How Do Users Search with Basic HTML5 Video Players?

Claudiu Cobârzan and Klaus Schoeffmann

Alpen-Adria-Universität Klagenfurt
9020 Klagenfurt, Austria
{claudiu,ks}@itec.aau.at

Abstract. When searching within a video for a specific scene most non-expert users employ a basic video player. The main advantage of such a player over more advanced retrieval tools lies in its ease of use and familiar controls and mode of operation. This means that the available navigation controls (play, fast forward, fast reverse, seeker-bar) will be used for interactive search and browsing. We compare the search behavior by type of interaction and speed of interactive search of two groups of users, each numbering 17 participants. Both groups performed the same tasks using an HTML5 video player but in different setups: the first group performed Known Item Search tasks, while the second performed Description Based Search tasks. The goal of this study is twofold. One: better understand the way users search with a basic video player, so that useful insights can be taken into consideration when designing professional video browsing and search tools. Two: evaluate the impact of the different setups (Known Item Search vs. Description Based Search tasks).

Keywords: video search, video browsing, user behavior, HTML5 video player.

1 Introduction

The amount of video data made available on the Internet is continuously increasing thanks in part to social media and sharing platforms, as well as to the wide availability and popularity of video recording devices in consumer electronics. A large portion of the videos are recorded by non-professionals that are driven by the most diverse of motives [5]. Those non-professionals will mostly employ simple video players, not only for viewing the content, but also for searching for specific sequences. This is because the available navigation features like play, pause, fast-forward, fast-reverse as well as random access using a seeker-bar are familiar and easy to use. Those simple features are in fact preferred by non-expert users as reported in [10]. However, they limit the user experience and interaction metaphors especially in mobile setups. Recent research has begun to investigate more appropriate controls for such situations [3], [4].

Popular web based sharing platforms like YouTube often provide only the most basic video player functionality (play/pause buttons and a seeker-bar) and

C. Gurrin et al. (Eds.): MMM 2014, Part I, LNCS 8325, pp. 109–120, 2014.

dump some of the typical VCR controls like the fast-forward and the fast reverse buttons. Current implementations of standard HTML5 video players also tend to favor those basic controls. For example, the Safari implementation offers the fast-forward and fast-reverse functionalities only when the player runs in full-screen mode.

This paper investigates the way those basic controls are used for video browsing (a combination of video playback and video search) within long videos when searching for a specific sequence. It continues the work in [8], which presents evaluation results from a user study performed with 17 participants that had to solve "Known Item Search" (KIS) tasks used for the Video Browser Showdown [7]. We add the result from a second user study, also with 17 participants, which had to solve "Description Based Search" (DBS) tasks with the same long videos as in the Video Browser Showdown [7], but with the target videos replaced by their textual descriptions. We compare and discuss the results from the two studies.

We are aware of only another study that evaluates search strategies, but concentrates on VCR-like (play, pause, stop, fast-forward, fast-reverse) controls [1] and not typical HTML5 navigation features.

2 Related Work

User interaction with VCR-like controls has been studied in [1]. The authors performed a user study in which the tasks were of type "Known Item Fact Retrieval" or "I'll know it when I see it" [2]. They consisted of finding video segments within a small archive based on a semantic question. The available navigation controls were *play*, *stop*, *pause*, *fast rewind*, *fast forward*, *step reverse* and *step forward*. Four search strategies were identified within the study: (1) *incremental linear search* (55%), (2) *decremental linear search* (10%), (3) *educated guess* (29%), and (4) *random selection* (6%). Within one file, the browsing behavior consisted of: (1) *straight viewing* (21%), (2) *speed switching*: linear viewing with switching back and forth between playback and fast-forward (46%), (3) *inaccurate shuttle determination*: fast-forward too far, fast-reverse, then play – or if too far back, fast-forward again (13%), (4) *accurate shuttle determination*: similar to (3) but with step-forward and step-backward (7%), (5) *halt and refine*: step-forward and play but pause sometimes to reflect on where they are (13%). The fastest approach proved to be *speed switching* and *halt and refine*, *accurate shuttle determination*, *straight viewing*, and *inaccurate shuttle determination*.

