Shaping Video Experiences with New Interface Affordances

by

Abir Al Hajri

B.Sc., Sultan Qaboos University, 2003

M.Sc., University of Wollongong, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

The University of British Columbia

(Vancouver)

December 2014

Abstract

Watching and creating videos have become predominant parts of our daily lives. Video is becoming the norm for a wide range of purposes from entertainment, to training and education, marketing, and communication. Users go beyond just watching videos. They want to experience and interact with content across the different types of videos. As they do so, significant digital traces accumulate on the viewed videos which provide an important source of information for designing and developing tools for video viewing interfaces.

This dissertation proposes the next generation video management interface which creates video experiences that go beyond just pushing the play button. It uses how people view and interact with contemporary video to design strategies for future video interfaces. This has allowed the development of new tools for navigating and managing videos that can be easily integrated into existing systems.

To help define some design guidelines for the video interface, a behavioural analysis of users' video viewing actions ($n = 19$) was performed. The results demonstrate that participants actively watch videos and most participants tend to skip parts of videos and re-watch specific portions from a video multiple times. Based on the findings, new fast navigation and management strategies are developed and validated in search tasks using a single-video history ($n = 12$), a video viewing summary ($n = 10$) and multiple-videos history ($n = 10$). Evaluation of results of the proposed tools show significant performance improvements over the state-of-the-practice methods. This indicates the value of users' video viewing actions.

Navigating other forms of videos, such as interactive videos, introduces another issue with the selection of interactive objects within videos to direct users to

different portions of the video. Due to the time-based nature of the videos, these interactive objects are only visible for a certain duration of the video, which makes their activation difficult. To alleviate this problem a novel acquisition technique (*Hold*) is created, which temporally pauses the objects while the user interacts with the target. This technique has been integrated into a rich media interface (*MediaDiver*) which made such interaction possible for users.

Preface

All of the research work presented in this dissertation was conducted in the Human Communication Technologies Laboratory (HCT) at the University of British Columbia, Point Grey campus. All user studies and associated methods were approved by the University of British Columbia Behavioural Research Ethics Board [certificates #: H10-01897, H08-03006, and H13-01589].

Parts of this dissertation have been published elsewhere. Earlier versions of Sections 4.4, 4.5, 4.6 and 5.1 have been published in A. Al Hajri, G. Miller, S. Fels and M. Fong [5]. I was the lead investigator, responsible for all major areas of concept formation, literature review, interfaces (pilot and main) design and implementation, experimental design, data collection and analysis, as well as manuscript composition. G. Miller was involved in the discussion and testing of the interface, and assisted with writing of the manuscript. S. Fels provided feedback on the design and manuscript. M. Fong helped in the implementation of some parts of the main interface.

An earlier version of Section 5.2 has been published in A. Al Hajri, M. Fong, G. Miller and S. Fels [6]. I developed and implemented the interface jointly with M. Fong. I was responsible for literature review, experimental design, data collection and analysis, as well as manuscript composition. M. Fong implemented the *VCR* algorithm, was involved in the discussion of the interface and provided feedback on the manuscript. G. Miller assisted with the interface design and testing, and writing the manuscript. S. Fels provided editorial feedback on the manuscript.

A version of Chapter 6 has been published in A. Al Hajri, G. Miller, M. Fong and S. Fels [8]. I was the lead investigator, responsible for all major areas of concept formation, literature review, interfaces (pilot and main) design and imple-

mentation, experimental design, data collection and analysis, as well as manuscript composition. G. Miller was involved in the discussion and testing of the interface, and assisted with writing. M. Fong helped in the implementation of some parts of the interface and provided feedback on the manuscript. S. Fels provided feedback on the design and manuscript.

An overview of the visualizations described in Sections 5.2.2 and 6.3 has been presented in a poster and published in A. Al Hajri, M. Fong, G. Miller and S. Fels [7]. I designed the poster and wrote the manuscript. G. Miller assisted with the revisions of the manuscript and poster. S. Fels was the supervisory author on this project.

Earlier versions of Sections 7.1, 7.2 and 7.2 have been published in A. Al Hajri, S. Fels, G. Miller and M. Ilich [4]. I formulated the mathematical model, performed and evaluated the model, was responsible for literature review, experimental design, data collection and analysis, as well as manuscript composition. G. Miller helped with writing the manuscript and provided assistance on simplifying the model using a vector notation. I have adapted M. Ilich's experimental interface to evaluate the model and test the *Hold* technique. M. Ilich originally proposed the *Hold* technique which I adapted, modeled and evaluated, and he provided feedback on the manuscript. S. Fels provided assistance on the manuscript revisions.

A version of Section 7.4 has been published in G. Miller, S. Fels, A. Al Hajri, M. Ilich, Z. Foly-Fisher, M. Fernandez and D. Jang [85]. I designed and implemented the interface jointly with Z. Foley-Fisher, M. Fernandez and M. Fong. I wrote the manuscript. M. Ilich and D. Jang implemented earlier versions of some aspects of the interface that were adapted in the produced system. G. Miller and S. Fels were the project leaders. G. Miller was involved in the discussion and design of the interface, and helped in the manuscript writing and revisions. S. Fels provided feedback on the design and manuscript.

An earlier version of Chapter 8 has been presented as an interactive demonstration in A. Al Hajri, G. Miller, M. Fong and S. Fels [7, 8] and GRAND '14. I designed, implemented and tested the application jointly with M. Fong. G. Miller and S. Fels were involved in the discussion of the design of the application. I wrote the manuscript for GRAND '14 and G. Miller provided assistance on the manuscript revisions. M. Fong was responsible for the experimental design and

data collection. G. Miller and I were involved in the discussion of the study design and helped in the data collection.

The video content of the screenshots in Figures 1.2, 1.4, 1.6, 4.2, 4.3, 4.4, 4.5, 5.7, 5.8, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.8, 6.9, 8.1, 8.3, 8.4, 8.5, 8.6, 8.8, and 8.9 is © copyright 2008, Blender Foundation.

The figures in Chapter 2 are used under permission as follows:

- Figure 2.1 from © [28] page 11. Used with permission from Chorianopoulos, K.
- Figure 2.2 from © [82] page 42. Used under permission from Brusilovsky, P.
- Figure 2.3 from © [69] page 566. Used under permission from Kim, J.
- Figure 2.4 from © [22] page 202. Used under permission from Ooi, W. T.
- Figure 2.6 from © [115] page 123. Used under permission from Kaasalainen, J.
- Figure 2.7 from ©Aderhold, L., Kao, K. & Hilden, J. (2009). Tabviz. Retrieved from http://tabviz.org/. By permission from Aderhold, L.
- Figure 2.8 from © [83] page 692. Used under permission from Milic-Frayling, N.
- Figure 2.9 from © [120] page 317. Used under permission from Shintani, T.
- Figures 2.11 and 2.12 from © [79] pages 1159-1160. Used under permission from Matejka, J.
- Figure 2.13 from © [92] page 1222. Used under permission from Huber, M. & Olivier, P.
- Figure 2.14from © [25] page 789. Used under permission from Chen, B. Y.
- Figure 2.15 from © [71] page 202. Used under permission from Girgensohn, A.

- Figure 2.16 from © [105] page 278. Used under permission from Shaw, R. & Liu, Y.
- Figure 2.17 from © [13] page 59. Used under permission from Robbins, D., Czerwinski, M., Bederson, B. & Cutrell, E.
- Figure 2.18 from © [52] page 3331. Used under permission from Anderson, J.
- Figure 2.19 from © [47] page 281. Used under permission from Balakrishnan, R.
- Figure 2.20 from © [53] page 845. Used under permission from Hasan, K.

Table of Contents

List of Tables

List of Figures

Glossary

1D One Dimensional

2D Two Dimensional

3D Three Dimensional

ANOVA Analysis of Variance, a set of statistical techniques to identify sources of variability between groups

GLMMS Generalized Linear Mixed Models, a set of statistical techniques to identify source of variability between groups for non-normal distributed dependent variables

K-MEANS Clustering method aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean

URL Uniform Resource Locator, used to describe a means for obtaining some resource on the World Wide Web

VCR View Count Record, a video navigation tool based on users' viewing statistics

VVB Video Viewing Behaviour, a Google Chrome extension used to track users' viewing activity on YouTube

XML Extensible Markup Language

Acknowledgments

First and foremost, I must acknowledge my praises and limitless thanks to God Almighty, the most gracious and merciful, for His help and blessing. I am certain that this work would not have materialized without His guidance.

Most of all, I would like to express my deepest appreciation to my supervisor Professor Sidney Fels, for his excellent guidance and advice throughout the course of my PhD research. I have learned so much, and without him, this work would not have been possible. Thank you so much for believing in my abilities, supporting me, and making this journey such a rewarding experience.

I wish also to express my heartfelt gratitude to Dr. Gregor Miller for his precious guidance, continuous encouragement and support that motivated me to move forward. I thank him for his constructive criticism, personal involvement in all stages of my work, and belief in my capabilities throughout this research. His comments and suggestions have been invaluable.

I would like also to thank Dr. Ali Mesbah, Dr. Karthik Pattabiraman and Dr. Konstantin Beznosov for agreeing to serve on my supervisory committee and Dr. Luanne Freund, Dr. Jane Wang and Dr. Ravin Balakrishnan for being on my dissertation examination committee.

I owe a deep debt of gratitude to my university (Sultan Qaboos University, Oman) for giving me this opportunity to pursue my PhD.

I must acknowledge as well my fellow colleagues at the HCT lab, for their friendship and support and special thanks to Matthew Fong who worked with me closely on some designs used in this dissertation. The last few years have been quite an experience and you have all made it a memorable time of my life.

I would like to take this opportunity to also thank all the participants who took

part in the different user studies. Without their contribution, none of this research would have been possible.

Last but not least, my wholehearted thanks go to my family and friends for their love and support during this phase of my life. Special thanks to my parents without whom none of this would have been possible. I remain indebted to them for their generous support throughout my professional career and particularly during the process of pursuing this PhD degree. Because of their unconditional love, patience, encouragement, and prayers, I have had the chance to complete this dissertation.

Dedication

To my parents, Aisha and Mohammed, who prayed constantly for me to reach this point in life. Thank you for your support, unconditional and endless love, guidance, encouragement and patience while I was thousands miles away pursuing my Ph.D. research.

Chapter 1

Introduction

Video has become a predominant part of a user's daily life, with the emergence of online video providers [1] such as YouTube™, Vimeo™, and Dailymotion™. The rapid growth rate of video on these providers is due to the recent proliferation of mobile devices with cameras, and faster data speeds which has changed the concept of users from just being consumers of content to content creators. This has led to more and more content becoming available online that billions of consumers can experience and enjoy with just a simple click.

Consuming video online, on mobile devices or on home computers is now a well-accepted form of communication and entertainment. The use of video is not limited to the entertainment media; it is being used in various ways to drive sales, entertain, communicate and educate. This variety in the types and purposes of videos has revolutionized how consumers access, view and interact with media content. Managing, navigating, accessing and sharing specific information from such content is not trivial and often imposes high cognitive workload and physical navigation burdens on the user. Different approaches and designs have been proposed by researchers to tackle some of these aspects, which are discussed in Chapter 2. However, due to the great increase in the quantity of video now available, the variability of content, the evolving nature of the use and consumption of video, and the change of users' viewing behaviour, we need better management

[1] YouTube:http://www.youtube.com/; Vimeo:http://www.vimeo.com/; DailyMotion:http://www.dailymotion.com/

and navigation tools to access what we want more efficiently and to provide us with mechanisms to find previously seen content for us to use or share.

To illustrate the motivation behind this research and how these requirements manifest in a real-world setting, we consider a scenario from an entertainment context since it makes up a quarter of the videos viewed online as shown in Chapter 3. Let us look at Tom who spends 25% of his online time watching entertainment videos.

Last week, as usual, Tom, watched comedy videos on YouTube. Tom watched a funny one-hour cat video mix and laughed at many parts. He thought, "I'd love to share the funny bits with Sally." He clicked on the seek bar, scrubbed around finding the funny bits and wrote down the time codes. He emailed Sally, "Hey, Sally, check these out: http://youtube/funnycat; 00:00:10:37 - 00:00:13:21; 00:03:23:10 - 00:04:10:21; 00:13:42:07 - 00:15:07:22; 00:28:10:11 - 00:28:47:00; 00:39:01:29 - 00:39:56:19; 00:51:19:10 - 00:52:10:21."

Tom noticed that it was already 9:00 PM. So he started watching his favourite show "The Amazing Race." His favourite team did well through most of the challenges. They almost won, but they missed one trick in the last challenge and the cowboys won. Tom remembered that this team was participating in a previous season but he could not remember which season. So he tried to click on one of the team members to retrieve more information about them. It was hard to select them since they were moving so fast. Tom paused the video and clicked on one of them, which gave him more details about that participant. "Oh, yeah here it is. They were in the 2010 season", he said.

The next day, Tom went to school and he talked with his friend Sally about the show and his favourite team's performance in the episode. He brought up the webpage on his phone and looked for the episode. He jumped around the video trying to find when his team was struggling in one of the challenges. "Here it is. Check it" Tom said. "It is easy. How they did not get it?" Sally asked. "I missed this episode, but you know what, it reminds me of a challenge from last season" Sally said. He said "Oh really. Which episode was that?" "Let me check. I think I have it in my shows list. Let me bring that up for you", Sally said. They searched for the video, and jumped around the video trying to find the shot. "No, No, I think it happened sometime when one of the sisters team jumped from the cliff" she said.

They were looking around trying to find it. "Oh! Here it is, here it is! Sally said. "That is really hard. Can you send it to me? I am creating a list of hard challenges as I am planning to participate in the next season" Tom said. "Yeah sure, here it is. Shared the video" Sally said. "Oh, no, I just want that challenge part" He said. Sally said "Sorry, current viewers don't enable that."

