The Influence of Context and Interactivity on Video Browsing[1]

Barbara M. Wildemuth, Terrell Russell, T. J. Ward, Gary Marchionini, & Sanghee Oh
Open Video Project, School of Information and Library Science
University of North Carolina at Chapel Hill

TREC VID 2005 Notebook Paper, Revised March 2006
UNC SILS Technical Report 2006-01

Structured Abstract

Approaches tested: *The goal of this study was to investigate the effects of providing context and interactivity in a retrieval system, supporting the browsing of search result sets. Thus, three systems were developed: (1) a basic system, modeled on the current results list provided by google video searching (runs UNC-BAS-1 and UNC-BAS-2); (2) a similar system, with the context of each shot provided by showing keyframes from the shots appearing just before and after the retrieved shot (runs UNC-CON-1 and UNC-CON-2); and (3) a system that builds on the previous system by offering several mechanisms of interactivity (runs UNC-INT-1 and UNC-INT-2).*

Comparative findings: *In terms of both performance and user perceptions, the Context+Interactive system was superior. While there were no differences in precision, recall was improved with this system, and users preferred it (based on several measures of user perceptions).*

Conclusions: *The effects of context on browsing search results were negligible, but should be explored further through re-examination of the definition and operationalization of the concept of context. Interactivity, in combination with context, had positive effects on browsing effectiveness; it was considered easy to use, even though it introduced more complexity into the interface.*

1 Introduction

One necessary part of any information retrieval process is the person's selection of items from a list of retrieved items. For example, if someone searches the Open Video collection, they may begin by entering a set of search terms or by selecting a portion of the collection to browse. Once that request is presented to the search engine, it responds with a list of videos corresponding to the query/selection. At this point, the user must browse through the list to select those that they would like to view or download.

This activity is explicitly included in a number of models of the searching process, e.g., the model proposed by Marchionini (1995). In that model, the search process consists of recognizing and accepting an information problem, defining and understanding the problem, choosing a search system, formulating a query, executing the search, examining the results, and extracting information from the results. The focus of this year's TREC VID effort by the Open Video team was on mechanisms for examining the list of retrieved results.

2 Background

Browsing is a term that is used to describe many types of behavior. The following section will place it in the broader context of information seeking behaviors, then will review the literature related

[1] This work is supported by National Science Foundation (NSF) Grant IIS 0099638.

to browsing of results from video collections. The context provided by the results display and the level of interactivity of the interface displaying the results may both affect the user's ability to effectively examine the results list and select videos of interest. Both of these effects are discussed here.

2.1 Browsing

While browsing may be as broad in scope as to encompass an entire information-seeking episode (as in the models proposed by, e.g., Rice, McCreadie, & Chang, 2001, and Marchionini, 1995), the concept also can be defined to focus on the behavior that occurs as a person is scanning the list of results retrieved from an information system (Allen, 1996). This type of browsing is often referred to as *scanning.* Marchionini (1995) defines it as "a perceptual recognition activity that compares sets of well-defined objects with an object that is clearly represented in the information seeker's mind" (p.111). Allen (1996) notes that this type of behavior is affected both by the information seeker's characteristics (such as the precision with which the information seeker mentally represents the target being sought) and characteristics of the information being represented by the system (such as the clarity of the representation of the objects being scanned).

Jörgensen (1999) points out the non-textual nature of images and their influence on the process of browsing image collections. Fidel (1997) argued that the information needs that motivate image searches are distributed along a continuum from a Data pole (where the image is sought as a data source) to the Object pole (where the image is sought for its value as an object), and that information needs closer to the Object pole are more likely to elicit browsing behaviors. Yeo and Yeung (1997) directly considered video browsing, identifying two different types: microscopic browsing (of individual shots or frames) and macroscopic browsing (of scenes or stories).

In general, as Yeo and Yeung (1997) point out, "the goal [of systems supporting video browsing] is to overcome the sequential and time-consuming process of viewing video" (p.49). Many of the systems developed by TRECVID participants, as well as other systems described in the literature, are progressing toward this goal, usually by using keyframes, organized in varying displays, to represent video objects (Smeaton, 2004). Within the Agile Views design framework (Geisler et al., 2001), representations of the video objects (i.e., the surrogates for objects in the collection) might be seen as previews of particular shots or overviews of scenes and stories (called visual summaries by Yeo and Yeung). We have also investigated the efficacy of a variety of such surrogates.

An early study (Wildemuth et al., 2002) examined the relative efficacy of slide shows with and without audio keywords, storyboards with/without audio keywords, and fast forwards. While no surrogate was universally judged "best," we were able to reject the slide show with textual keywords from further consideration. Because the fast forward surrogate was relatively novel, it was investigated further (Wildemuth et al., 2003). We compared four fast forward surrogate speeds, with sampling rates of 1:32, 1:64, 1:128, and 1:256 and concluded that the tradeoff between efficiency and performance was balanced at the 1:64 speed. Another issue that arose from our early studies was the way in which people might use poster frames in combination with brief textual descriptions of the videos. Through eye tracking, we compared people's use of two alternative overviews: one with the poster frame on the left and the text on the right, and one with the layout reversed (Hughes et al., 2003). Most participants used the text as an anchor from which to make their relevance judgments and the images as confirmatory evidence of their selections. A follow up study (Hughes, 2003) found that most participants used both types of surrogate in combination—hypothesizing relevance with one and using the other to confirm or refute their hypothesis. These studies parallel results from other researchers and demonstrate the need for a variety of surrogates to effectively support browsing of video collections.