In contrast, the study in [8] on user interaction by the means of a basic HTML5 video player, focused on the use of the default controls in non-full-screen mode: *play* and *pause* buttons and the *seeker-bar*. The users had to perform "Known Item Search" tasks: a short video sequence of 20 seconds was initially presented and then it had to be located within an hour long video. As far as we are aware of, this is the only recent study on interactive search and behavior while using modern players employing mainly *play*, *pause* and the *seeker-bar* as main

navigation aid. This is somewhat surprising since in recent years numerous video browsing tools have been proposed (a detailed review can be found in [9]).

Five navigation methods while interacting with the HTML5 video player were identified: (1) *Playback* (36%), (2) *Forward@Playback* (12%), (3) *Forward* (36%), (4) *Reverse@Playback* (7%) and (5) *Reverse* (10%). The best results in terms of completion time were obtained by users applying *linear forward search* with *seeker-bar dragging* in *non-playback state*.

3 Navigation Patterns: "Description Based Search" (DBS) vs. "Known Item Search" (KIS) Tasks

In [8], a preliminary report on the behavior of 17 users performing "Known Item Search" tasks within hour long videos, can be found. The tool used during the tests was a simple HTML5 video player (see Figure 1d). Following, we compare those findings with those of a newly performed user test also with 17 participants. The same HTML5 video player was used, but in a different setup: the users had to identify exactly the same sequences as in the KIS tasks but were presented only with the textual description of the video sequences that had to be located and not with the actual footage.

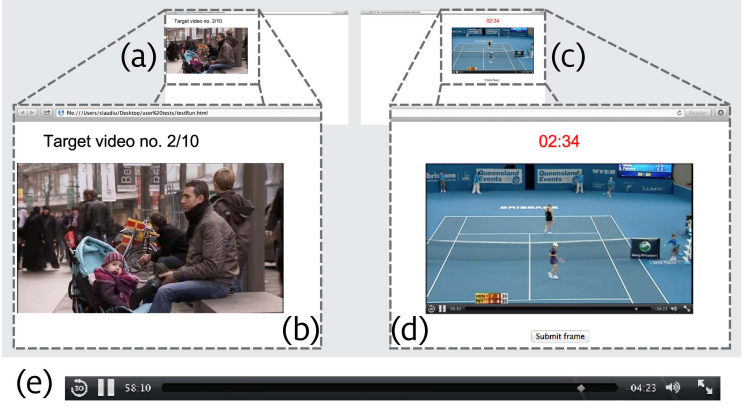


Fig. 1. The interface used in the *KIS* study. (a) and (b) the HTML5 video player during the first stage of a trial with the automatic playback of the target scene. (c) and (d) the HTML5 video player during the second stage of a trial: the effective search. (e) close up of the provided interaction possibilities of the video player.

3.1 User Studies

The data used in both studies is identical and consists of one-hour Dutch news videos. It comes from the public dataset of the Video Browser Showdown [7] in 2013 and it is currently available on its website. The same hardware was

used in both studies: an 17-inch MacBook Pro laptop with the resolution set at 1920×1200 pixels to which a wired optical mouse was connected. The applications' interfaces were presented in a Safari web-browser window in full-screen mode.

The first study in [8] (which we will call the *KIS* study throughout the rest of the paper) had 17 participants (2 female, 15 males) with ages between 23 and 52 years, all of which were daily computer users. Each had to complete 10 search tasks consisting of finding a 20 seconds video sequence within an hour long news video. For each task, the short sequence was played back once within an automatic playback player on the left side of a full screen window as shown in Figure 1a.

All interaction elements were removed during the playback (see close-up in Figure 1b). Once the playback ended, the users had up to 3 minutes to find the presented scene within the corresponding long video which was presented on the left side of the screen within a player with basic controls (play, pause, seeker-bar - Figure 1d and Figure 1e for a close-up).