In order to offer Tom and others the accessibility and functionality emphasized in this scenario, we need to design a video interface that allows users to (1) watch videos, (2) easily navigate videos, (3) access previously seen content within videos, (4) quickly find previously seen content, (5) share specific portions from videos, (6) generate summaries of previously seen videos, and (7) easily select interactive objects within videos. Some strategies, features and design guidelines need to be developed and added to the current video interfaces to overcome some of the challenges of meeting these requirements. Figure 1.1 illustrates what we imagine the future video viewing interface will look like and the different features that need to be added to the current video viewers.

Creating a next generation video interface that meets the above requirements brings up some challenges, including knowing how people navigate and interact with the new types of content, as well as matching it to the cognitive mechanism that people have when dealing with time-based media. As users view and navigate various videos, their viewing patterns, annotations comments and so on, can be thought of as digital "footprints" left on the videos. Some of these footprints are generated explicitly when users intentionally make a specific action around the media while viewing the content. This may include, for example, rating a video, commenting, adding annotation, tweeting, or voting. However, the majority of the footprints are generated unintentionally (i.e. implicit-user metadata) by the nature of simply interacting with the media without requiring any additional actions from the user. For example, a user's physiological response, facial expression, eye movements, visiting a video, viewing, and video interaction clickstreams, such as play, pause, skip, replay, or seek/scrub. This data can be examined to characterize video watching and navigation patterns that can be employed in designing new tools and interfaces for personalized video viewers. Moreover, the content being viewed from each video and how frequent it has been viewed can signify important meaning for such content. This information can then be used to assist the design

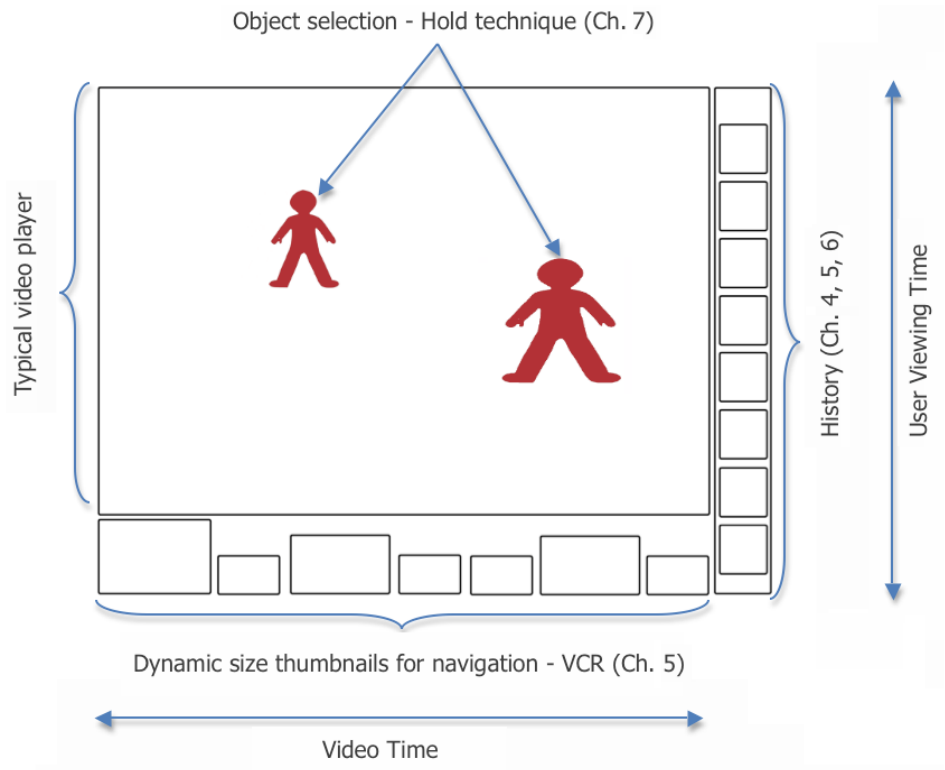


Figure 1.1: An illustration of a future video interface with the contributions made by this dissertation. The interface allows users to view and navigate videos while keeping track of their viewing history and providing access to this data. Users' viewed content can be accessed using the history component. Navigating the playing video can be achieved using the *VCR* and the *Hold* technique. (Note: This figure lays down the elements and interactions defined in this dissertation.)

of new tools to navigate and search video content.

Viewing pattern statistics are important to facilitate interface design to match how people watch videos. There is a significant difference between linearly watching a feature length movie (such as found on Netflix) and short educational modules or comedy videos, such as on edX[2] or YouTube, for example. Thus, characterizing the contemporary watching patterns would enable developing personalized tools

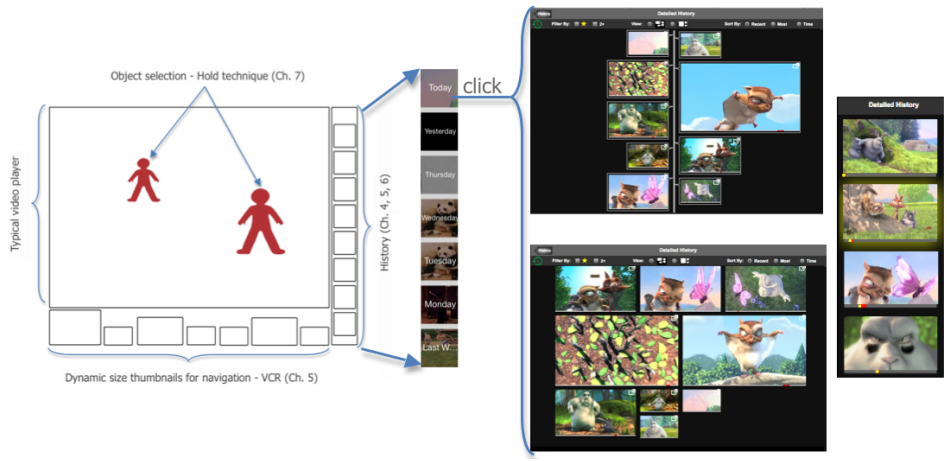
[2]https://www.edx.org/

that could satisfy users' needs. In the first part of this dissertation, we study users' viewing behaviour on YouTube through web browsers on a desktop platform that is discussed in more detail in Chapter 3. We explore whether users' activity while watching videos can provide some insights that can be turned into new tools for navigating video content and searching previously seen content. Through this study we show that a change exists in the way we view and experience videos besides just sitting and watching videos passively from start to end without any interaction during the playback of the content. We have been able to explore how people interact with videos from different categories. And, in contrast to previous research, we have looked at the behaviour of each individual user and how often they perform the different interactions while viewing. This enables us to determine seven different behaviours a user may exhibit while viewing videos, which in return helped us define some design guidelines for a more personalized video interface tailored to these behaviours as shown in Figure 1.1.

Researchers [44, 69, 111] have put substantial effort in extracting meaning from users' digital footprints and turning it into targeted applications. Many applications have been developed in the literature by mining this kind of data, as discussed in more detail in Chapter 2. These applications indicate that users' digital footprints provide a rich resource that can be leveraged for viewers. Nonetheless, this data is not accessible to users and only researchers use it to define a set of tools for consumers. We are interested in providing this data to the users themselves to create more personalized experiences, and investigate how they will use it, what other applications may emerge, and whether this data is going to change the way people view videos.

Our study (Chapter 3) has showed a high revisitation of videos to access previously seen content, which implies that users very often go back to videos they have seen to search for specific portions or information. Providing users with what they have seen from each video can help them to easily find what they are looking for. We call this kind of data **Video Viewing History**, which is simply an archive of each interval a user has watched from any video. History of users' actions has been widely investigated for multiple purposes in different domains including web browsers [21, 60, 75], documents editing [9, 56], workflow [49, 91], tutorial generation [15] and information visualizations [51, 54]. Researchers have introduced

and developed different tools that keep and visualize records of users' actions for later use. In comparison to the user's actions history in these domains, video viewing history is more difficult since we are not only dealing with the temporal nature of users' experience, we are also dealing with the temporal nature of the media itself; whereas, history for web browsing, for example, has a user time and no media time. This additional complication with video history, therefore, requires more sophisticated representation and visualizations to communicate such data than those proposed for the history in the other domains.



(a) (b) (c)

Figure 1.2: The history component of the video interface (a). Clicking on one of the thumbnails from history (b) brings a detailed history in another screen. Three different designs were developed and tested for the detailed history as show in (c).

To provide users (e.g. Tom and Sally in the scenario) access to their previously seen content that can be used for different purposes, we need to design an interface that allows users to watch videos, keeps track of their viewing history as well as providing access to such data. This will allow us to explore the usability of viewing history and investigate how it can improve users' task performance such as search and navigation. To achieve this, a history component (Figure 1.2(b)) is added to a video interface (Figure 1.2(a)). Through testing and learning from users, a series of modifications (Figure 1.2(c)) and experiments are applied to the design of this

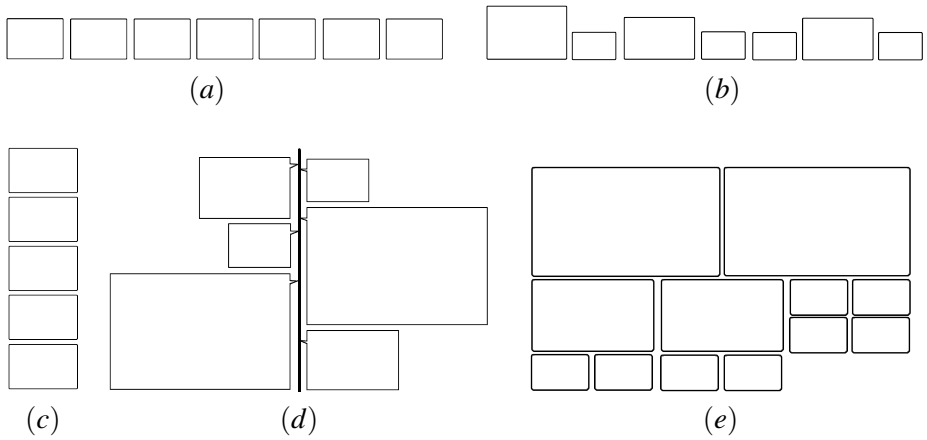


Figure 1.3: Viewing history visualizations where each interval is represented by a quadrilateral shape: the bigger the size the more that interval is watched. Four designs are proposed and evaluated in this dissertation: (c) *History Timeline* and (b) View Count Record (*VCR*) for single-video history, and (d) *Video Timeline* and (e) *Video Tiles* for multiple-video history. *VCR* is compared against the state-of-the-practice navigation method filmstrip (a).

component as detailed in Chapters 4, 5, 6 and 8. This leads to two designs for a single-video history: History Timeline (Figure 1.3(c)) and View Count Record (VCR) (Figure 1.3(b)); and two designs for multiple-video history: Video Timeline (Figure 1.3(d)) and Video Tiles (Figure 1.3(e)). Each quadrilateral shape in these designs refers to an interval that a user has watched that is represented by a seek-able and play-able thumbnail (discussed in details in Section 4.3). These designs try to present and communicate the viewing history in a clear and comprehensible way to users so that they do not hinder the navigation and search tasks. We have shown that our designs outperformed the state-of-the-practice approaches in a search task for previously seen events in videos. Using these designs enables us to check how they can help utilize and manage users' viewing histories, and evaluate the benefits this brings. They help us to satisfy requirements 2, 3, 4, 5, and 6 of the interface specified earlier.

Going back to the scenario about Tom, video navigation is another issue that needs to be tackled. Navigating a video using its timeline to find specific content

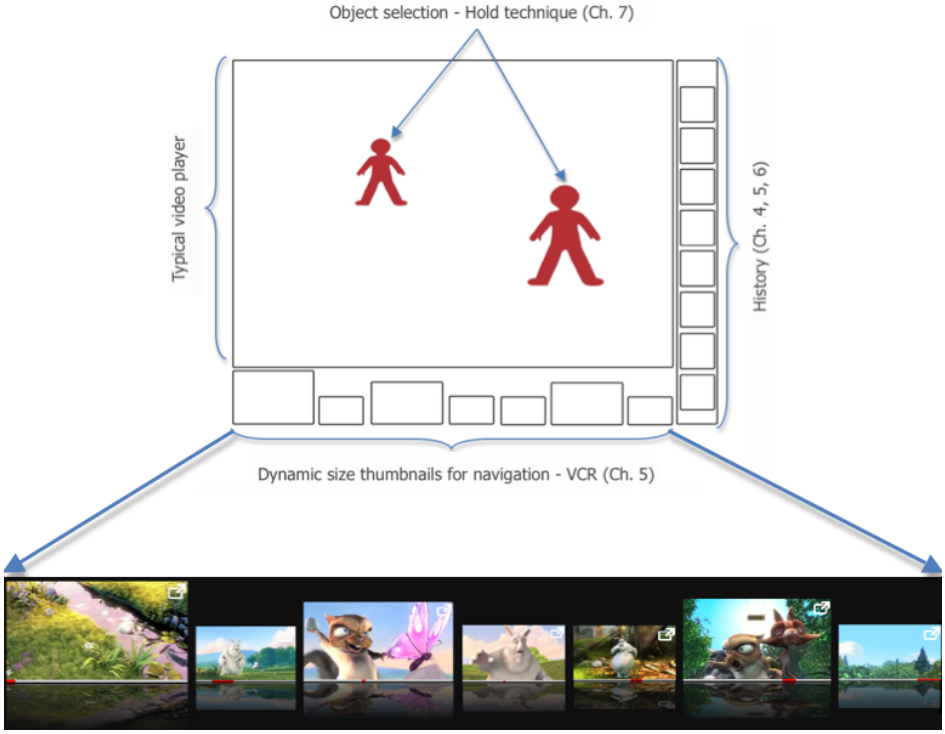


Figure 1.4: The VCR component in the video interface is used for navigating the playing video.

can be demanding and time consuming. Researchers have proposed different interaction techniques to alleviate this problem allowing users quickly navigate and search video content as discussed in Chapter 2. Our interface adopts users' viewing heuristics to improve video navigation. This leads to the design of the VCR component shown in Figure 1.4, which is detailed in Chapter 5. Thus, viewers can use either the history or VCR components to navigate, find previously seen content and summarize the video based on their viewing heuristics; hence, it fulfills requirements 3, 4, 5 and 6. Our evaluation results have shown that a user's viewing history provides quick navigation and fast search tools for previously seen content.