2.2 Context

The current study also investigated the influence of *context* on browsing. Most discussions of context (e.g., Sonnenwald, 1998) are concerned with the context in which information seeking behavior is conducted. The current study is concerned with a different type of context: the context of a

particular item within a set of retrieved items (Kwasnik, 1992). In most retrieval systems, including video retrieval systems, the results retrieved are displayed in a list, preferably a list ranked by the relevance of each item to the query. For example, see the current "list" display on google video (http://video.google.com). However, there are two aspects of video that may make this practice less than fully effective. First, many systems (such as those designed for TREC VID) are based on individual shots, which are very short in duration and represent only a very small portion of the original video program. Second, the shots were intended to occur in a particular sequence, i.e., they are temporal in nature, individually, and were designed to be shown over time (e.g., Ngo, Pong, and Zhang, 2001, worked with "temporal slides" of video, and Chua and Ruan, 1995, included "sequencing" as one of the activities supported by their video retrieval and sequencing system). When these two characteristics of video objects are taken into account, it can be inferred that a list of shots ranked by their relevance to the query may not be optimally useful. Each shot in the list is disassociated from the shots that occurred just before and after it in the original full-length video program (i.e., it is being viewed and evaluated "out of context"). In systems designed to retrieve shots from news video through text searching of the sound track, it is highly likely that relevant shots occur just before and after the shots actually retrieved through a keyword search of the transcript. The loss of the context of each shot (i.e., the shots surrounding it in the original program) is likely to be detrimental to the effectiveness of the retrieval system overall. The existence and scale of such an effect was investigated in this study.

2.3 Interactivity

Interactivity has also been identified as a system characteristic that can potentially increase its acceptance. In the design of the current study, we took Sundar's (2004) advice in defining interactivity as an attribute of the technology, rather than as a perceptual variable (i.e., perceived interactivity). Our focus is on "interactivity-as-product," or the "set of technological features [that] allow users to interact with the interface or system itself," rather than with other people through a computer-mediated communication system (Stromer-Galley, 2004, p.391). By keeping the definition of interactivity focused on the system's characteristics, we were able to avoid conflating the independent variable (i.e., differences in interactivity) with dependent/outcome variables based on user perceptions (i.e., perceived ease of use, perceived usefulness, and flow).

Jaffe (1995) defines interactive systems as "computer-based information systems which allow users some measure of control over the content and/or sequence of presentation" (p.10). Steuer (1992) provides a similar definition of interactivity: "the extent to which users can participate in modifying the form and content of a mediated environment in real time" (p.84). The concept of user control as a key component of interactivity is a theme also discussed by McMillan (2002), Downes and McMillan (2000), and Craven et al. (2001). Interface features that put control in the hands of the users increase the interactivity of the system. In addition, the rapidity with which a system responds to a user action increases the interactivity of the system (Downes & McMillan, 2000). These aspects of control and rapid feedback are summarized in Marchionini's (1995) definition of interactivity as "the number and rate of choices and actions the user makes and takes during information seeking" (p.110). The system used in the current study operationalized interactivity as the real-time display of a magnified view of keyframes on mouseover of a thumbnail view.

3 Three Systems Supporting Browsing of Search Results

As described above, the goal of this study was to investigate the effects of context and interactivity on the browsing of search result sets. Thus, three systems were developed: (1) a basic system, with a simple results list display; (2) a similar system, with the context of each shot provided; and (3) a system that provides both context and interactivity. Each of these three systems is described in more detail below.

The MySQL full text search engine was used for all three systems evaluated in this study. Their default list of stopwords was accepted and the research team set the minimum word length at three characters. In computing a relevance score, MySQL takes into account the number of words in a

record, the number of unique words in that record, the total number of words in the collection, and the number of records that contain a particular word. The search results were ranked based on the relevance score computed by MySQL.

Each of the systems is a web application built on entirely open source software. They were built using the RubyOnRails framework. Rails is a Model-View-Controller (MVC) framework written in the ruby programming language. This framework allowed us to prototype our ideas quickly and get results onscreen soon after a brainstorming session. We employed a number of AJAX (Asynchronous – Javascript – and – XML) calls to make the interfaces more interactive and responsive. This was made much simpler with the built-in tools available within Rails. We used the MySQL database and Apache, running in FCGI mode, for our webserver. The system was stable, very quick, and handled multiple users very well.

3.1 The basic system (runs UNC-BAS-1 and UNC-BAS-2)

In order to provide a baseline, we developed a system (see Figure 1) in which the search results display included only the basic elements: a poster frame (i.e., keyframe) at the left and the full ASR text for the shot at the right. The poster frame was 215 pixels wide (displayed at a screen resolution of 1024x768).