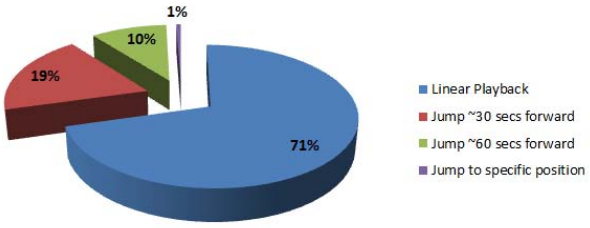
The second study (which we will call the *DBS* study throughout the rest of the paper) also had 17 participants (11 female, 6 males) with ages between 20 and 36 years and all daily computer users. They had to find the exact video sequences as the users in the *KIS* study but instead of the playback of the target scene, they were presented also on the left side of the screen with its textual description. For example, the textual description of the target video no. 2 (a frame of which is shown in Figure 1b) is as follows: **Find the video sequence showing people in a shopping street where a man (Tom Sluyts) takes care for his girl, sitting in a baby buggy and dressed with pink clothes, before being interviewed. A couple is coming out of a shop; the man is obviously walking to wrong direction**. No time limit was imposed on reading the description. The users had a *Start test* button that allowed them to start the 3 minutes timer and get access to the long video on the left side of the screen within the player with the same basic controls as the one in the *KIS* study (play, pause, seeker-bar - Figure 1d and e for a close-up).

In both studies, the basic controls of the HTML5 player could be used to navigate within the long video in search of a frame belonging to the target segment (in the case of the *KIS* study) or fitting the presented description (in the case of the *DBS* study). In both studies, the participants used the *Submit frame* button below the player to check whether the current displayed frame actually belonged to the target video or it fitted the description. False submissions were signaled by setting the background red for 4 seconds, while correct ones were signaled by a green background and the achieved score being displayed for 10 seconds. The next trial was available after a successful submission or after the allowed 3 minutes search time period expired.

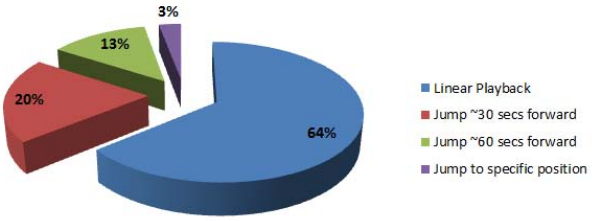
3.2 Discussion

In the following we will discuss and compare the interaction types as well as the interaction speed we have observed during the two discussed tests. Also some interesting particular cases that came up during the two tests will be presented.

Search Start Interaction. To our surprise, the users in both the *KIS* and *DBS* studies approached the tasks almost in the same manner, as can be seen in Figure 2. The users in the *DBS* study (see Figure 2a) started in 71% of all 170 tasks with playback from the beginning of the video performing linear search forward by forward navigation. The users in the *KIS* study (see Figure 2b) employed the same approach in 64% of their 170 tasks. In 29% of the tasks in the *DBS* study respectively 33% of the tasks in the *KIS* study, the users preferred to start the search with a jump within the video. For *DBS*, a 30 second jump was recorded for 19% of the tasks, while for the *KIS* study, the same jump of 30 seconds appeared for 20% of the tasks. A 60 seconds jump was recorded for 10% of *DBS* tasks and 13% of *KIS* tasks. Random positioning, which usually marks some kind of educated or instinctive guess, was seldom employed. The users trusted their luck only for 1% of the *DBS* tasks and 3% of the *KIS* tasks.



(a) *DBS* study

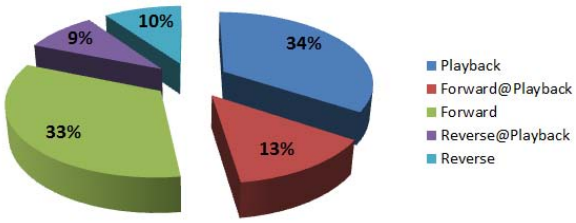


(b) *KIS* study

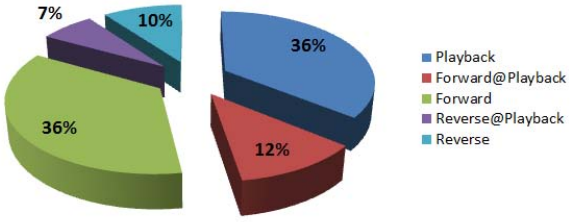
Fig. 2. Interaction methods used to start the search