With the diverse selection of videos now available for users, navigation extends beyond the use of the controls available in a standard video player. Other forms of navigation have emerged that have introduced new challenges for users. For example, one of the forms of video that has emerged and is widely used in market-

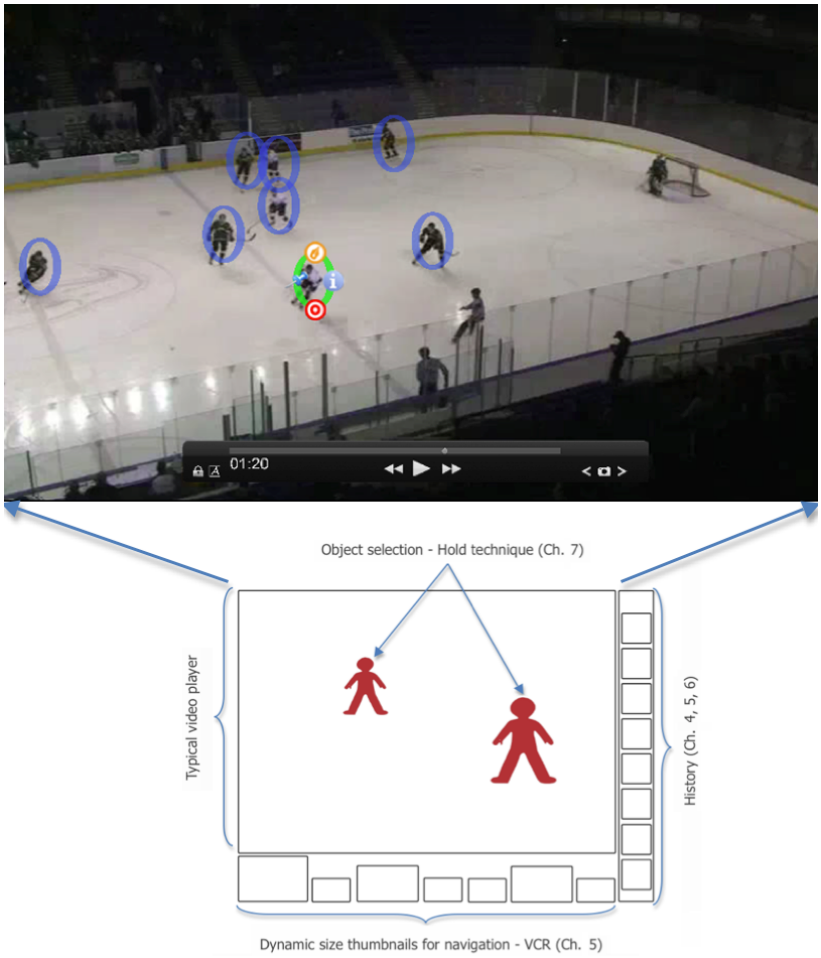


Figure 1.5: The *Hold* technique in the video interface is used for selecting and activating interactive objects for navigation.

ing and education is interactive video where interactive spots or annotations within videos are introduced. Clicking on these spots or annotations directs viewers to another video or piece of information where they can engage and spend more time on that specific information. However, due to the time-based nature of these videos, the embedded clickable objects are visible or active only for a certain duration of the video in contrast to web pages, in which hyperlinks are present at all times. Therefore, the activation and selection of these hotspots becomes difficult. This

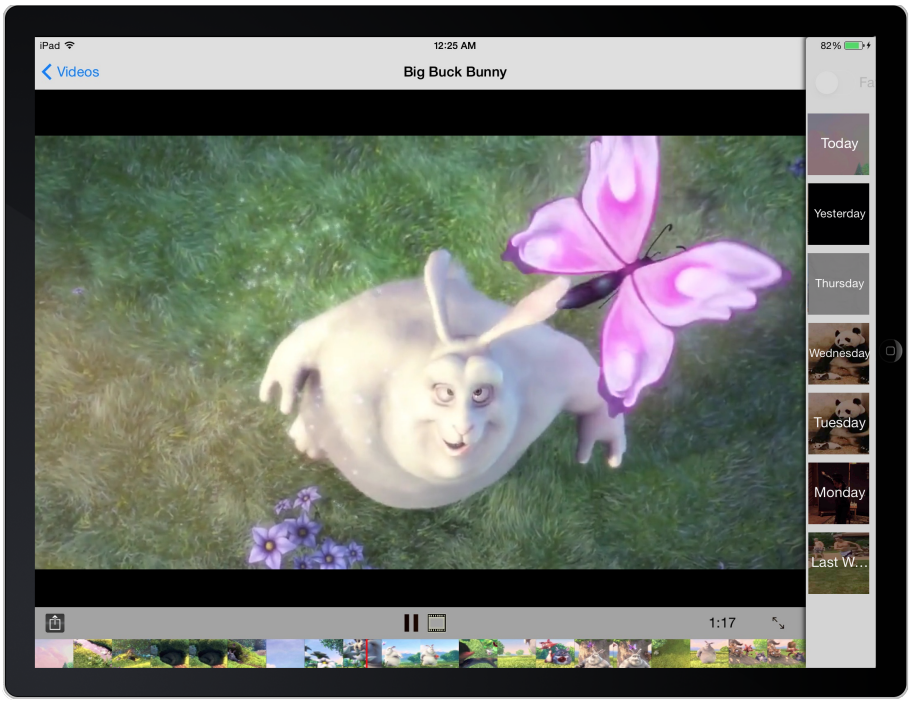


Figure 1.6: The history-based video viewer application with some content.

can also be seen from Tom's scenario where he tried to click on the team members to retrieve more information. To alleviate the navigation problem in this kind of video, we have looked at the factors that affect the selection of these hotspots to allow users easily navigate the video content and hence satisfy requirements 3 and 7. Therefore, as our third main contribution in this dissertation, in Chapter 7 we derive and validate a new mathematical model that estimates the time needed to select a moving object based on its size, speed, direction, and angle of movement. Based on this model, a novel target acquisition technique called *Hold* is developed to ease the problem of selecting moving objects. We have verified that *Hold* outperforms the traditional technique of chasing the object to be selected and showed how it can be used as context switch that enables multiple functions in a multi-stream rich media sport videos interface (*MediaDiver*) as shown in Figure 1.5.

Based on the guidelines we have created and the various interface elements we have evaluated, this dissertation has proposed a version of the next generation

video interfaces that overcomes some of the challenges discussed earlier as shown in Figure 1.1. Our version of a video interface that has brought together all the design elements except for the selection technique *Hold* is called *Mevie* (Figure 1.6). Chapter 8 describes *Mevie* in detail, including, how the designs were integrated into a mobile platform. It helped us to evaluate the components when put together in one interface and to explore people's reactions when these elements are introduced on a video interface. The results revealed that users welcomed the changes brought into the interface and started thinking differently about videos and the way they could be used. We started to see how a change in the video interface introduces various potential applications and usages of videos.

1.1 Contributions

This dissertation provides three main contributions to the research domain of Human Computer Interaction (HCI). These contributions include studying users' video viewing behaviour(Chapter 3), creating and validating some design strategies for future video interfaces to match how users view and interact with video (Chapters 4, 5, 6, and 8), and developing novel selection technique for a moving target (Chapter 7). Publications resulting from this work are listed in Appendix A and the main contributions are summarized here.

Contribution 1: Behavioural analysis and characterization of how users interact with online videos. We developed a Google Chrome extension that allowed us to collect users' viewing behaviour while watching videos on YouTube on a desktop platform. A five-month collection of traces was analyzed to characterize how people navigate and view videos and to determine their viewing behaviour. Based on our sample (n = 19), we showed that people are actively watching videos and demonstrated that people re-watch all video types equally and they do it often. The data revealed that when a user accesses a video a second time, it is mostly to refer back or re-watch something that has been previously seen and not to be resumed from where it was left. We also showed that the drop-off has little to no correlation with the video length and popularity, indicating the subjectivity of users' interest in the content. We found that most participants were re-watchers (i.e. watch portions of a video multiple times), skippers (i.e. jump around a video to find specific

information or to pass over irrelevant and uninteresting parts) or both.

Contribution 2: Creating and validating some design strategies for a history-based video viewing interface to enable users to more quickly navigate and find moments from previously seen videos. We designed and implemented a video viewer interface that keeps track of users' viewed portions from each video and provides access to this data. In order for users to be able to access this data, different visualizations for history were designed and evaluated. By applying users' viewing history we were able to provide a fast navigation and search tool. We have solved it in these contexts:

i. **For single video history.** We used a chronological ordered list of thumbnails (Figure 1.3(c)) as a visualization of the viewed portions within a video, illustrated in Section 5.1. Using this visualization we showed that participants were faster in finding answers to specific questions from previously seen content.

ii. **For multiple video history.** We designed two different visualizations for history: *Video Timeline* (Figure 1.3(d)) and *Video Tiles* (Figure 1.3(e)), detailed in (Chapter 6). Searching for previously seen events was faster using both designs in comparison to the state-of-the-practice method.

iii. **For visualization of the history in a filmstrip.** We employed the viewing statistics of a video to construct a histogram-based filmstrip we call *VCR* (Section 5.2). We showed that searching for specific events in a previously seen video using the *VCR* (Figure 1.3(b)) outperformed the state-of-the-practice filmstrip (Figure 1.3(a)).

Contribution 3: Developing a novel selection technique and formulating a mathematical model for moving target acquisition. To alleviate the burden of selecting interactive objects in the new form of videos, we modeled the time needed to select a moving object, and based on this model a novel acquisition technique was developed.

i. **Formulating a mathematical model.** We derived a new model to estimate the time needed to select a target moving in a 2D environment based on the target size, speed, direction and angle of movement.

ii. **Validating the model.** An experiment was designed to validate the proposed model for a moving target. The data showed a very good correlation with the model, verifying its validity.

iii. **Applying the selection technique in a working interface.** To test the feasibility of our proposed selection technique (*Hold*), we designed and implemented an interactive interface, called *MediaDiver* for experiencing, viewing and annotating complex video domains and its associated metadata content. It enables viewers to interact, browse, explore and annotate (i.e. tag) a multi-stream sport video.

1.2 Dissertation Outline

This dissertation establishes the research needed to meet many of the requirements for the next generation of video interfaces. Thus, it is structured around the main contributions made in this dissertation. We begin in Chapter 2 by reviewing what has been done in the literature and how researchers tackled the different issues. Chapter 3 details our behavioural analysis on how users view and interact with online videos. Our first prototype of a history-based interface and the evaluation of the usability of the interface for creating short trailers are described in detail in Chapter 4. Based on the findings from the evaluation, some modifications have been applied to the interface, which are described in Chapter 5. Chapter 5 also presents two different visualizations for a single video viewing history and details the studies conducted to evaluate these visualizations in a search task in comparison to the state-of-the-practice method. Chapter 6 then describes new modifications to the interface, details two new visualizations for multiple-video history, and presents the evaluation of the performance of these two designs in a search task. In an effort to reach a large audience, we have integrated our concept and designs into a mobile application. A detailed description of this application with its evaluation is presented in Chapter 8. In Chapter 7, we describe another form of video navigation and tackle the selection problem of moving objects in interactive videos. A new mathematical model is derived in Chapter 7, validated and used to propose a new selection technique. Finally in Chapter 9, we summarize the dissertation contributions, describe directions for future work, and provide some concluding remarks.

The appendices provide some additional material. Appendix A lists the publications and interactive demonstrations associated with the dissertation. Appendix B provides a list of questionnaires that were used in each user study in this research.

Chapter 2

Related Work

Most people have experienced videos on different platforms and various social media websites. They use their devices trying to find or explore some videos and they end up viewing, navigating, re-watching, and sharing several videos. Further, to navigate between different videos has been made even easier on social websites such as YouTube. When a person watches a video on these websites, they are offered other videos that he/she can navigate to. These are offered either through a list of recommended videos, or hyperlinks and hotspots within a video that a user can click to be directed to that video and he/she can start viewing it. This new kind of interactivity changes the concept of watching a video sequentially and passively to what is known as interactive videos. In interactive videos, a video is connected to multiple types of media allowing the viewer to have access to additional information when requested. Just as the World Wide Web (WWW) changed the reading and publishing of web pages through the introduction of hyperlinks, video linking changes the linearity of video, offering viewers richer information and more immersive experiences.

Even with a single interactive video, various people end up with different experiences because they could navigate around it using the embedded hotspots. Recording users' navigation experiences using their digital footprints gives them the possibility to navigate back through the video sequences they had already selected, allowing them to re-experience specific sections multiple times, reuse portions of it, save, search or even share it with others. People can have a rewarding

way to enjoy their experience all over again. To offer such a system, we need to understand (1) how people navigate and view videos, (2) how this navigation behaviour and interaction history can be used, (3) how to present this information to users, and (4) how users' experience and tasks performance can be improved. In this chapter we are looking at what has been done in the literature and how researchers tackled these issues. We start by looking at what has been discovered about users' navigation and interaction behaviour while viewing videos (Section 2.1). We then survey the different applications of recording and utilizing users' viewing behaviour in Section 2.2. For users to be able to use their viewing history, the interface needs to provide a visualization that reflects what they have seen and allows them to easily access and use it. Hence, in Section 2.3 we check different visualizations proposed by researchers in the literature. To offer users pleasant experiences while viewing videos, in Section 2.4 we looked at how the navigation within videos can be made quick and effortless, and how to ease the selection of interactive objects within videos in Section 2.5.

2.1 Studying Video Viewing Behaviour

There is a growing body of research and interest in understanding video viewing behaviour which has been largely motivated by the popularity of online videos, interest in the social web, sharing, and the use of videos online. Every second, hundreds of hours of user-generated content are uploaded and millions of people are enjoying this content on different platforms (e.g. TV, mobile, desktop, tablet). A large amount of metadata is left on these videos every time users view them, which provides a rich resource that can be leveraged for viewers. This data can be mined, aggregated and analyzed to understand people's consumption practices, how they experience media, and to develop models and tools that improve users' task performance.