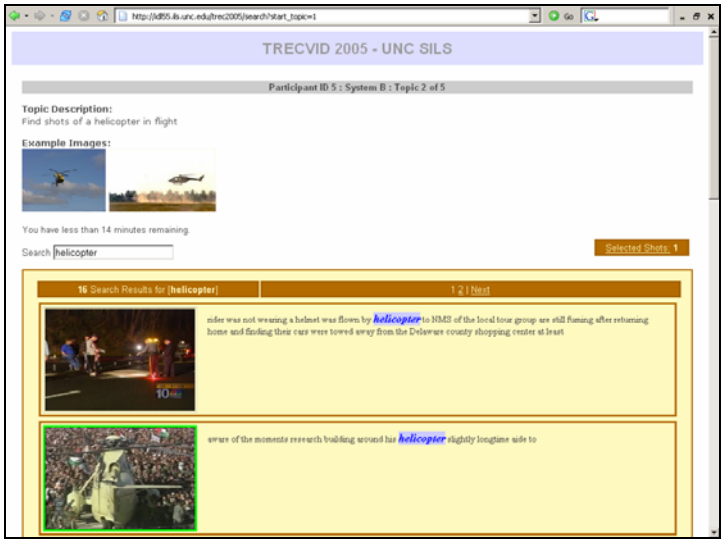

Figure 1. Results list from the basic system

3.2 The system with context (runs UNC-CON-1 and UNC-CON-2)

The sequence of shots from news stories is pertinent for browsing, because it is highly likely that the shots surrounding a shot retrieved based on relevant text are also relevant to the topic. One remaining design question is the size of the "window" surrounding the relevant shot, i.e., *how many* shots before and after a retrieved shot are also likely to be relevant? To resolve this question, a subset of topics selected from TREC VID 2003 were re-run, and the positions of any relevant shots within six shots in either direction (before or after) of the retrieved shot were recorded. It was found that, of the 249 relevant shots identified, 22% were the retrieved shot. If the three shots before the retrieved shot and the three shots after the retrieved shot were included in the "window," the user would view over 75% of the relevant shots. Thus, to provide context, we included three shots before each retrieved shot and three shots after each retrieved shot in the user's view of the results list (see Figure 2). Each of these keyframes was 106 pixels wide. The retrieved shot is bordered in light blue. In addition, the ASR text from the shots before and after each retrieved shot was displayed.

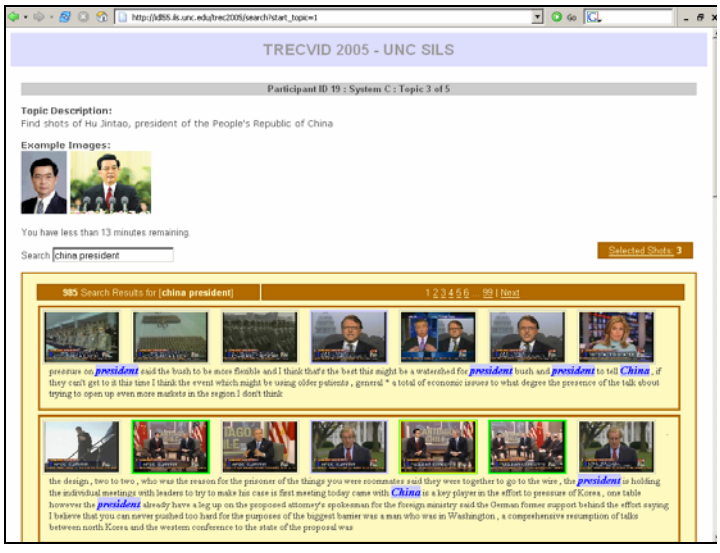

Figure 2. Results list with context

3.3 The system with context plus interactivity (runs UNC-INT-1 and UNC-INT-2)

The third system augmented the system with context by providing some interactive features. These features included: the highlighting (with a gold border) of each keyframe/shot on mouseover, and the display of each keyframe/shot at a larger size at the left, on mouseover. While these interactive features cannot be fully experienced through a single screen shot, this system's interface is shown in Figure 3. In this system, the large poster frame was 215 pixels wide; the smaller keyframes were each 79 pixels wide.

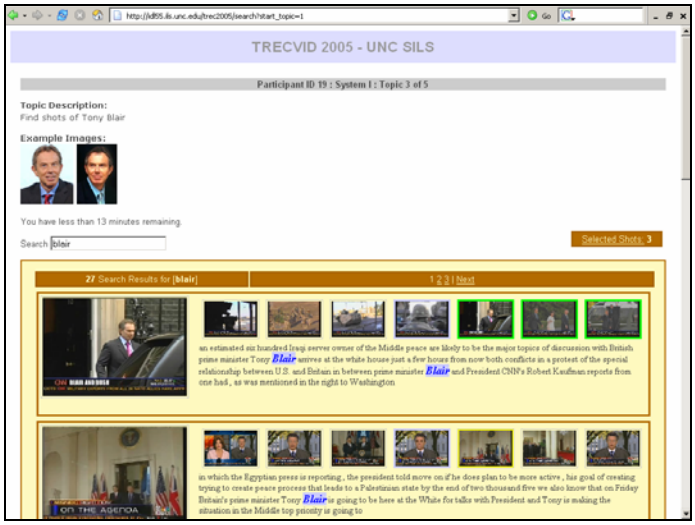

Figure 3. Results list with context and interactivity

4 Study Methods

4.1 Research design

The focus of the study is on the influence of context and interactivity on people's browsing of search results. A within-subjects research design was used to evaluate the three different search result displays in terms of precision, recall, usefulness, ease of use, and the experience of flow during interaction.