Interaction Types. The application logs revealed for both our studies two classes of user behavior:

- *Click & Play*: the play button of the HTML5 video player is pressed and then multiple clicks are performed on the seeker-bar towards the end (*forward*) or towards the beginning (*reverse*) of the video. The pause button is sometimes used (usually when a certain frame is examined or when it is submitted for evaluation)
- *Dragging*: the player's seeker-bar is used to browse the content in both directions. The play and pause buttons are rarely used.



(a) *DBS* study



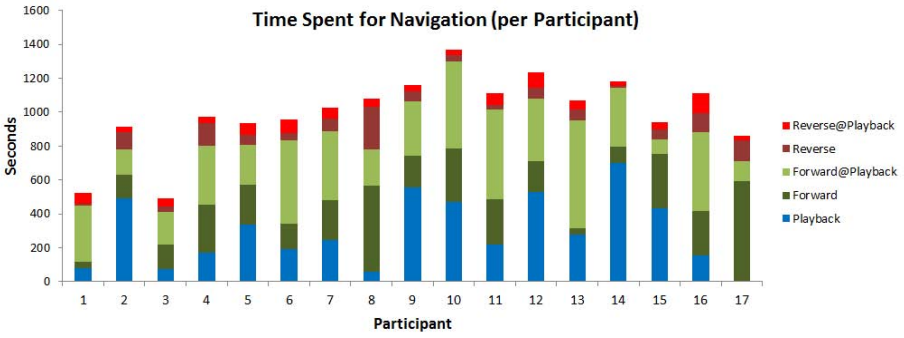
(b) *KIS* study

Fig. 3. Navigation methods used in both studies

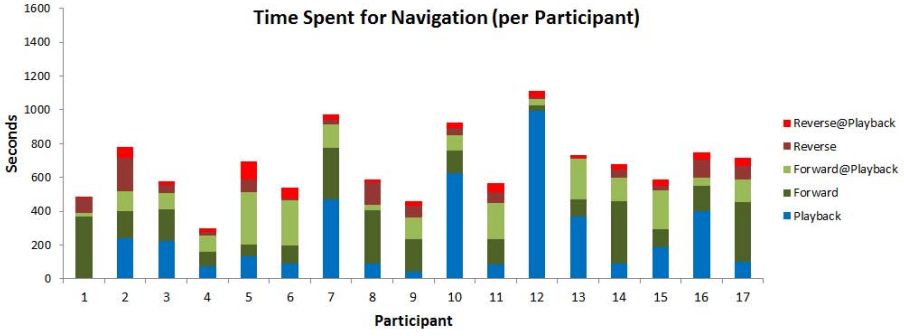
The navigation methods are surprisingly similar in both *DBS* and *KIS* studies, as can be seen in Figure 3. Playback, dragging or clicking to a future point in time while in playback (*Forward@Playback*) or pause state (*Forward*) accounted for aprox. 81% of the navigation in the *DBS* study (see Figure 3a) and aprox. 83% of the navigation in the *KIS* study (see Figure 3b). Reverse positioning in pause (*Reverse*) or playback (*Reverse@Playback*) states accounted for only 19%, respectively 17% in *DBS* and *KIS*.

Navigation Strategies. Individual users showed in both studies varied strategies (see Figure 4). Overall, the users in the *DBS* study were significant slower

than the ones in the *KIS* study. The approaches attempted in the *DBS* study appear to be more consistent (Figure 4a), while the ones in the *KIS* study appear a little bit more diverse (Figure 4b). Participant 17 in the *DBS* study and participant 1 in the *KIS* study did not use playback at all, while participants 14 in the *DBS* and participants 10 and 12 in *KIS*, used playback for most of the search time. Reverse positioning was used by some of the participants of the *KIS* study almost equally long as forward positioning (participants 2 and 16). In contrast, all the participants in the *DBS* study showed a strong preference for forward navigation. All the participants in the two studies, preferred positioning in paused mode over positioning in playback mode. If we compare Figure 4 with Figure 5, it becomes apparent that users that employed a lot of *Dragging* (e.g. participants 1, 4 and 9 in the *KIS* study) had significant less frames than those preferring *Click&Play* (e.g. participants 7, 10 and 12 also from the *KIS* study) and they were also a lot faster - in fact, they had the fastest overall submission in the *KIS* study. The most effective approach proved to be interactive search using the seeker-bar without playback.