Two resources of metadata have been used in the literature to extract meaning by analysis of the user activity on the video: explicit-user interactions and implicit-user interactions. Explicit-user metadata are collected by asking users to make a specific action around their points of interest such as rating a video, commenting, adding annotation, tweeting, and voting. Users intentionally provide this infor-

mation during/after viewing or experiencing the media. However, implicit-user metadata is any information that is generated by the nature of simply interacting with the media without requiring any additional action from the user. For instance, visiting a video, viewing, and video interaction clickstreams, such as play, pause, skip, replay, or seek/scrub. This interaction data is automatically gathered by the application while users naturally interact with videos.

2.1.1 Explicit-User Metadata

Some researchers have focused on the meaning of activities and behaviours surrounding the social experience (e.g. users' comments, ratings, annotations, remixes and micro-blogs) to understand media semantics, to get some insight into users' behaviour and how to utilize this data. For example, Shaw and Davis [107] showed that users' annotation and retrieval requests can be analyzed and used to develop a better video representation structure. They also demonstrated that the analysis of annotations and the re-use of video segments in re-mixes [108] can be used to understand media semantics. Shamma et al. [106] explored the benefit of the micro-blogs (Twitter activity) in structuring a TV broadcast. They found that the analysis of the comments from the Twitter stream can be used to predict changes in topics in the media event and comments reflect the topics of discussion in the video. Hence, this can be used to create summaries of broadcasts. Olsen and Moon [94] also used explicit user data to generate summaries; however, they applied users' ratings to select segments instead of users' tags or comments. Users' voting was also used by Risko et al. [98] to visualize the important parts within a video. They developed a lecture video annotating tool with which students indicate the importance of a segment from a video by clicking a button. The responses were then aggregated from all students to highlight important parts of the video. Yew et al. [122] also used users' ratings but for a different application. They used the users' rating score of a video to train a Naive Bayes classifier to determine the genre category of a YouTube video. A 75.5% category prediction accuracy was produced using their classifier, which indicates the importance of explicit-user metadata in reflecting the properties of the video content. In this dissertation, we are more interested in the implicit interactions rather than the explicit ones.

2.1.2 Implicit-User Metadata

Implicit-user data can be collected from user navigation behaviour and interaction with the video or from user unconscious physical actions such as eye movements, heart rate and brain neuron reactions that can be gathered with electroencephalography (ECG). Both approaches have been applied in the literature to explore video viewing patterns and usage opportunities.

Users' Physiological Response

Money and Agius [87] looked at users' physiological responses, including electrodermal response (EDR), respiration amplitude (RA), respiration rate (RR), blood volume pulse (BVP), and heart rate (HR), while viewing a video to develop a video summarization technique. Analysis of viewers' facial expressions was also used to provide a promising video summarization tool [65]. Peng et al. [96] analyzed users' behaviour, including eye movements and facial expression, while watching videos to estimate users' interest in the content. Using the interest estimation model, the authors were able to identify interesting parts within a video which could then be easily compiled into video summaries. This approach showed an effective utilization of users' physiological behaviour in video summarization. However, it is not practical since it requires a video camera and additional sensors, which needed to be worn while viewing videos.

Users' Interactions

Users' natural interactions with a video, such as play, pause, skip, seek, replay, and revisit, provide a large amount of metadata that can be examined to understand video watching and navigation patterns. This information can then be used to improve users' task performance and provide insights on how interfaces should be designed. Different researchers have looked at users' interactions within a video to determine the interesting or important parts of the video that can be used later to create representative thumbnails, as a tool for navigation, or as a summarization tool. Despite the excitement, relatively few attempts have been made to explicitly analyze users' implicit interactions to understand how users interact with videos while viewing. Yu et al. [126] analyzed users' logs from a video-on-demand sys-

tem to understand users' behaviour, content access patterns and their implications on the design of media streaming system. They looked at users' daily and weekly access patterns, users' arrival rate, and streaming session length. For the access patterns, they found a correlation between users' work habits and the peaks of the number of users accessing the system. Daily data reached its peaks at noon breaks and after work, while the weekly peaks were reached on Sundays. The average session length was found to be quite short, where 52.55% of the sessions were terminated within the first 10 minutes. The authors explained that this was due to users being uninterested in the content and they scanned the beginning of the videos to quickly determine their interest. This result predicts a design guideline for a streaming system where, in order to serve a minimum of 50% of users' session, the system needs to cache the first 10 minutes of the videos. They also found a negative correlation between session length and video popularity, whereby more popular videos were watched in shorter session times. Work done by Yu et al. demonstrated how often people watch videos, how long they watch, and the relation between popularity and session length. They did not look at the actual interactions within the videos such as play, pause, etc.

Furthermore, Huang et al [59], Hwang et al. [61] and Yin et al. [124] examined users' actions while viewing videos from a video-on-demand system. However, in contrast to Yu et al. [126], they also studied in-video actions such as play, pause, unpause, seek and stop events. Haung et al. [59] analyzed a nine-month trace of data from MSN video service. They studied users' interactions such as pause, resume, fast-forward, and fast-backward. They found that users generally watch large portions of short videos, while less than 20% of users stayed on long videos (i.e. duration more than 30 minutes) and watched more than 60% of these videos. This indicates that most users do not watch videos in their entirety. Moreover, they discovered that 80% of the sessions were watched without any interaction from users. Similar results were also found by Yin et al. [124]. For long videos (i.e. more than 30 minutes long), approximately 40% of sessions had some interactivity from users. However, as the authors mentioned, their data was mostly for videos between 5 and 15 minutes. They did not have any videos of length between 30 to 48 minutes. Likewise, Hwang et al. [61] found that most users only watched a fraction of a video, and they quit the videos before the end. For long videos,

around 50% of videos were abandoned before reaching 40% of the video length, while for shorter videos only 20% of the videos were quit before viewing 40% of the video. Hwang also looked at the correlation between a video's popularity and how much of it was viewed. They discovered that there was little to no correlation between them. On the other hand, Yin et al. [124] studied users' interactions with a video-on-demand system during the 2008 Beijing Olympics. Similar to Yu et al. [126], the peak hours of access was observed after work and more interestingly, the peak days were on the opening ceremony day of the Olympics and the day when a popular Chinese athlete withdrew due to an injury. Yin et al. also looked at how the popularity of videos (i.e. number of views) changes over time and the results showed that in such a multi-day event, popular content changes frequently, whereby the top 5 videos were completely new every day, which was due to the real-time, event-driven nature of the content. The top 5 videos were discovered to be strongly related, whereby these segments belonged to the same logical event. For the video length, they found an inverse correlation between a video length and the viewed percentage of a video where longer videos were dropped-off earlier than shorter ones. This result coincides with Huang's [59] findings. Yin's data also showed a strong correlation between video length and session length for short videos. Overall, viewing time (i.e. session length) was under 600 seconds irrespective of the actual video length. In terms of activity while watching, 80% of the sessions had no user actions (e.g. play, pause, seek) similar to findings by Huang et al. [59]. However, these findings contradict the observations made by Gkonela [44], Chorianopoulos [28], Kim [70], and our reported results in Chapter 3.

Gkonela and Chorianopoulos [28, 44] analyzed users' interactions in a controlled lab experiment for three different types of video, which were Documentary, How-to, and Lecture. To control video playback, they employed three custom buttons: play/pause, GoForward (i.e. skip forward 30 seconds), and GoBackward (i.e. rewind or jump backward 30 seconds). They found that users used skip forward button most of the time (812 out of 1,258 interactions) and they explained that it was due to time pressure of the experiment in which users had to answer questions within 5 minutes of a 10 minutes video length. When trying to find the answers to the provided questions, users were allowed to skip different portions of the video before settling down on a region that contained the intended answer. Looking at

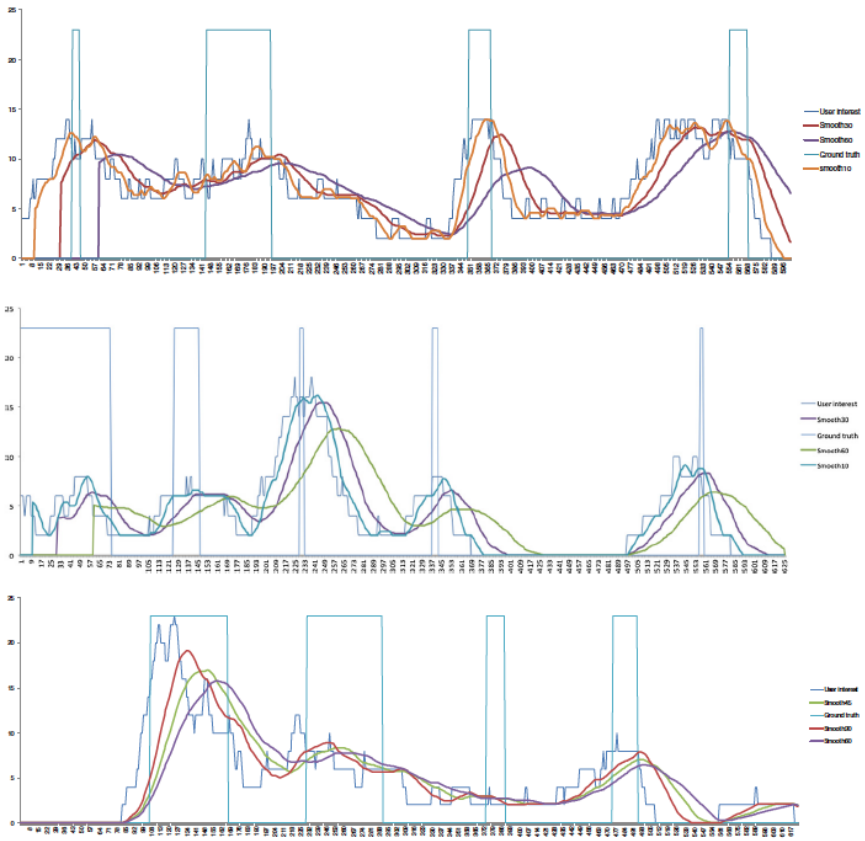


Figure 2.1: Chorianopoulos [28] results of the matching between users' interactions (Replay30) time series and ground truth interest points within videos. Approximately 70% of the interesting segments were observed within 30 seconds before a re-watching local maximum.

the re-watching behaviour, for all videos, they found a high match between the peaks (i.e. local maximum) in the users' re-watching time series graph and the established ground truth. Approximately 70% of the interesting segments were observed within 30 seconds before a re-watching local maximum as shown in Figure 2.1. This work shows interesting observations where users tend to skip videos a lot when trying to find specific information, and they re-watched parts of interest multiple times. However, these results could not be generalized since 1) custom buttons were used that do not allow authors to capture the real natural users' navi-

gation behaviour, 2) the experimental design forces users to exhibit some specific behaviour since they were tasked to find answers for some questions which does not give users the freedom to navigate videos as they would naturally do, and 3) the three types of videos, have a similar purpose to gain knowledge, which does not map the general users' behaviour because users may exhibit different behaviour when they view videos not for a learning purpose.

Li et al. [74] also investigated viewing patterns in a laboratory setting. They investigated the usability of different features of an enhanced video interface and how often participants used these features across six different video contents. Their enhanced interface contained different features to navigate and view videos, such as traditional video controls (e.g. play/pause, fast forward, and rewind), rich indexes for navigation (e.g. table of contents, video shot boundaries, and personal marks), and speeded-up playbacks (e.g. time compression and pause removal). Participants were asked to watch long videos (40 - 60 minutes) in a limited amount of time (30 minutes) using the different features. The results showed that participants applied these features differently based on the video content. For informational audio-centric, video indexing tools such as a table of contents or a list of personal notes provide a valuable tool for fast viewing of the content. However, for informational video-centric videos, the list of thumbnails and shot boundary frames helped participants to quickly view videos. For narrative-entertainment videos, participants preferred to have a fast playback of the video and to jump around commercials. Similar to Gkonela and Chorianopoulos [28, 44], Li et al. [74] used a task that might have forced participants to apply specific patterns while viewing rather than the normal behaviours participants would exhibit naturally.

Another work that explicitly analyzed users' interactions was done by Kim et al. [70], who looked at specific interactions in one category of videos. Kim et al. analyzed students' interactions in educational videos to understand how learners use video content and how that affects their learning experience. They looked at in-video drop-off rate and users' interactions including play and re-watch. They discovered that the drop-off rate increases as the video gets longer. Moreover, they found that students returned and re-watched parts from videos and most of these re-watch actions occurred around parts that are confusing or important (similar to [28, 44]). 61% of the re-watch peaks accompany a visual transition in a video. Based

on the observations around the peaks, Kim et al. identified five categories of users' activity: starting of new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation. Peaks that occur due to following a tutorial step were found to be significantly higher than those occurring for starting new material and non-visual explanation. Kim's work has proven that users actively view videos and they do actually re-watch parts from a video multiple times. Their work provides a good start in understanding video viewing behaviour; however, they have not analyzed other interactions such as skip, replay, and revisit. Moreover, they only looked at educational videos that have a similar purpose as in [28, 44], and their results were based on a specific population (i.e. students). In Chapter 3, we report the results of analyzing the different behaviours while navigating non-specific categories of videos with no specific purpose of viewing and using different sample of users than just students. We demonstrate how these behaviours occur in different video categories and for each user not only for the collective users' behaviour. We also show how the re-watching behaviour exists across all other categories. This contradicts with other researchers' [28, 44, 69] claim of the occurrence of this behaviour in educational and how-to videos only.