Each participant conducted one training search and four TREC topic searches on each of three different systems. The topics were organized into six blocks; the order of the blocks was counterbalanced, as follows:

Block 1: Topics 161 (people with banners or signs), 149 (Condoleeza Rice), 156 (tennis players), 166 (palm trees)
Block 2: Topics 160 (something on fire), 153 (Tony Blair), 165 (basketball players), 155 (map of Iraq)
Block 3: Topics 157 (people shaking hands), 152 (Hu Jintao), 171 (soccer goal being made), 172 (office setting)
Block 4: Topics 158 (helicopter in flight), 150 (Iyad Allawi), 159 (George W. Bush and vehicle), 164 (ship/boat)
Block 5: Topics 167 (airplane taking off), 151 (Omar Karami), 162 (people entering/leaving a building), 168 (road with cars)
Block 6: Topics 163 (meeting at table), 154 (Mahmoud Abbas), 169 (tanks or other military vehicles), 170 (tall building)

Each participant searched either Blocks 1, 2, and 3, or Blocks 4, 5, and 6. Each participant searched all three systems; the order in which the participants searched the systems was counterbalanced.

4.2 The video collection

Each participant searched a database that consisted of the English-language videos from both the development set and the test set of videos.[2] Thus, the database included approximately 74 hours of video from CNN, NBC, and MSNBC, plus several hours of NASA's Connect and/or Destination

[2] Both sets of videos were used in order to provide the participants with a more realistic searching context. The development set was not used for any training of the system prior to the study being conducted. Only the shots selected from the test set were submitted to NIST for evaluation, so only those shots are included in the calculations of mean average precision for the submitted runs. All other data reported here included both the shots selected from the test set and the shots selected from the development set by the study participants.

Tomorrow TV series'. The text included in the database was generated through automatic speech recognition and supplied by NIST.

4.3 The participants

The 38 participants were recruited via email to a variety of listservs in the School of Information and Library Science (SILS) and the School of Journalism and Mass Communication (SJMC), both at the University of North Carolina at Chapel Hill. In addition, participants were encouraged to invite their acquaintances to participate (i.e., a quasi-snowball recruiting scheme was used).

4.4 Study procedures

After giving informed consent, each person participated in an evaluation session, consisting of:

- a pre-session demographic questionnaire, asking for some basic information about the participant and his or her experience with video and video collections;
- 5 search topics on one of the systems (including a training topic), each followed by a very brief questionnaire, asking about the participant's familiarity with the topic and their satisfaction with the process of searching that topic;
- a questionnaire about that system, described below;
- 5 search topics on another system (including a training topic), each followed by a very brief questionnaire;
- a questionnaire about that system;
- 5 search topics on another system (including a training topic), each followed by a very brief questionnaire;
- a questionnaire about that system; and
- a brief post-session questionnaire, asking for the user's reactions to the system, for comparisons of the systems in terms of which was easiest to learn, which was easiest to use, and which the participant liked the best, and concluding with two open-ended questions, asking for the user's thoughts about the best and worst aspects of each system

The questionnaires that followed completion of all the searches on a particular system included three measures: a measure of the usefulness of the system (6 items; Davis, 1989), a measure of the ease with which the system could be used (6 items; Davis, 1989), and a measure of the participant's experience of flow during the interaction (8 items; Ghani, Supnick, & Rooney, 1991). The first two of these measures – perceived usefulness and perceived ease of use – are key components of the Technology Acceptance Model and have been used extensively to evaluate users' acceptance of a variety of technologies (Ma & Liu, 2004). A person's experience of flow has only recently been incorporated into studies of system use/acceptance, and consensus has not been reached on its definition and operationalization (Webster, Trevino, & Ryan, 1993; Finneran & Zhang, 2005). Nevertheless, improving a user's experience of flow during the searching process has been advocated as a design goal (Hearst et al., 2002). The current study incorporated Ghani et al.'s (1989) original measures of two aspects of flow – enjoyment and concentration.

The questionnaire responses were captured on paper. The user's interactions with the system, included which shots were selected as relevant, were captured in transaction logs.

The study procedures took approximately two hours to complete. Each participant received $20 in appreciation of their efforts.

4.5 Data analysis

Because NIST limits the number of runs that will be evaluated, the shots selected from the test set of videos by a few of the participants were submitted to NIST for relevance judging. The supplemental run included the shots submitted by the remainder of the study participants. The shots submitted from the development set of videos were evaluated for relevance by members of the Open Video team, following the same procedures as those used by NIST.