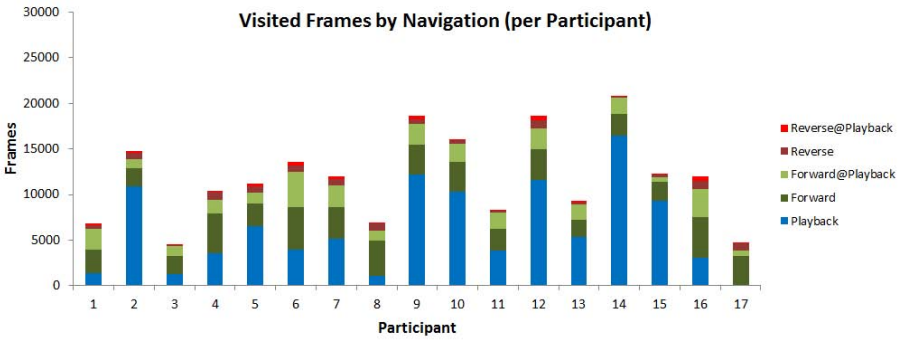


(a) *DBS* study

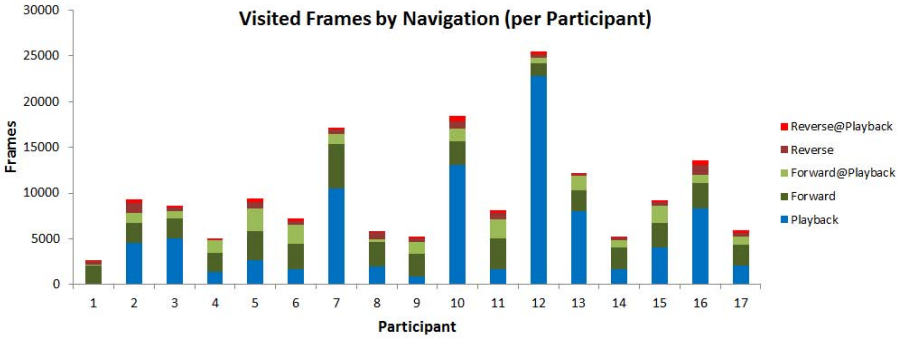


(b) *KIS* study

Fig. 4. Time spent for a specific search method (per user)



(a) *DBS* study



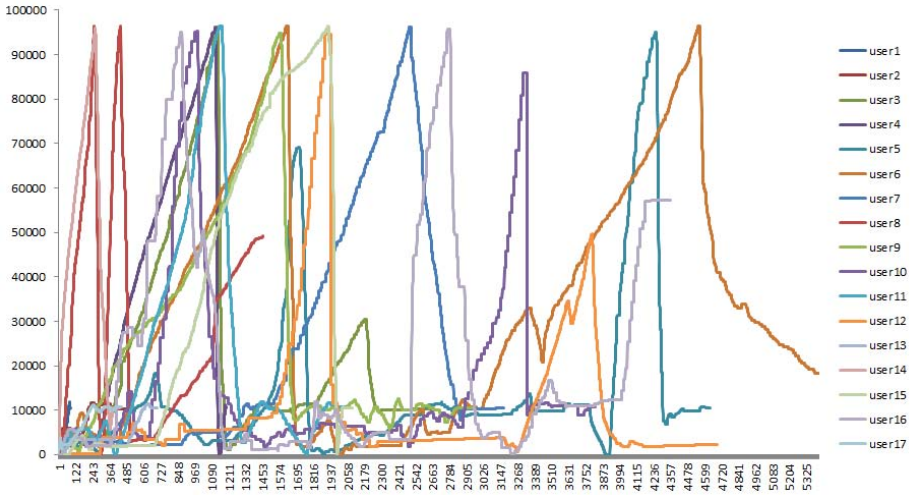
(b) *KIS* study

Fig. 5. Number of frames visited per participant by using a specific search method