2.2 Leveraging Implicit-user Interactions

A large body of work has been going on analyzing users' behaviour logs to be used for a targeted application. This work includes looking at users' interactions and using the wisdom of the crowd to propose new tools for video navigation, summarization, creating representative thumbnails, and design guidelines for viewing interfaces and content producers. Martin and Holtzman [77] emphasized the importance of implicit interactions in developing models to filter media delivered to a user to be more personally relevant and socially grounded content. They used the percentage of a video watched, the amount of time spent viewing, and the amount of time spent interacting with a video to assess the popularity and the relevance of the content being viewed to the users' social networks. These data then can be used to filter and propose other related content to the user based on their own behaviour, preference, and their social peers. Martin and Holtzman focused on news content

and did not explore their model applicability to other types of content that could exhibit different users' behaviour. Yew et al. [123] recognized the importance of using the number of play, pause, fast forward, and rewind events to identify the genre of comedy videos. Feeding this information to their Naive Bayes classifier showed a 6.8% accuracy increase in category prediction when compared to their previous work [122], which used explicit-user metadata. They were able to achieve 82% accuracy using only the collective wisdom (i.e. collective user interactions). This means that predictive models can be made more accurate via the examination of users' viewing behaviour. Yu et al. [126] and Yin et al. [124] discovered the importance of using the amount of video being watched for network providers. They found that the session length for most users was roughly 10 minutes, which they suggested to be used as a guideline for caching mechanisms. Thus, the initial 600 seconds of a video (or of the most viewed segments) would be cached instead of caching large videos entirely.

Some researchers have used the number of views (i.e. revisit actions) a video received to identify quality of content [36], and to predict video popularity [112]. Crane and Sornette [36] used the number of views (i.e. revisit actions) a video received to identify high quality content or videos that attract attention and keep their appeal longer over time. They analyzed 5 million videos posted on YouTube and found that using this view count they could identify high quality videos from junk videos amongst the viral videos within YouTube. Similarly, Szabo and Huberman [112] were able to predict a video's popularity thirty days ahead using view counts on YouTube.

Another group of researchers has looked at the number of re-watch actions within videos to infer video segments of interest or importance, which then can be used for generating a representative video thumbnail [28, 44, 73, 108], to place content on network providers [61], and to make mash-up video summaries [108, 111, 125]. Leftheriotis, Gkonela and Chorianopoulos [28, 44, 73] proposed a thumbnail generation method that is based on the peaks of the re-watching view count. They used the three most popular scenes in a video (i.e. most re-watched) as the proposed thumbnails to represent a video being played. They found that these representative frames matched the segments of interest in a video. Shaw and Schmitz [108] proposed using the number of reuse of segments in re-mixes to select a representative

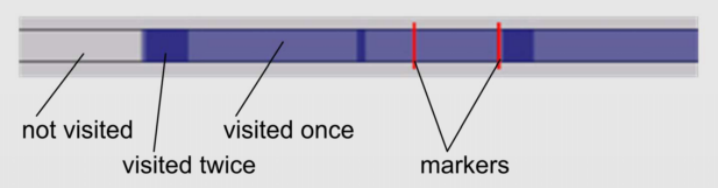


Figure 2.2: Mertens [82] footprint bar visualizing viewing history of a user.

thumbnail. They used the top local maxima of these reuse numbers to select segments that can be combined in a video summary. Hwang et al. [61] looked at how viewing patterns can be applied to place the most viewed video content in provider networks rather than using the complete video to reduce their storage and network utilization. Their approach showed substantial savings in storage and network bandwidth when compared to a simple caching scheme. Moreover, according to Yu et al. [125], using the peaks in the number of re-watch graphs to generate a video summary can offer a shorter path in each video. They found that there are segments of a video clip that are commonly interesting to most users. Syeda-Mahmood and Ponceleon [111] used the number of re-watch actions along with some explicit user interactions where they asked users to provide their sentiment (e.g. bored or interested). Both sets of data were used as states in a Hidden Markov Model to determine the interesting segments from a video. To generate a preview using these segments, Syeda-Mahmood and Ponceleon analyzed the audio track to precisely decide the length of each segment and their start and end time around the peaks, after which they can be automatically combined into a meaningful video preview. When testing these previews as teasers for other users to rate for which preview they are going to watch the entire video, they found that the ratings of the videos to be watched entirely changed after watching these previews. This indicates the effect of the method for creating video previews. All the aforementioned work uses collective data to generate these summaries or previews. Furthermore, we show in Chapter 5 how personal re-watching behaviour can also be used as a tool for generating video previews.

Some have used the viewing history by employing the number of times each frame or second in a video has been viewed (i.e. view count) as a navigation tool for a video [69, 82]. Mertens et al. [82] used users' traces or footprints as an

Figure 2.3: Kim et al. [69] Rollercoaster timeline visualizing viewing history of multiple users (collective wisdom). The height of the timeline at each point shows the amount of navigation activity by learners at that point. The magenta sections are automatically-detected interaction peaks.

overlay over a video timeline as shown in Figure 2.2. They used different brightnesses of a colour to indicate how often each part of a video was viewed, which reflects how a video was consumed. Clicking on any of these highlighted portions on the timeline seeks the video to the corresponding time. Kim et al. [69] applied Mertens' approach, but they visualized the view count as a histogram (or what they call Rollercoaster) as illustrated in Figure 2.3. The height at any point in the timeline indicates how often that part was viewed. Using the height instead of colour brightness makes it easier for users to spot the peaks or where the important parts are located within the video. However, neither visualization tells users the content of these parts without the need to navigate to each specific point. In Chapter 5, we propose a better visualization, which applies the view count (i.e. viewing heuristics) as the basis similar to Mertens et al. [82] and Kim et al. [69].

Carlier et al. [22] and Shamma et al. [104] used another form of implicit interaction. Carlier et al. [22] looked at users' zooming and scrolling actions while watching videos on small screens. By exploiting collective users' wisdom, they re-targeted a high-resolution video for display on small screens as shown in Figure 2.4. They used the selected regions after applying Gaussian Mixture Models, Minimum Covariance Determinant and a re-framing technique to find and stabilize regions of interest from the different frames. Their experimental re-targeted videos automatically produced using the crowd-sourced data were only slightly worse than those produced by hand by an expert. This indicates that even zooming and scrolling interactions provide a potential application based on the trust of the crowd. However, these need a custom interface that offers these features, which is still not available on the commonly known video social websites (e.g. YouTube). On the other hand, Shamma et al. [104] looked at the implicit social sharing activity (e.g. the number of pauses, rewinds, fast forwards and session length) that occurs

Figure 2.4: Carlier el al. [22] employed users' zooming and scrolling actions while watching videos on small screens to re-target a high-resolution video for display on small screens. Their approach overview is shown: four frames and a few viewports (first row), heatmaps and detected regions of interests (ROIs) (second row), and re-targeted frames including re-framing techniques (last row).

while sharing a video in a synchronous sharing tool to predict videos' popularity. Their prediction model reached a 95.8% accuracy using a small training set to predict whether a video would have more than 10 million views. Even though 100 of their total 1,580 videos had over 10 million views, their model was able to correctly predict 81 videos. Their data showed no correlation between YouTube view count and number of times a video was revisited, nor between video length and session duration. However, there was a correlation between session duration and YouTube view count, which might be a dominant feature of the predictive model.

In this dissertation, we are looking at how viewing history can propose fast navigation, quick search, easy sharing, and effortless video authoring. Even though most of the aforementioned research has tackled some of these activities, no one has provided users access to their own viewing history. The data is only available to the researchers. Our approach is to give users access to their viewing history and explore how they can utilize this data and what benefits it brings.

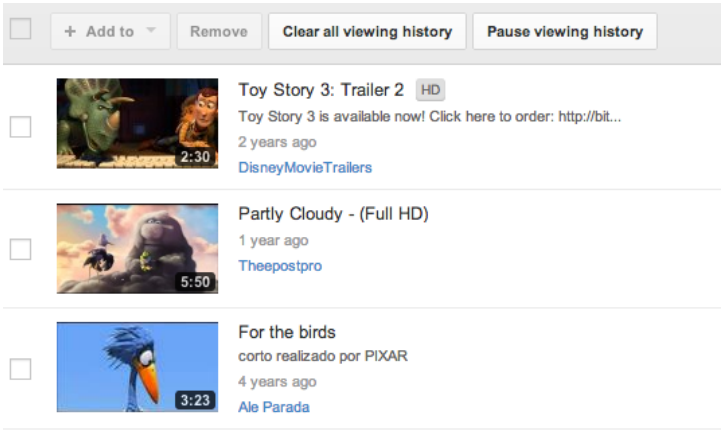


Figure 2.5: *YouTube*'s history is a Timeline-based visualization. It consists of a list of thumbnails ordered chronologically.

2.3 Viewing History Visualization

The human visual system can perceive graphical information such as pictures, videos and charts in parallel; however, text can only be perceived sequentially [55], since the human brain processes visual input earlier than textual input [113]. Nadeem et al. showed that the use of visual aids in history mechanisms is more effective than the use of only textual data [90], thus having a visualization of the video navigation history can help extensively when searching for information. They also demonstrated that the use of history mechanisms may have a significant effect on user satisfaction and performance when revisiting previously viewed content. These findings reveal that it is important to develop and enhance history visualization mechanisms. However, since there is little work on historical video navigation, we are going to explore this from the perspective of web browsing history.

In most web browsers, history is represented as a list of the visited web pages' titles sorted by date, popularity or aggregated by some time period. The history menu opens in a new window where pages are visualized as titles or with thumbnail images. Researchers have tried various visualizations to simplify searching within the history, which can be divided into three categories: timeline-based, graph-based and Three Dimensional (3D) visualizations.

2.3.1 Timeline-based Visualization

In the timeline-based visualization, history consists of a linear scroll-able list of thumbnails, which appear in a reverse chronological order: the most recently visited page is at the top of the list, and clicking on any of these thumbnails (or icons) redirects the user to the corresponding web page. Most web browsers, YouTube, Netflix, Hodgkinson [57] and Girgensohn et al. [43] use this visualization to represent the history of users' visited content as shown in Figure 2.5, where clicking on one of the thumbnails navigates to the corresponding content. Li et al. [75] and Hupp et al. [60] used this approach for their detailed histories with the addition of a list of coloured icons next to each thumbnail describing the users' performed actions. However, this visualization faces problems when multiple tabs or multiple browsers are opened at the same time. Vartiainen et al. [115] developed the Rolling History for mobile devices where they proposed four directions of navigation control. Instead of having one reel of thumbnails, they used two: horizontal and vertical, shown in Figure 2.6. All opened browsers are in the vertical reel; the currently active browser appears in the middle, and its history is visualized in the horizontal reel. Blankenship et al. proposed TabViz[1], which uses a fan-shaped browser tab visualization where concurrent active tabs are represented in a radial hierarchical structure, visualizing the parent of each opened tab (Figure 2.7). Khaksari proposed a Grid as another solution for multiple tabs opened at the same time [68]. The Grid consists of a number of labeled tabs, where each tab corresponds to the relevant tab in the browser. Each vertical column of thumbnails is mapped to the history of corresponding tabs in the background. According to the article, this visualization reduces cognitive workload, increases enjoyability and reduces user frustration.

2.3.2 Graph-based Visualization

Milic-Frayling et al. [83] deduced that high effectiveness during search requires users to have a mental map of both the hierarchical structure and the access sequence of web pages. Using the timeline-based visualization, back-tracking or visiting new content from the currently viewed history item would affect the struc-

[1] http://tabviz.org/

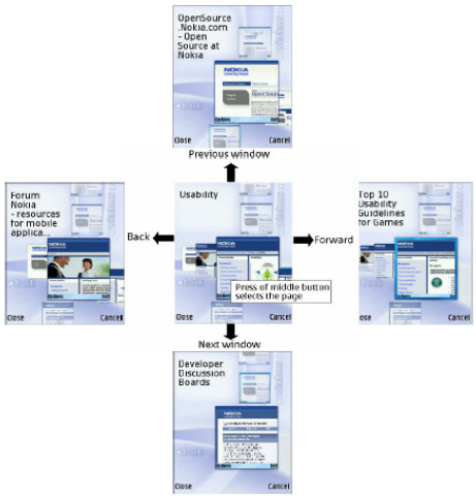


Figure 2.6: *Rolling History* [115] is another Timeline-based visualization, which used four directions of navigation control to cover the history of multiple tabs or browsers opened at the same time.

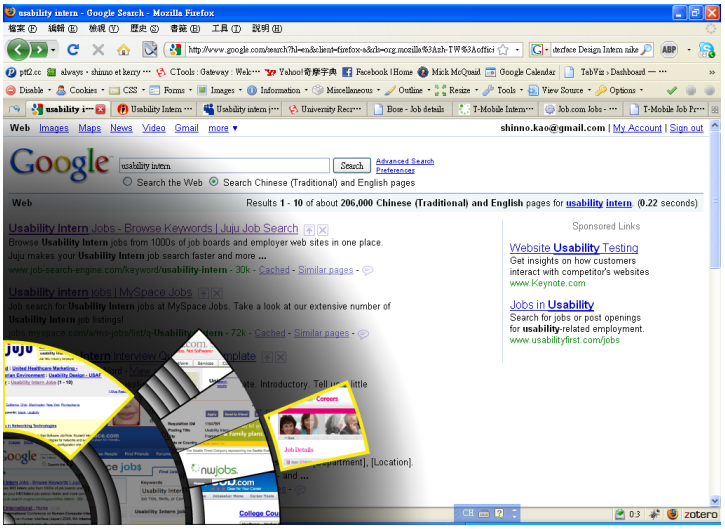


Figure 2.7: *TabViz* is a Timeline-based visualization that employed a fan-shaped hierarchical visualization to show the history of multiple tabs.