NIST calculated mean average precision (MAP) for each run submitted (two runs for each system). From the data about the relevance of each shot submitted, precision and recall were also calculated for each search conducted by each participant. Precision was calculated as the proportion of the submitted items that were judged to be relevant. Calculation of recall (Saracevic et al., 1988) was based on the assumption that the full set of relevant items in the collection is represented by the relevant items identified by NIST assessors plus any additional relevant items identified by Open Video assessors. While this approach to performance measurement is somewhat unusual within the context of TREC and other traditional information retrieval experiments, we believe that it is more able to take into account the variability in searcher performance, which can be as great as an order of magnitude (Borgman, 1989). The systems were then compared in terms of precision and recall, using analysis of variance. Bonferroni t tests were used for post hoc analysis of any differences found to be statistically significant.

Questionnaire data were also aggregated over the participants, for each system. Measures of usefulness, ease of use, and flow were compared across systems, using repeated measures analysis of variance. Post hoc t tests were conducted to analyze any differences that were found to be statistically significant. Data from the post-session questionnaire provided direct measures of user perceptions of the relative value of each system. Qualitative data from that questionnaire are included in the discussion of the results, to help us understand the quantitative results.

5 Results

5.1 Characteristics of the participants

The participants included 22 women and 16 men. Their average age was 30.3 years, and ranged from 18 to 58. None of the Open Video team members participated; 24 of the participants were affiliated with SILS but were not on the research team, and 14 had no relationship with the research team.

While the majority of participants were affiliated with SILS (n=24), several other disciplines were also represented. Two were affiliated with the School of Journalism and Mass Communication and two with the sociology department. Other departments represented, with one participant each, included allied health sciences, business, business and German, chemistry, communications and psychology, medicine, statistics, and surgery and environmental sciences. Two participants were not affiliated with any university department.

The participants averaged 9.3 years of experience with online searching (s.d.=3.4). All but one use a computer daily, and 35 of the 38 search daily. They are less experienced with using videos. Eleven of the participants watch videos less than once a week (24 weekly and 3 daily), and 21 never search for videos/films (13 occasionally, 3 monthly, and 1 weekly).

The participants varied in their familiarity with TV news. Seven of them do not watch any TV news while, at the other extreme, two of them watch TV news more than once each day. The remainder were distributed over weekly viewing (11), viewing more than once a week (11), and daily viewing (7). The participants rated their knowledge of current affairs as 3.7, on average (s.d.=0.9), where 1 indicates no knowledge and 5 indicates a great deal of knowledge.

5.2 User performance

As described above, the precision and recall data reported in Table 1 were calculated across all 38 users. In other words, the mean precision is the average of the overall precision achieved *by each person* on all the topics they searched. Recall calculations were based on the assumption that all the relevant shots were found by the aggregated efforts of all the TREC VID participants, including the Open Video team's assessment of shots from the development set.

Table 1. Summary of performance, by system

	Precision		Recall		Time per topic (in seconds)		Shots submitted	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
Basic	0.85	*0.24*	0.03	*0.04*	338.1	*194.0*	5.4	*6.2*
Context	0.78	*0.29*	0.04	*0.06*	333.8	*196.7*	10.8	*14.3*
Context+Interactive	0.79	*0.26*	0.05	*0.08*	358.0	*217.3*	11.1	*12.7*

Note: These data include all the search runs submitted to NIST, as well as all the user searches not included in the submitted runs. They also include shots selected from the development set, as well as the test set.

The differences across systems in precision were not statistically significant ($F=2.10$, $p=0.1237$). The differences across systems in recall were statistically significant ($F=5.08$, $p=0.0066$). Post hoc Analysis indicated that the Context+Interactive system supported better recall than the Basic system. The differences across systems in the amount of time spent searching were not statistically significant ($F=0.62$, $p=0.5407$). The differences across systems in the number of shots submitted were statistically significant ($F=11.33$, $p<0.0001$). Post hoc Analysis indicated that more shots were submitted with the Context and Context+Interactive systems than with the Basic system.

A subset of search results (six runs) was submitted to NIST to be evaluated in terms of mean average precision. Two runs represent each of the three systems. The results of those analyses, as reported by NIST, are shown in Table 2. Both runs for the Context system show higher average precision than the runs for the Basic system; both runs for the Context+Interactive show higher average precision than the runs for the Context-only system. Precision at 10 shots also shows the same trend; the results are more mixed for the other precision calculations.

Table 2. Results reported by NIST, aggregated by system

	Average precision	Precision at 10 shots	Precision at 30 shots	Precision at 100 shots	Precision at 1000 shots
Basic, run 1	0.030	4.2	5.1	5.1	5.1
Basic, run 2	0.020	3.0	3.3	3.3	3.3
Context, run 1	0.036	4.3	7.7	8.5	8.5
Context, run 2	0.032	3.9	6.2	6.2	6.2
Context+Interactive, run 1	0.039	4.5	8.3	9.7	9.7
Context+Interactive, run 2	0.046	4.7	7.0	7.3	7.3

5.3 User perceptions

In addition to the measures suggested by NIST, several measures of user perceptions were taken in relation to the three systems. These included questionnaires on perceived usefulness, perceived ease of use, and flow (enjoyment and concentration), as well as several other questions.

