Testers:** Lillian Williams, SUNY at Buffalo and all UB Workshop participants: Schomburg Center for Research in Black Culture, NY Public Library, Randforce Associates, University of Illinois (Urbana Champaign, Springfield, Chicago)

**Informedia User Interface:** Ron Conescu, Neema Moraveji

**Informedia Processing:** Alex Hauptmann, Dorbin Ng

**Informedia Library Essentials:** Bob Baron, Bryan Maher