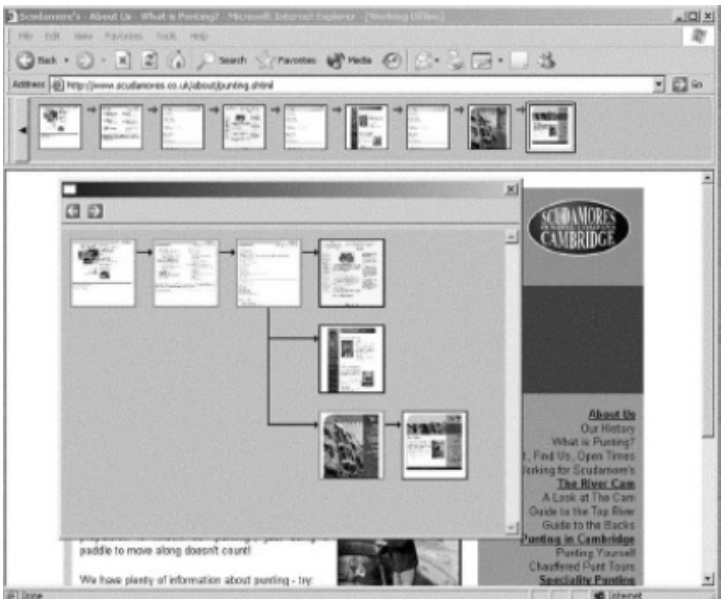


Figure 2.8: *Tree* or directed graph [83] are Graph-based visualizations which visualize each visited page as a node and links as the edges between nodes.

ture of the history, creating confusion and affecting the searching task. This can be solved using a 2D graph-based visualization where the history is presented as a horizontal tree (e.g. [83, 84]). A new branch is generated from the parent node whenever a user back-traces and visits a new page, as shown in Figure 2.8. Mayer [80] used a directed graph where each visited page is a node and the edge between them is the link. Pages that are visited multiple times are visualized using a single node to avoid repetition. Mayer used the size of the node to represent the time spent on the corresponding web page.

2.3.3 3D Visualization

The final category is the use of 3D visualizations for browsing history. Frecon et al. [41] developed WEBPATH, which visualizes the graph of the users' browsing history using a 3D representation. Each visited page is a cube labeled with the page title on top and the surface of the cube shows an extracted image from the HTML description of the page. Users have control over which image can be used

(a) (b) (c)

Figure 2.9: 3D visualization: (a) *WebBook* [120] represented each web page as a traditional book page; (b) *Circle* mode [120] placed thumbnails of the visited web pages at the circumference of a circle; (c) *Cube* mode [120] is a 3D visualization which placed thumbnails of pages on the faces of a cube.

to represent the page. This visualization has not been evaluated to check its usability and performance. Card et al. proposed WebBook [20], which used a book metaphor to aggregate web pages in virtual 3D books. A web page is represented as a traditional page in a WebBook as shown in Figure 2.9(a). However, in terms of search speed, this design might not perform well due to the need for flipping or visiting most of the pages that precede the desired page. Yamaguchi et al. [120] also proposed two other visualizations: a circle mode and a cube mode as shown in Figure 2.9 (b) and (c) respectively. In a circle layout, the thumbnails of web pages are placed around the circumference of a circle while in the cube layout thumbnails are put on the surfaces of a cube. However, it is not clear how scaleable the cube is, since the surface area is limited.

None of the approaches proposed to date have been evaluated or used to visualize a detailed video navigation history aside from our work described in Chapter 5 and Chapter 6. Research within video history (e.g. [43, 84] and YouTube) used the aforementioned approaches only to visualize the previously watched videos but not how these videos were navigated (i.e. intervals). Other researchers [69, 82] visualized the viewing heuristics as footprints over the video timeline. Mertens et al. [82] have used the video timeline itself to visualize users' viewing 'footprints' using different brightness levels, where the more a portion gets viewed the darker that area gets in the timeline(Figure 2.2). It is analogous to more footprints on that area. Using this visualization, the user can easily see which portion of the video has been

watched the most. Similarly, Kim et al. [69] visualized users' viewing heuristics in the timeline; but they used a histogram graph (they call Rollercoaster, shown in Figure 2.3) instead of colour brightness where height represents how often each portion has been viewed. Their visualization shows the intensity and range of each peak making it easier for users to spot the commonly revisited parts in the video. However, using Kim or Mertens' visualization, it is impossible to tell the content of these peaks without seeking to them and checking their content. In Chapter 5 we describe how to visualize a single video history, and in Chapter 6 we demonstrate how this can be extended to visualize a detailed multiple-videos navigation history.

2.4 Video Navigation

Navigating a video space or even long videos can be demanding and time consuming. There have been many interaction techniques proposed in the literature to alleviate this problem in order to quickly navigate and search video content, simplify access and improve efficiency; here, we will only mention a few. Search in videos falls into two categories: unknown-item search and known-item search. Unknown-item search is when users simply explore a video to check what it contains without having a specific goal in mind. However, known-item search is when a user watches or navigates a video to find specific scenes or information. Known-item search can be subdivided into two subcategories: unseen scenes, and previously seen scenes. The search for previously seen-moments needs some knowledge about which parts the user already saw. In this dissertation we are interested in known-item search and more specifically for previously seen parts of a video.

Figure 2.10: The standard navigation tools used in most video systems are the video controls (play, pause, seek, fast forward, and rewind).

Using simple video controls (play, pause, seek, fast forward, and rewind) (Figure 2.10) to navigate and search a video is time consuming and not efficient since they require users to continuously interact with the video timeline and buttons (e.g.

(a) (b)

Figure 2.11: Video navigation using representative previews in (a) Netflix, and (b) YouTube [79]

scrub/seek, fast forward) to find a certain portion or information from the video (i.e. known-item search). For fast search, users need to remember the approximate temporal location of each event. There are many video systems proposed and discussed in the literature that have adopted a variety of interaction tools within their systems to improve video navigation. We are going to mention a few of these that fall within three categories: the use of representative thumbnails for navigation, direct manipulation of the video content, and the utilization of a user's interaction behaviour in a navigation tool.

2.4.1 Navigation using Representative Previews

A simple improvement is to modify the standard video timeline with some visualization cues. Netflix and Hulu address this problem by using a preview thumbnail at the location of the cursor on the video timeline (i.e. temporal location) to ease the search task (see Figure 2.11(a)). The previous problem still exists since the preview appears based on the cursor location and users need to keep hovering over the timeline while monitoring the displayed thumbnail. Simple linear video navigation can be accomplished through representative thumbnails, such that selecting a thumbnail directly positions the main video at the specific time corresponding to that thumbnail. Some researchers [31, 38, 97, 102, 121] applied this approach in DVD systems as well. Filmstrip (e.g. used in some videos in YouTube (Figure 2.11)(b)) also applies this approach, which consists of a strip of equally time-spaced thumbnails. This approach helps users spot the scene they are looking for.

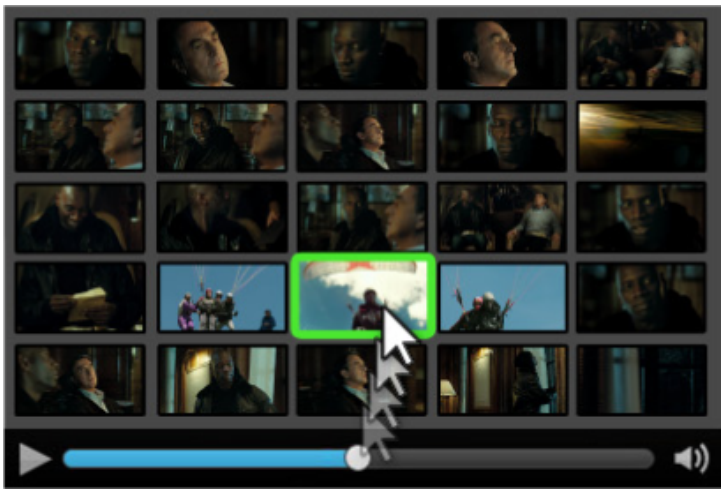


Figure 2.12: Swifter [79]: a video scrubbing technique that displays a grid of pre-cached thumbnails during scrubbing actions

Ramos et al. [97] proposed a navigation tool called a Twist-Lens Slider, which applies a filmstrip with a fish-eye view. They used the pressure on the pen as input to zoom in the filmstrip, which provides a better view of the thumbnails under the pen position and reduces the overlapping between thumbnails. Nevertheless, looking for target scenes of interest, which are not at or nearby these defined thumbnails, could leave users with the same problem as with the traditional video controls if they do not remember its relative temporal location to the available thumbnails.

Matejka et al. [79] tried to solve the continuous scrubbing problem in the search by displaying more thumbnails using a grid instead of a single thumbnail when scrubbing a video (called Swifter, shown in Figure 2.12). Comparing their approach to previous ones applied in Netflix and YouTube, showed significantly better performance where users were 48% faster in locating a scene. Certainly, increasing the number of thumbnails available for users to search from will increase the probability of finding the thumbnail of interest from the grid and the time needed to search the grid. Moreover, having a fixed number of thumbnails in the interface might affect the searching performance based on the video length. In Matejka's approach, users still need to continuously scrub the video timeline to reveal the grid of thumbnails; however, it is less frequent than a single thumb-

Figure 2.13: Panopticon [63, 92]: a video surrogate system that displays multiple sub-sequences in parallel to present a rapid overview of the entire sequence to the user.

nail as in Netflix. Thus, it still needs users to somehow remember roughly where the intended scene is located or they will just have to scrub the entire video timeline. Jackson et al. [63] and Nicholson et al. [92] applied a similar concept to Matejka [79] where they displayed a grid of thumbnails for the complete video without excluding any frames from the video (see Figure 2.13). They achieved this by using video surrogates instead of static thumbnails where each surrogate represents approximately ten seconds (depending on the video length). These surrogates are playing in parallel and each surrogate plays in a loop, to provide users with a rapid overview of the video without the need to seek/scrub. When they compared their navigation tool, Panopticon, against YouTube and a tool similar to Matejka's tool (Swifter), Panopticon was significantly faster in finding answers from unseen videos. Having multiple surrogates playing at the same time overloaded participants with information and it is distracting if the goal was not to search the video but rather enjoying the content of the video. Their results showed that users actually spent significant time looking at the navigation tool rather than the video player itself and this justified the distraction from the video viewing. The authors

Figure 2.14: Cheng et al. [25] Smartplayer helps users rapidly skim through uninteresting content and watch interesting parts at a normal pace.

tried to remove the video player and used Panopticon as a viewer and navigation tool at the same time; however, the parallel playing of the surrogates still was overwhelming and would distract focusing on one specific surrogate. Panopticon was found efficient for searching video but not practical for watching videos. A similar concept to a rapid preview of the video content was proposed by Cheng et al. [25]. They proposed Smartplayer (Figure 2.14), a playback mechanism to help users rapidly skim through uninteresting content and watch interesting parts at a normal pace. However, the determination of the importance or interest of the content is determined by the content-producer, which might not match that of the consumers of the video. SmartPlayer provides an interesting technique to quickly obtain an overview of a video's content, but does not necessarily improve the performance for finding tasks.

2.4.2 Navigation by Direct Manipulation of Content

Since the video content changes as time passes, the separation between the video content and the controls used to manipulate or navigate the video introduces a

Figure 2.15: Kimber et al. [71] Trailblazer allows users to navigate a video by directly controlling (i.e. dragging) objects in the video or on the floor plan.

target acquisition problem. Thus, some researchers have focused in introducing new techniques to manipulate the video timeline rather than using the traditional video controls (e.g. play, pause, fast forward, rewind, and seek). They allow users to navigate the video by directly manipulating the video content along its natural movement path rather than using the timeline slider [37, 45, 67, 71, 97]. Objects in a video, which are detected based on their history of movement, are dragged backward and forward in time to manipulate the video timeline. Kimber et al. [71] developed a direct manipulation by graphing a visualization path for the movement of each tagged object and enabled the user to scrub the video by dragging the object in the video window (Figure 2.15). Similar to Kimber's technique, Karrer et al. introduced the DRAGON interface [67], which provided a mechanism that matches the direction and movement amplitude of an object of interest to the direction and amplitude of a human interface device. It used the optical flow of the video to create the movement trajectory of objects of interest. DRAGON allowed users to have a frame-accurate navigation through the video, which is not provided by the

timeline slider. Dragicevic et al. [37] and Goldman et al. [45] also allowed users to directly manipulate video content to go forth and back through the video in a similar way to Kimber [71]. Nevertheless, this approach does not help users to quickly find certain information or scenes from the video unless they remember its relation to the currently viewed content. Monserrat et al. [88] extended this approach by allowing both dragging and selection of objects in blackboard-style lecture videos. Clicking on an object seeks the video to the starting timestamp of that object and starts playing from that moment; while dragging the object to the left or right performs rewind and fast forward action. Their experiment showed that this was significantly faster than a video player with a single thumbnail pop-up based on cursor location (e.g. similar to Netflix). However, still this approach has the same problem if the intended scene is not temporarily close to the current displayed content or objects.

2.4.3 Navigation Applying Users' Wisdom

People in real scenarios re-watch parts of videos that are important, interesting, affective, or hard to understand [70, 100]. This can be seen most simply in YouTube's or Vimeo's feature for sharing a video from a specific start time, providing an exact use case of what we propose. This watching behaviour leaves digital footprints on the video frames creating a non-flat video histogram emphasizing the interest of each part of the video. Thus it provides a potential tool that can be used to facilitate fast navigation and scenes search. Using this metadata to enhance the tools discussed in Section 2.4.1 can improve users' performance while searching as well as personalize the tool based on users' own behaviour. Instead of using video previews that are generated systematically, personal or the collective wisdom can be applied. In Chapter 5 we demonstrate how using personal wisdom significantly improved the search task within videos.