The questionnaire recommended by NIST in 2004 was used to collect data on user perceptions immediately after each search. The results from this questionnaire, aggregated by system, are shown in Table 3 (on next page).

The differences across systems were not statistically significant for the questions concerning participants' familiarity with the topic, the usefulness of the example images/videos, or the amount of time needed to find shots. The difference between systems in the ease of finding shots was statistically significant ($F=10.22$, $p<0.0001$). Post hoc Analysis indicated that both the Context and the Context+Interactive systems were more effective than the Basic system. Participants' satisfaction also varied by system ($F=4.59$, $p=0.0106$). Post hoc Analysis indicated that the participants were more satisfied with the Context+Interactive system than with the Basic system.

Table 3. User perceptions, based on post-search questionnaire (1, not at all, to 5, very much)

	Basic		Context		Context+Interactive	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
I was familiar with this topic before I did the search.	3.6	*1.3*	3.6	*1.4*	3.8	*1.4*
The example images/videos given with the topic description were useful for searching.	3.8	*1.1*	3.8	*1.3*	3.8	*1.3*
I found that it was easy to find shots that are relevant for this topic.	2.1	*1.2*	2.7	*1.4*	2.8	*1.4*
For this particular topic I was satisfied with the results of my search.	2.5	*1.4*	2.8	*1.4*	3.0	*1.5*
For this topic, I had enough time to find enough answer shots.	3.9	*1.4*	4.2	*1.1*	4.0	*1.3*

After completing the four assigned searches (plus one training search) for each system, each participant completed measures of usefulness (6 items), ease of use (6 items), and two dimensions of flow (4 items each). The results from these measures are shown in Table 4. (Note that lower scores indicate more positive attitudes.)

Table 4. User perceptions, based on post-system measures

	Basic		Context		Context+Interactive	
	Mean	*s.d.*	Mean	*s.d.*	Mean	*s.d.*
Perceived ease of use	2.9	*1.0*	3.0	*0.8*	2.6	*0.9*
Perceived usefulness	3.2	*1.1*	3.3	*0.9*	2.8	*1.0*
Flow (enjoyment)	4.4	*1.3*	4.3	*1.5*	3.6	*1.4*
Flow (concentration)	3.3	*1.1*	3.7	*1.1*	3.2	*1.1*

Note: Lower scores indicate more positive attitudes.

Differences in perceived ease of use, perceived usefulness, flow (enjoyment), and flow (concentration) were all statistically significant. Post hoc analysis of perceived ease of use ($F=5.01$, $p=0.0092$) indicated that the Context+Interactive system was perceived as easier to use than the other two systems. Post hoc analysis of perceived usefulness ($F=4.09$, $p=0.0208$) indicated that participants found the Context+Interative system more useful than the Context system. Post hoc analysis of flow (enjoyment) ($F=7.03$, $p=0.0016$) indicated that participants experienced more enjoyment when using the Context+Interactive system than when using either of the other systems. Post hoc analysis of flow (concentration) ($F=3.61$, $p=0.0321$) indicated that participants were able to concentrate more effectively when using the Context+Interactive system than when using the Context system.

After working with all three systems, the participants completed one additional questionnaire (a series of five-point scales; see Table 5). The participants found the length of the training session, and the systems' response time adequate. They found the systems easy to learn to use, and they understood the nature of the searching task. They were only lukewarm about the efficiency of searching using these systems. They found video searching to be somewhat different than other searching they perform. Finally, they perceived the systems as only moderately different from each other.

Table 5. Responses to final questionnaire (1=not at all, to 5,=very much)

	Mean	*s.d.*
Did you find that the length of the training session for the system(s) you used was sufficient?	4.4	*1.0*
Did you find that the system(s) response time was fast enough?	4.6	*0.9*
Learning how to use the system(s) was easy.	4.3	*0.8*
The system interface(s) allowed you to do the search task efficiently.	3.0	*1.2*
To what extent did you understand the nature of the searching task?	3.9	*1.0*
To what extent did you find this task similar to other searching tasks that you typically perform?	3.2	*1.1*
How different did you find the systems from one another?	3.3	*1.1*

The participants were then asked for direct comparisons of the three systems; their responses are shown in Table 6. The Context+Interative system was found easier to learn to use, easier to use, and was liked the best overall when directly compared with the other two systems.

Table 6. Comparison of systems in post-session questionnaire (number of participants giving each response)

	Basic	Context	Context+Interactive	No difference
Easier to learn to use	9	5	14	9
Easier to use	7	6	24	1
Liked the best overall	4	5	28	1

6 Discussion

The primary goal of the current study was to investigate the role of context and interactivity on users' browsing of the results list obtained by searching a video collection. Thirty-eight participants conducted four searches on each of three systems: a basic system, displaying each item in the list with a poster frame and the text accompanying the shot; a system with added context, displaying the poster frames and text for shots in the sequence before and after the retrieved shots; and a system with both context (as just described) and interactivity, operationalized as the highlighting and expansion of individual poster frames as the user rolled the cursor over them.