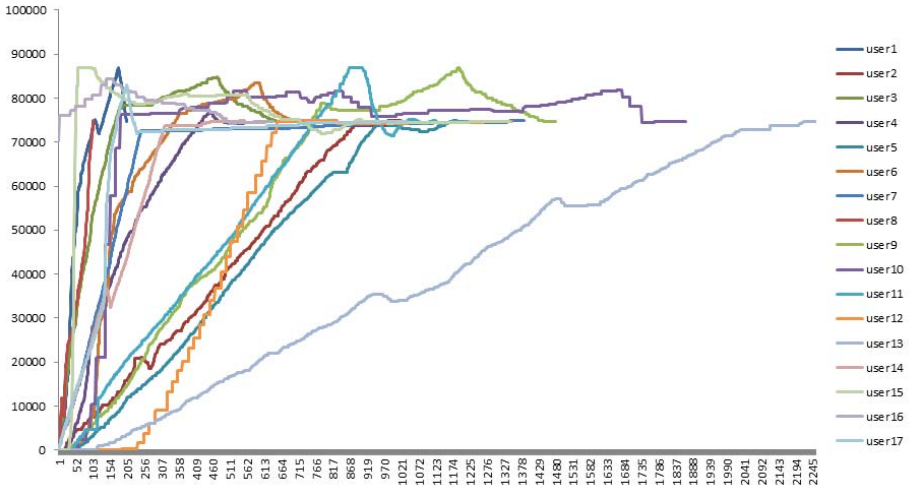
Interesting Particular Cases. The analysis of the log files from the two user studies has revealed some interesting situations regarding the approaches individual users employed while searching for a certain video sequence, as well as interesting similarities when the navigation patterns of the users were plotted together.

Not all the tasks within each study were approached in the same manner. Figure 6 shows how the participants in the *KIS* study approached two different tasks: one that they described in the follow up interviews as being "difficult" (see Figure 6a for details regarding the scene) and one that was considered "easy" (details in Figure 6b).

For the "difficult" task, many users searched repeatedly from the beginning towards the end since they were not able to identify the target scene (see the "saw-tooth" pattern in Figure 6a). Others, correctly located the target scene at the beginning of the long video, but had difficulties in pinpointing its exact location because there were multiple very similar sequences (see the erratic movement between 0 and 10000 on the *0y* axis also in Figure 6a).

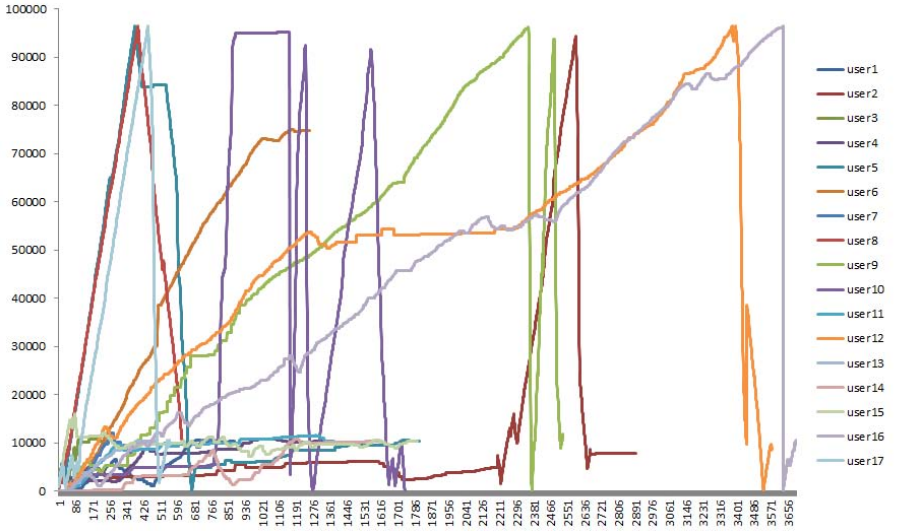


(a) **Navigation diagram of all participants for a "difficult" task in the *KIS* study**: in split screen a reporter talks to an anchor person in a studio. Almost identical scenes with the two people appear multiple times within the news program in and outside the target area. This made it hard to locate the exact scene.