According to Shamma et al. [105], the more affective a scene is, the more the corresponding interval of the video is viewed or consumed. Thus, the cumulative seamless users' interaction history could be leveraged for the benefit of future viewers (i.e. social navigation). Yu et al. [125] used low-level feature extraction along with users' footprints (i.e. view count) to rank each scene and offer users

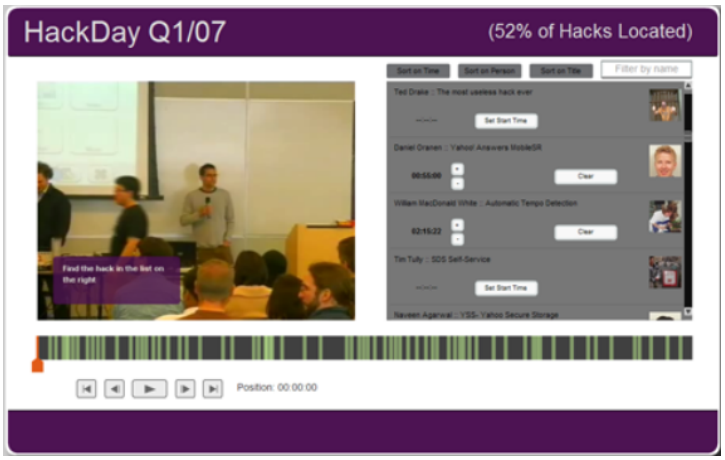


Figure 2.16: Shamma et al. [105] HackDay TV system applies users' footprints on the video timeline to visualize what has been consumed from a video to help users navigate the video. Colour intensity indicates how often each portion has been used in remixes.

other scenes that have some correlation with what they are watching. This provides users with a quick navigation method to similar scenes. Clicking on any of the provided scenes jumps the user to the corresponding time in the video. Mertens et al. [82] have used the video timeline itself to visualize users' viewing footprints using different brightness levels, which lets users quickly navigate to the most viewed scenes (Figure 2.2). Shamma et al. [105] have also applied similar approach on their HackDay TV as shown in Figure 2.16. However, this does not supply a visualization of the video, which inhibits search (when searching for a previously seen event, users need to remember its approximate location). Likewise, Kim et al. [69] used Rollercoaster, shown in Figure 2.3, to visualize users' viewing heuristics in the timeline. Their visualization shows the intensity and range of each peak making it easier for users to spot the commonly revisited parts in the video. They also applied non-linear scrubbing by extending the rubber band analogy to control scrubbing speed and support precise navigation [78, 98]. It is similar to the Smartplayer [25] concept where the important parts got longer exposure than other parts. They applied friction while scrubbing around peaks to slow down scrubbing in these areas, allowing users to get a more comprehensive view of these portions

of the video. Kim's tool faces the same problem as Mertens' [82] where there is no preview of the peaks' content. We show in Chapter 5 how extending this approach by including previews along heuristics improved the search task.

2.5 Object Selection in Videos

Most commercial brands today are using the growth in popularity of online video sites to get their name and message out to a wider population of consumers. However, their video marketing is not what it used to be. Simple video commercial pop-ups are being substituted by introducing interactive spots or annotations within videos. Clicking on these spots direct viewers to the commercial ad where they can engage and spend more time on the brand's marketing video. Many online video sites offer companies a platform for creating this kind of interactivity within videos. For example, YouTube annotations give brands the opportunity to make their own interactive advertisements.

One of the main issues with this form of video is the selection of the interactive links or objects within these videos to traverse to the next piece of information. Due to the time-based nature of these videos, the embedded clickable anchors or annotations are visible or active for only a certain duration of the video in contrast to web pages in which hyperlinks are present at all times. Therefore, the activation and selection of these hotspots becomes difficult and are affected by the shape, size and the location of the object at a given time of the user's selection. For some, selecting a stationary object or a graphical element can be difficult. Due to the moving behavior of the objects in videos and camera panning or zooming, the selection becomes challenging. In some application domains, such as sports or racing, the difficulty is exacerbated where objects or regions are rapidly moving making the selection hard to be achieved. It can become frustrating when the user tries to chase the moving object or when the selection results in mistakenly activating a hyperlink associated with another object. In order to correct this mistake, the user may need to stop the video clip, rewind it to a previous time point, and try to select the object again. Thus, there is a need for methods and techniques that should help the user to select and activate a hotspot associated with objects in interactive videos.

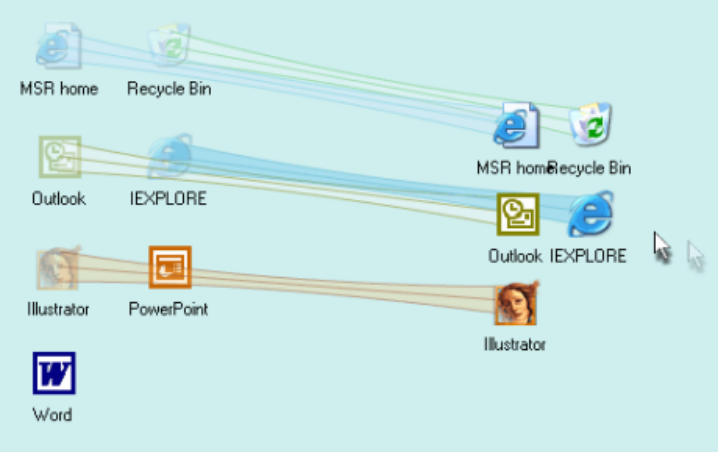


Figure 2.17: Baudisch et al. [13] Drag-and-Pick selection technique.

2.5.1 Static Object Selection

Researchers have proposed interaction techniques to help users select targets by reducing the index of difficulty as implied by Fitts' extended models [39]. One approach consists in decreasing the distance from the cursor to the target (D): this was applied in Drag-and-Pick [13], Object Pointing [50] and Delphian desktop [11]. Drag-and Pick[13] moves potential targets closer to the cursor depending on the cursor's directional movement as shown in Figure 2.17. Guiard et al. designed another technique called Object Pointing [50] where the cursor skips empty space between targets by jumping from one target to another depending on directional movement and the closest target, thus considerably reducing D. Asano et al. furthered the research of Object pointing to Delphian desktop [11] by adding the prediction of a target using the peak velocity and trajectory of the cursor. These techniques are affected by the layout of objects on screen and tend to work best when targets are sparse on the screen. Baudisch et al. [14] took another approach in their Starburst method to reduce D by using empty screen space to increase the effective width and decreasing the distance. This technique optimizes movement time but prevents cursor interaction with the empty space.

Several other methods also focused on modifying the effective width either by increasing the target width or cursor area. McGuffin and Balakrishnan [81]

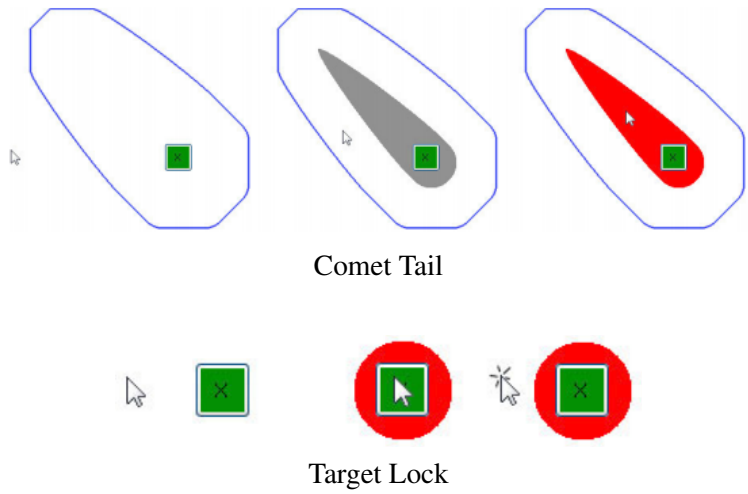


Figure 2.18: Gunn et al. [52] Comet Tail and Target lock selection techniques.

investigated the effectiveness of dynamically increasing the target size as the cursor nears the target. They found users could benefit from the expansion even when it occurred at the last 10% of the movement time towards targets. This effort was furthered by Zhai et al. [127] who showed that users had similar performance even when they could not predict the target expansion. Bubble targets [32], Comet Tail [52] and Target Lock [52] enhanced targets with a bubble or a comet tail as the cursor reached the target instead of expanding the target's actual size (Figure 2.18). Expanding target size out-performs the regular technique for the selection of single isolated targets but they do not perform well with clustered or dense areas of targets as selection ambiguity and visual distraction arise.

Kabbash and Buxton [66] increased the cursor area to increase the effective width. This technique introduced ambiguity when multiple targets were present on screen. Worden et al. [118] furthered Kabbash and Buxton's approach by proposing an additional single hotspot at the center of the cursor that is activated when multiple targets are present under the cursor to alleviate ambiguity. Other techniques such as Bubble cursor [47] and Starburst [14] tried to maximize the activation area (effective width) of each target by partitioning the empty space effectively. The Bubble cursor, shown in Figure 2.19, could cause visual distraction as the cur-

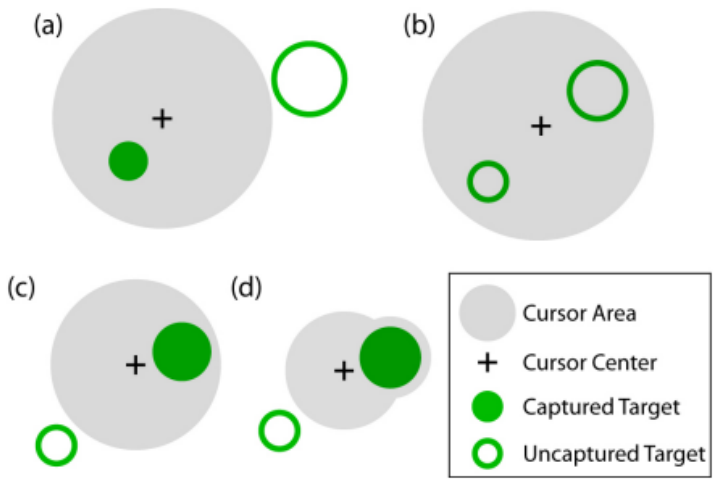


Figure 2.19: Grossman et al. [47] Bubble cursor selection technique. (a) Area cursors ease selection with larger hotspots than point cursors. (b) Isolating the intended target is difficult when the area cursor encompasses multiple possible targets. (c) The bubble cursor solves the problem in (b) by changing its size dynamically such that only the target closest to the cursor center is selected. (d) The bubble cursor morphs to encompass a target when the basic circular cursor cannot completely do so without intersecting a neighboring target.

sor expands/shrinks unpredictably, especially with moving targets. Chapuis et al. applied the expansion of the cursor area in relation to the cursor moving speed in a technique called DynaSpot [24]; however, this technique does not address the speed of the targets.

There also have been some efforts [16, 32, 117, 118] to improve the target acquisition time by changing the control-display (*CD*) gain (the ratio between distance moved by the physical input device and distance moved by the visual cursor). By increasing the *CD* gain in the empty space and decreasing it when the cursor is reaching or over targets, the motor space D/W ratio is decreased. Semantic pointing [16] and Sticky icons [32, 118] used this technique and showed that they out-perform regular selection. However, they faced problems when distractors are present, as this reduces cursor movement time degrading performance compared to regular pointing. The angle mouse [117] handled a dense cluster of targets and

was able to avoid distractors. However, as stated in the paper, it only improved the performance of people with motor impairments.

2.5.2 Moving Object Selection

There has been little research done in this area to investigate and propose new techniques that allow users to easily select moving objects. Mould and Gutwin [89] suggested using target feedback to assist users in targeting moving objects. In their proposal, the object is highlighted once the cursor is over it, indicating that by clicking, the user can successfully select that object. Their results showed that this method helped in reducing the error rate. Nonetheless, the feedback from their participants indicated that it causes visual distraction especially when objects are close to each other, leading to continual highlighting change. Furthermore, the time needed to select the object did not improve by using this technique as illustrated in their experimental results.

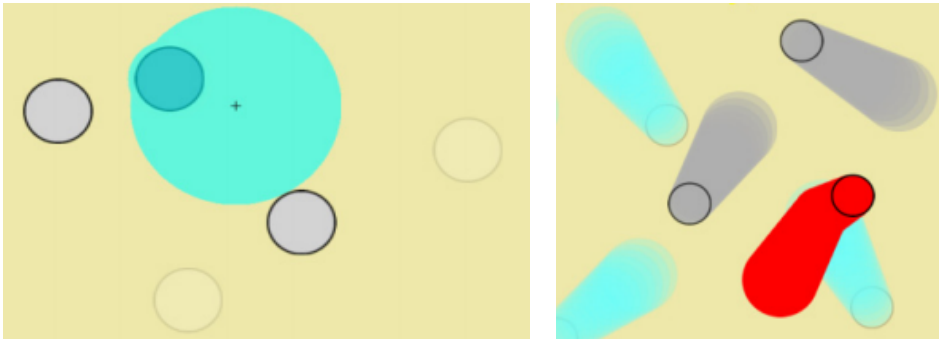
Comet Tails and Target Lock, shown in Figure 2.18, are two other techniques that were introduced by Gunn et al. [52] to alleviate the difficulty of selecting moving objects. The Comet Tails technique works by providing a larger sensitive area behind the object on which the user can click to select that object. In the Target Lock technique, when the cursor moves or is placed over an object, the object will be highlighted and remain highlighted even when the cursor moves outside the object region (trigger the lock effect). Clicking anywhere would complete the selection of that object. In order to select another object using the Target Lock technique, the user needs to move the cursor over the new object and it will be locked. These techniques showed promising results in terms of time and error rate when compared to the chase technique. However, as discussed by the authors, the Target Lock resulted in erroneous selections due to the movement of non-target objects under the cursor. This could become even worse when objects are moving too fast because when the cursor is triggered over any of these objects, they will get highlighted. This continual alternation of highlighting the objects without users' control can cause frustration and visual distraction.

In relation to Hypervideo, to avoid disturbing the video being played, some researchers tried to present the links or hotspots outside the video player area.

In HyperCafe [101, 102], hyperlinks were displayed as pictures or text outside the video window enabling users to select them without the struggle of chasing those hyperlinks. Hyper-Hitchcock [43] also provided the same approach used in HyperCafe where the links were placed on the timeline. Even though watching the video was not disturbed by the links, the user's attention was continually drawn from the story presented on the Hypervideo because they had to observe the area where the links are shown.