The overall finding is that the system providing both context and interactivity was superior to the other two systems, in terms of both performance and user reactions. While the Context+Interactive system was not necessarily superior to *both* of the other systems on *all* measures, *neither* of the others outperformed it on *any* measure.

There was no difference between systems in the precision achieved, but the Context+Interactive system was superior to the Basic system in terms of recall and mean average precision. In addition, study participants submitted more shots from the Context+Interactive system (and the Context system) than from the Basic system. From these results, we can infer the advantages of making more shots visible in the results list display. As users were able to view more shots more efficiently (because more shots were on the screen at one time), they also selected more shots and, thus, improved the recall they achieved with their searches.

Perceived usefulness is a consistently valid predictor of the acceptance of a system (Ma & Liu, 2004). The study participants perceived the Context+Interactive system to be more useful than the Context system; the Basic system fell in between and its usefulness was not reliably different than

either of the other two systems. The participants' open-ended comments on the final questionnaire provide some insight into this finding. They often commented that the small size of the keyframes on the Context system made it difficult to accurately assess the relevance of the shot. On both the Basic system and the Context+Interactive system, larger poster frames were available. Thus, the difference in usefulness appears to be closely associated with the size and visibility of the keyframes provided. The design tradeoff between size and number of keyframes on each page should be carefully considered in future systems. We suggest that the user-controlled expansion of individual keyframes is one way to trade off size and screen real estate.

Study participants consistently found the Context+Interactive system to be easier to use than the other systems. On the questionnaire following each search, they rated the Context+Interactive and Context systems higher than the Basic system, in terms of the ease with which shots could be found. The participants' perceptions of ease of use on the Davis (1989) questionnaire indicated that the Context+Interactive system was easier to use than either of the other two systems. When the three systems were directly compared on the final questionnaire, participants indicated that the Context+Interactive system was both easier to learn and easier to use than either of the other two systems. In some ways, this finding is surprising, since the interactivity incorporated into the interface of the Context+Interactive system made it more complex than either of the other two systems. We would conclude that the interactivity incorporated in this system was "natural" enough for users that they did not perceive it as difficult to incorporate into their repertoire of system use behaviors.

The Context+Interactive system also supported an experience of flow better than the other systems. The measure of enjoyment indicated that the Context+Interactive system was more enjoyable to use than either the Context system or the Basic system. The measure of concentration indicated that participants were able to concentrate on the task better with the Context+Interactive system than with the Context system (the Basic system was in between, and was not significantly different than either of the other two). As with perceived usefulness, the negative perceptions of the Context system may be attributable to the small images available in its interface.

Two questions were used to assess participants' affective responses to the three systems. On the questionnaire administered after each search, the participants indicated that they were more satisfied with the Context+Interactive system than with the Basic system (the Context system was in between, and was not significantly different than either of the other two). On the direct comparison after using all three systems, more participants said that they "liked" the Context+Interactive system than either of the other two systems. It is likely that the participants, all of whom are experienced online searchers, have come to expect the context and the interactivity incorporated in this system, and experienced disappointment when interacting with the Basic and Context systems.

A remaining issue is the potential value of providing context in a list of results from a search of a video collection. While the Context+Interactive system was consistently superior, the Context system was viewed more positively than the Basic system on only one measure: the single question about ease of finding shots on the post-search questionnaire. The performance results lead to a somewhat more positive conclusion about the Context system: it appears to outperform the Basic system on mean average precision and leads to more shots submitted than with the Basic system; but it is equivalent to the Basic system on the other performance measures. From these findings, we conclude that context is not as important as interactivity for improving the effectiveness of browsing through video results lists. Because context as operationalized in this study (i.e., by displaying sequential shots) may be unique to news video, this aspect of a display warrants further study.

7 Conclusion

An important step in the searching process is the examination of the results retrieved. In this step, the searcher browses through the results to make judgments about their relevance and to extract information from those found to be relevant. Because video is costly (in terms of time) to download, displays of results lists should be optimized to make the process of browsing more effective. The role of context, defined as the shots appearing just before and after the retrieved shot, and the role of

interactivity, defined as users' control over the display, were manipulated in the current study to investigate their effects on user performance and perceptions. It was found that the combination of context and interactivity positively affected both performance and user perceptions.

Future studies should focus on two things. First, they should examine the role of context through more varied operationalizations of that idea and with different genre of video. While the positive effects of context found in this study were minimal, it is possible that our system implementation did not maximize its effects or that its effects are more evident with other types of videos. Second, future studies should explore the limits of interactivity and its positive effects on user behaviors and perceptions. The interactivity incorporated in our system was minimal; we could easily imagine more interactivity than was possible for us to implement within the time constraints of a single TREC study. Other types of interactivity, and their effects on browsing effectiveness, should be examined.