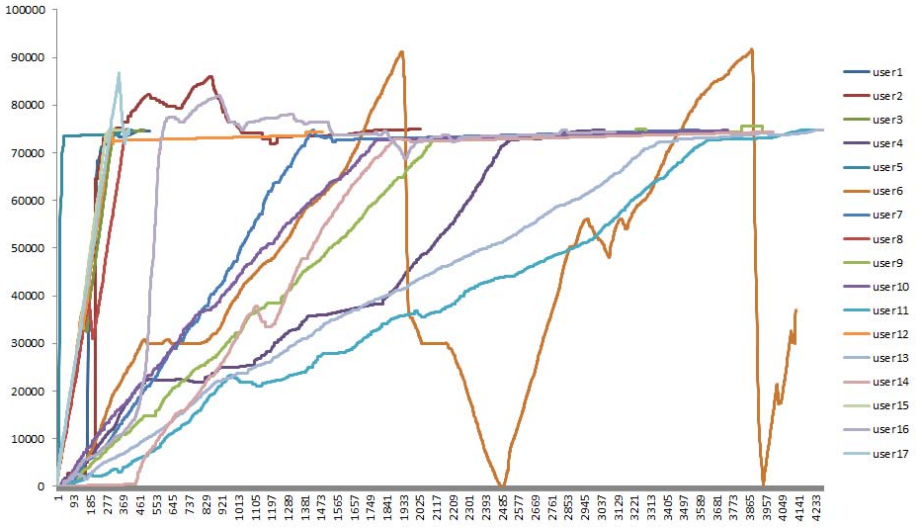


(b) **Navigation diagram diagram of all participants for an "easy" task in the *KIS* study**: a goal scene from a soccer match is shown. The player scoring the goal wears a white t-shirt and blue shorts. The coach (Ariel Jacobs) from Anderlecht gives an interview in front of an UEFA Europa League sponsor wall.

Fig. 6. *KIS* study: "Difficult" vs. "Easy" tasks



(a) **Navigation diagram of all participants for a "difficult" task in the *DBS* study**: in split screen a reporter talks to an anchor person in a studio. Almost identical scenes with the two people appear multiple times within the news program in and outside the target area. This made it hard to locate the exact scene.



(b) **Navigation diagram diagram of all participants for an "easy" task in the *DBS* study**: a goal scene from a soccer match is shown. The player scoring the goal wears a white t-shirt and blue shorts. The coach (Ariel Jacobs) from Anderlecht gives an interview in front of an UEFA Europa League sponsor wall.

Fig. 7. *DBS* study: "Difficult" vs. "Easy" tasks

For the "easy" task, most of the users made an "educated guess" that the target scene has to be located towards the end of the video and acted accordingly. The majority moved very quickly to the end and then concentrated on finding the exact scene (the lines which are the closest to the *0y* axis in Figure 6b). The others also navigated towards the end, but at a slower pase (the next group of lines more towards the middle in Figure 6b). A single user (**user 13**) chose to ignore the "obvious" hint provided by the target scene preview and applied linear search for the entire duration of the test. This is a user who changed his strategy in mid-test session and switched from an awkward approach in which he searched from the beginning to the end of the video and then reverted the direction from the end towards the beginning. After recognizing it is a failing strategy, he switched for the rest of the session to linear search, which he stubbornly applied even when there were indications that other approaches might be more appropriate.

In Figure 7 we present the approaches made by the participants in the *DBS* study for solving the exact same pair of "difficult" and "easy" tasks. The similarities between the approaches taken by the participants can easily be seen. We have basically the same behavior for both of the tasks. For the "difficult" one, some of the users struggle to locate the scene, hence the repeated dragging towards the end of the video and back to the beginning (the same "saw-tooth" patterns in Figure7a similar to the ones in Figure 6a. Others recognized the scene in the beginning, hence the movement between 0 and 20000 on the *0y* axis also in Figure 7a).