8 References

Allen, B. L. (1996). *Information Tasks: Toward a User-Centered Approach to Information Systems.* San Diego: Academic Press.

Borgman, C. L. (1989). All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management, 25*(3), 237-251.

Chua, T.-S., & Ruan, L.-Q. (1995). A video retrieval and sequencing system. *ACM Transactions on Information Systems, 13*(4), 373-407.

Craven, M., et al. (2001). Exploiting interactivity, influence, space and time to explore non-linear drama in virtual worlds. *ACM SIGCHI 2001 Conference Proceedings,* 30-37.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Downes, E. J., & McMillan, S. J. (2000). Defining interactivity: A qualitative identification of key dimensions. *New Media & Society, 2*(2), 157-179.

Fidel, R. (1997). The image retrieval task: Implications for the design and evaluation of image databases. *New Review of Hypermedia and Multimedia, 3,* 181-199.

Finneran, C. M., & Zhang, P. (2005). Flow in computer-mediated environments: Promises and challenges. *Communications of the Association for Information Systems, 15,* 82-101.

Geisler, G., Marchionini, G., Nelson, M., Spinks, R., & Yang, M. (2001). Interface concepts for the Open Video Project. *Proceedings of the 2001 ASIST Annual Meeting, 38,* 58-75.

Ghani, J. A., Supnick, R., & Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. *Proceedings of the Twelfth International Conference on Information Systems (New York, December 16-18, 1991),* 229-237.

Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K.-P. (2004). Finding the flow in Web site search. *Communications of the ACM, 45*(9), 42-49.

Hughes, A. (2003). *When Do You Look and Why Do You Read? An Exploratory Study Regarding User Interaction with Visual and Textual Surrogates when Searching for Video Information***.** Unpublished Master's Paper, University of North Carolina at Chapel Hill.

Hughes, A., Wilkens, T., Wildemuth, B., & Marchionini, G. (2003). Text or pictures? An eyetracking study of how people view digital video surrogates. *Proceedings of the International Conference on Image and Video Retrieval (CIVR 2003),* 271-280.

Jaffe, J. M. (1995). Media interactivity, cognitive flexibility, and self-efficacy. Ph.D dissertation, University of Michigan.

Jörgensen, C. (1999). Access to pictorial material: a review of current research and future prospects. *Computers and the Humanities, 33*(4), 293-318.

Komlodi, A., & Slaughter, L. (1998). Visual video browsing interfaces using key frames. In *CHI 98 Summary, Human Factors in Computing Systems,* 337-338.

Kwasnik, B. H. (1992). A descriptive study of the functional components of browsing. In Larson, J., & Ungeer, C. (eds.), *Engineering for Human-Computer Interaction.* Elsevier Science Publishers, 191-203.

Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. *Journal of Organizational and End-User Computing, 16*(1), 59-72.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press.

McMillan, S. J. (2002). A four-part model of cyber-interactivity. *New Media & Society, 4*(2), 271-291.

Ngo, C.-W., Pong, T.-C., & Zhang, H.-J. (2001). On clustering and retrieval of video shots. *Proceedings of the ACM International Conference on Multimedia*, 51-60.

Rice, R. E., McCreadie, M., & Chang, S.-J. L. (2001). *Accessing and Browsing Information and Communication.* Cambridge, MA: MIT Press.

Saracevic, T., Kantor, P., Chamis, A. Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science, 39*(3), 161-176.

Smeaton, A. F. (2004). Indexing, browsing, and searching of digital video. *Annual Review of Information Science & Technology, 38*, 371-407.

Sonnenwald, D. H. (1998). Evolving perspectives of human information behaviour: Contexts, situations, social networks and information horizons. In Wilson, T. D., & Allen, D. K. (eds.), *Exploring the Contexts of Information Behaviour*. London: Taylor Graham, 176-190.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73-93.

Stromer-Galley, J. (2004). Interactivity-as-product and interactivity-as-process. *The Information Society, 20*, 391-394.

Sundar, S. S. (2004). Theorizing interactivity's effects. *The Information Society, 20*, 385-389.

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior, 9*, 411-426.

Wildemuth, B. M., Marchionini, G., Wilkens, T., Yang, M., Geisler, G., Fowler, B., Hughes, A., & Mu, X. (2002). Alternative surrogates for video objects in a digital library: users' perspectives on their relative usability. Presented at the European Conference on Digital Libraries (ECDL), Milan, Italy, September, 2002. http://www.open-video.org/ovadmin/ECDL2002.020620.pdf.

Wildemuth, B. M., Marchionini, G., Yang, M., Geisler, G., Wilkens, T., Hughes, A., & Gruss, R. (2003). How fast is too fast? Evaluating fast forward surrogates for digital video. Paper presented at the ACM/IEEE Joint Conference on Digital Libraries, Houston, May 2003. http://www.open-video.org/ovadmin/p221-wildemuth.pdf.

Yeo, B.-L., & Yeung, M. M. (1997). Retrieving and visualizing video. *Communications of the ACM, 40*(12), 43-52.