For the "easy" task most of the users in the *DBS* study also made the "guess" that the target scene lies somewhere towards the end of the long video. Some moved quickly (the lines which are most close to the *0y* axis in Figure 7b), while others applied linear search while dragging and finally concentrated on the last part of the video. The study also had a participant who had a hard time in locating the target scene in this test (the single "saw-tooth" line in Figure 7b corresponding to **user 6**). He had to start two times from the beginning after unsuccessful browsing two times through the video.

4 Conclusion

We have presented and compared the results from two user studies which focused on interactive search using basic HTML5 video players with limited navigation features. The adopted strategies vary quite significantly, especially in the case of the *KIS* study, but they also share some common characteristics. Most users (and especially the ones participating in the *DBS* study) favored linear forward search with seeker-bars positioning, as it helped alleviate the fact that the target scenes were introduced only by their textual description. Reverse search was seldom used and most of the times did not lead to success. The alternative to reverse search that almost all the participants adopted was to start fresh from the beginning. Linear search with and without playback was preferred since it helps to remember visited segments. This is especially important when the user does not clearly recollect the content in the target scene.

Our two studies also show that linear forward search with seeker-bar dragging in non-playback state is the most efficient in terms of search time, since the users can concentrate on solving the actual task. This means that video search tools do not necessarily have to provide playback feature and instead can confidently employ static images for interactive search.

Acknowledgments. This work was funded by the Federal Ministry for Transport, Innovation and Technology (bmvit) and the Austrian Science Fund (FWF): TRP 273-N15 and by Lakeside Labs GmbH, Klagenfurt, Austria, and funding from the European Regional Development Fund (ERDF) and the Carinthian Economic Promotion Fund (KWF).

References

1. Crockford, C., Agius, H.: An empirical investigation into user navigation of digital video using the vcr-like control set. International Journal of Human-Computer Studies 64(4), 340–355 (2006)
2. Huang, A.-H.: Effects of multimedia on document browsing and navigation: an exploratory empirical investigation. Information & Management 41(2), 189–198 (2003)
3. Hudelist, M., Schoeffmann, K., Böszörményi, L.: Mobile Video Browsing with a 3D Filmstrip. In: Proceedings of the 3rd ACM International Conference on Multimedia Retrieval, pp. 299–300. ACM, New York (2013)
4. Hudelist, M., Schoeffmann, K., Böszörményi, L.: Mobile Video Browsing with the ThumbBrowser. In: Proceedings of the 21st ACM Conference on Multimedia. ACM, Barcelona (accepted for publication 2013)
5. Lux, M., Huber, J.: Why did you record this video? An exploratory study on user intentions for video production. In: 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 1–4 (2012)
6. Over, P., Awad, G., Michel, M., Fiscus, J., Sanders, G., Shaw, B., Kraaij, W., Smeaton, A.-F., Quénot, G.: Trecvid 2012 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In: Proceedings of TRECVID 2012, NIST, USA (2012)
7. Schoeffmann, K., Bailer, W.: Video browser showdown. SIGMultimedia Rec. 4(2), 1–2 (2012)
8. Schoeffmann, K., Cobârzan, C.: An Evaluation of Interactive Search with Modern Video Players. In: IEEE International Conference on Multimedia and Expo (ICME), pp. 1–4. IEEE, San-Jose (2013)
9. Schoeffmann, K., Hopfgartner, F., Marques, O., Böszörményi, L., Jose, J.-M.: Video browsing interfaces and applications: a review. SPIE Reviews 1(1), 018004 (2010)
10. Scott, D., Hopfgartner, F., Guo, J., Gurrin, C.: Evaluating novice and expert users on handheld video retrieval systems. In: Li, S., El Saddik, A., Wang, M., Mei, T., Sebe, N., Yan, S., Hong, R., Gurrin, C. (eds.) MMM 2013, Part II. LNCS, vol. 7733, pp. 69–78. Springer, Heidelberg (2013)