Similar to the Target Lock technique [52] described earlier, Sundstrom [110] proposed an approach where the cursor is automatically linked to a hotspot in a Hypervideo stream. When the user places the cursor within a sensitive region, the cursor is linked or locked to that hotspot and will automatically track it. The decision to activate the object associated with the hotspot is left to the user, and they will not need to manually chase that object over time. The user will be assured that they have activated the intended hotspot even when it is completely or partially overlapped by other objects or it is out of the scene. In the case of overlapping hotspots, the system provides three methods to determine which object is to be linked to the cursor. The cursor can either be linked to the object with the closest center to the cursor, linked to the one on top, or not linked. This approach may have some potential applications as described by the author; however, the time and the precision needed to position the cursor over the intended object remain as challenges. Moreover, the ability to activate the hotspot even when it is out of the scene could cause some confusion.

In order to mitigate the challenge of selecting moving objects without introducing any distraction for the user, we proposed a novel selection technique called Hold (described in Chapter 7) which temporarily pauses the content while selection is in progress to provide a static target. It helps the user to select objects while watching a video without the need to use a separate pause button each time they need to do so or to chase a moving object. The Hold technique works as follows: when a user clicks the mouse button down, the moving objects temporarily pause while the user interacts with objects. When they release the button, the objects start moving again. This method was evaluated against the classical chase technique and the results have shown that it outperforms the latter for small and fast moving objects in terms of time needed to select the object and the error rate.



Target Ghost *Comet Ghost*

Figure 2.20: Hasan et al. [53] Comet and Target Ghost moving target selection techniques.

Hasan et al. [53] proposed two techniques for the selection of moving objects: Comet and Target Ghost (Figure 2.20). Their Comet technique is similar to the one proposed by Gunn [52] where they changed the model in which the size of the comet is determined by the speed and the size of the target. The Target Ghost; however, has the selection of static object concept similar to our Hold technique described in Chapter 7. By clicking on the shift key, a stationary ghost or a proxy of the objects will be displayed while the video will continue playing. This allows the user to select while not missing any information from the video. The results on One Dimensional (1D) targeting showed that Comet outperformed other techniques in selecting moving objects, while the Target Ghost did not improve the selection. In the Two Dimensional (2D) targeting, the Target Ghost outperformed all the selection techniques that had been tested. The authors mentioned that even with these good achievements, the two techniques have limitations that need to be taken into consideration. With a large number of objects on the scene, both Comet and Target Ghost could increase the clutter, resulting in visual distraction and selection degradation.

Our Hold technique showed promising results in reducing the time needed to select moving objects. It introduced a variety of metaphors for selection when applying the Hold metaphor that can be left for future work. Hold is one of the major contributions of this research work and to the HCI community.

2.6 Summary

To summarize, few researchers have examined users' behaviour while viewing videos [28, 44, 59, 61, 70, 124, 126]. They have looked at this from a collective perspective not at the individual level. Most of this research is either category specific [28, 44, 70], population specific [28, 44, 59, 70], or event specific [124]. The different aspects that have been analyzed are drop-off rate [59, 61, 70, 70, 124, 126], access peaks [124, 126], session length [124, 126], re-watching peaks [28, 44, 70], number of skips [28, 44], and whether any interactions occurred while viewing [59, 61]. None of the research has examined all of these aspects together, which can give different insights into users' behaviour. The majority of the research on video viewing has looked at how to utilize this metadata. Researchers have considered utilizing users' viewing history for filtering [77], caching [124], categorization [123], summarization [108, 111, 125], navigation [69, 82], and popularity prediction [104, 112]. However, most of these tools are not available for consumers or individuals. Researchers investigated these areas to provide content producers with new tools to define, place and deliver content to users. To be able to give users access to this information, mechanisms to visualize and manage this data need to be defined. To our knowledge, there is no research done in this area aside from overlaying the time series graph of the re-watching view count on the video timeline [69, 82].

The related works show the promise of utilizing what users have seen in the past for the design of new tools and features within the video space. We have identified three areas to contribute to this thread of research: (1) a thorough analysis of viewing behaviour considering the different aspects mentioned in the literature and introducing others to comprehensively understand users' behavior and how these behaviours correlate; (2) a history-based video viewing interface that allows users to view and quickly navigate videos, offers users access to and management of their metadata; (3) a new technique to ease the selection of objects within videos for faster activation or manipulation.

Chapter 3

Video Viewing Behaviour

Nowadays users can easily access hundreds of videos with just a simple click. They can enjoy and experience videos from different platforms and on various devices. As users view and navigate various videos, a substantial amount of digital footprints are left on these videos, which can be turned into useful information [29] and provide a rich resource that can be leveraged for viewers. Nonetheless, this data is not accessible because online video platforms do not share it. Researchers [29, 43, 70] have designed their own plug-ins or systems to capture users' interactions with a video player which are then mined, aggregated and analyzed to understand people's consumption practices, how they experience media, and to develop models and tools that improve users' task performance.

Although previous studies [28, 44, 59, 61, 70, 124, 126] have analyzed different sets of interactions within videos, to the best of our knowledge, nobody has done studies on general (i.e. any category or domain-independent) videos looking at how users' behaviour changed from traditional watching where users watch videos once, passively and sequentially. For example, Kim et al. [70] studied learners' (i.e. students) behaviour in an education scenario, and Chorianopoulos et al. [29, 44] examined instructional videos. While they found certain behaviours, for example re-watching exists and how to apply it later but they did not look at other types of video (e.g. comedy, sports, etc) and domain-independent users. In contrast, we are looking at the different categories available in YouTube. We are looking at as many of the possible behaviours we can just from tracking users' interactions. We

are not actively changing someone's behaviour; we are just looking at what they are doing on YouTube using a very simple plug-in (described in Section 3.3.1) on a desktop platform. YouTube was chosen to study users' viewing behaviour since it is the most popular video streaming site (60% of all online videos are watched through YouTube), it hosts videos from various categories and more than 4 billion videos are viewed daily on it[1].

In [69, 70], it has been suggested that re-watching exists only in the category of education and instructional videos. Our data suggests otherwise. Our participants exhibit these behaviours across all video types and in Section 3.7.2 the analysis confirms this. Others [59, 124] claimed that users mostly watch videos passively and do not interact with video content; however, in Section 3.7.6 we show that more than 65% of our collected videos were actively watched with frequent interaction from participants. Moreover, none of the previous work has looked at individual user's behaviour. They were mostly looking at the aggregation of all users' data. In this chapter we show the behaviours exhibited by each participant and across all video types.

In comparison with previous work, our study (1) analyzes general videos; (2) analyzes behaviours at an individual level; (3) analyzes behaviours in each video category; and (4) provides personal viewing pattern based on exhibited behaviours. The important results came from: (1) analysis in (Section 3.7.6) which shows inclusively that participants are actively watching videos; (2) analysis in (Section 3.7.2) which shows that participants re-watch all video types equally and they do it often; (3) analysis in (Section 3.7.4) which demonstrates that when participants accessed videos the next time it is mostly to refer or re-watch something that has been previously seen and not to be resumed from where it was left; (4) analyses in (Sections 3.8 and 3.9) which show that the drop-off has little to no correlation with the video length and popularity, indicating it is subjective with users being uninterested in the content; and (5) the characterization of the individual personal behaviours in Section 3.6.

This chapter starts by describing different interactions logged in our user study in Section 3.1. The types of behaviour are then described in Section 3.2, fol-

[1] http://www.statisticbrain.com/youtube-statistics/

lowed by our user study design in Section 3.3. For the study findings, Section 3.6 discusses the different categorizations being identified for individual viewing behaviours. Then, sections 3.7 to 3.10 provide detailed analysis of each behaviour, which helped in identifying these sets of behavioural groupings. Finally, Section 3.12 and Section 3.13 discuss the limitations and the directions for future research based on the observed findings.

3.1 Logged Video Interactions

There is a defined set of functions users can apply to interact with video content in a familiar video player. These interactions allow users to control the playback of a video. We are interested in analyzing how users watch any general video by recording the following actions:

Play: When a user starts playback in any part of a video.

Pause: Every time a user pauses or stops the video playback.

Seek/Scrub: Every time a user jumps to any specific time in the video and continues the playback.

Change Video: Each time a different video is being played.

These actions allow us to capture which videos a user has watched and more precisely which parts of a certain video have been viewed. They also show how a user has navigated each video, which can then be used to characterize their viewing pattern.

3.2 Types of Viewing Behaviours

The goal of this chapter is to present the results of users' watching patterns and how often they occur so that we can characterize their viewing activity. Various viewing behaviours may emerge from users while navigating and watching videos as indicated in Chapter 2. In this section, we define six behaviours and interactions that we are going to use and analyze in this chapter. These behaviours are determined based on the video interactions defined in Section 3.1. The six behaviours analyzed are as follows:

Skip: Any time a user seeks and starts watching from a point in a video that has not been seen before, as shown in steps 2 and 3 of the skip behaviour in Figure 3.1.

Re-watch: Any time a user watches a portion of a video that has been seen before. It can be either explicit or implicit re-watch.

Explicit Re-watch: Any time a user seeks and starts watching from a point that has been seen before (e.g. rewind) (step 2 of the explicit re-watch in Figure 3.1).

Implicit Re-watch: Any time a user seeks and starts watching from a point that has not been seen before but continues watching and encounters some parts that have been seen before, as illustrated in step 3 of the implicit re-watch in Figure 3.1.

Drop-off: Within a single session, if a video is abandoned before the end or is not watched entirely.

Uninterrupted: If a video is viewed sequentially and entirely in a single session without any actions from the user (e.g. skip, rewind in the middle, or drop-off), as shown in the uninterrupted behaviour in Figure 3.1.

Replay: If a video is watched entirely without interruption for the nth time in the same session where $n > 1$ (step 2 of the replay behaviour in Figure 3.1).

Revisit: If a video is visited again in another session.

Some of these behaviours require a definition of a *session*, specifically how long a gap is required between activity records before a new session is started. Session length varies between users, as it depends on how long a user spends watching videos, the number of videos watched, the length of the viewed videos and when a user opens and closes the YouTube page. Yin et al. [124] define a session length as the time from when the user hits the play button for a single video until they click on the stop button and subtracting any paused time. Almeida et al. [10] classify any requests from the same media file to correspond to a single session, while Hwang et al. [61] consider consecutive requests for the same video that are less than T

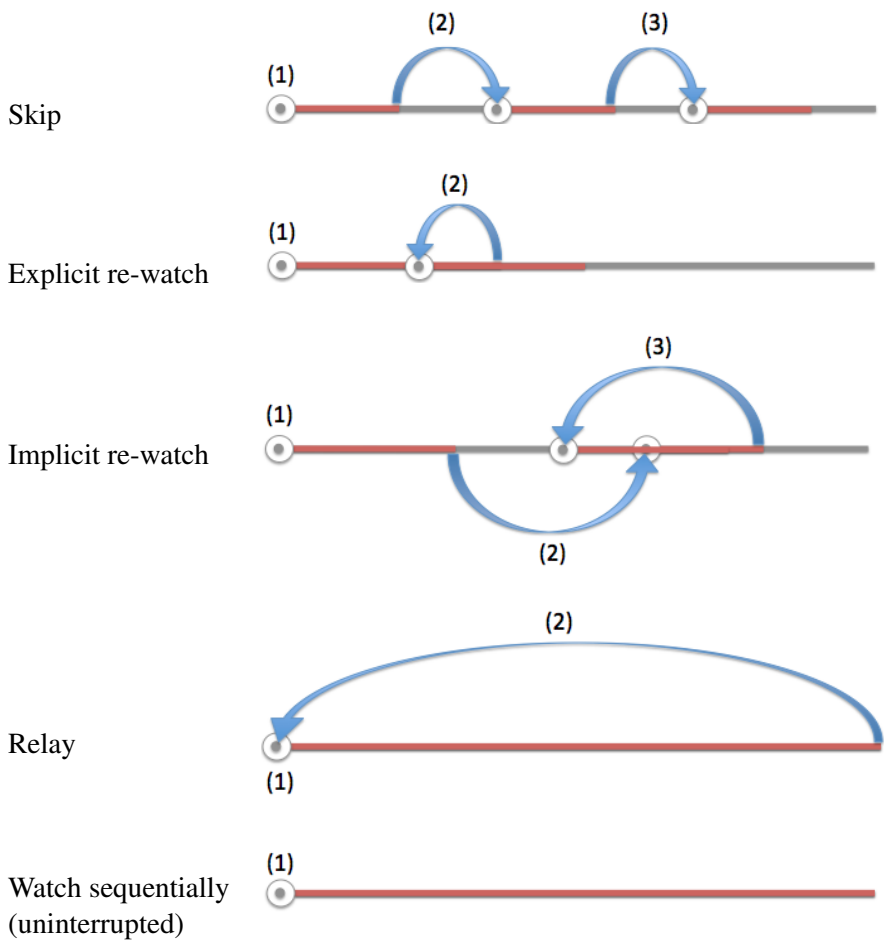


Figure 3.1: Illustration of video viewing behaviours: the white circles represent the start position of a user's viewing interval; grey lines represent the video timeline; red lines indicate the duration and temporal location watched; and the blue arrows indicate an action taken by the user to seek to another time. The numbers indicate the order of user actions.

apart to be in the same session, otherwise it is treated as a new session. Also, if a different video is requested then it is a new session. T was determined based on the normalized length viewed (NLV) of a video and they found that the change in NLV decreases when $T = 1$ hour and $T = 4$ hours. They used $T = 4$ hours since the number of sessions beyond that was negligible in their data. In web browsing,

Catledge and Pitkow [23] measured the time between all events for all users and found that a lapse of 25.5 minutes or greater can be used to indicate the beginning of a new session. Krishnan et al. [72] used 30 minutes of inactivity to indicate the end of one session and the beginning of the next, following the standard notion of a visit/session in web analytics. Our definition of a session is similar to [23, 61, 72] where we measured the inactivity period between each record for all records across users to determine a session boundaries. Based on the collected data we found that 80% of the records had less than 10 minutes of inactivity from their consecutive records, as shown in Figure 3.2. Thus, in this chapter we consider two consecutive records in the same session if the inactive time between them is within 10 minutes, otherwise these two records belong to different sessions.

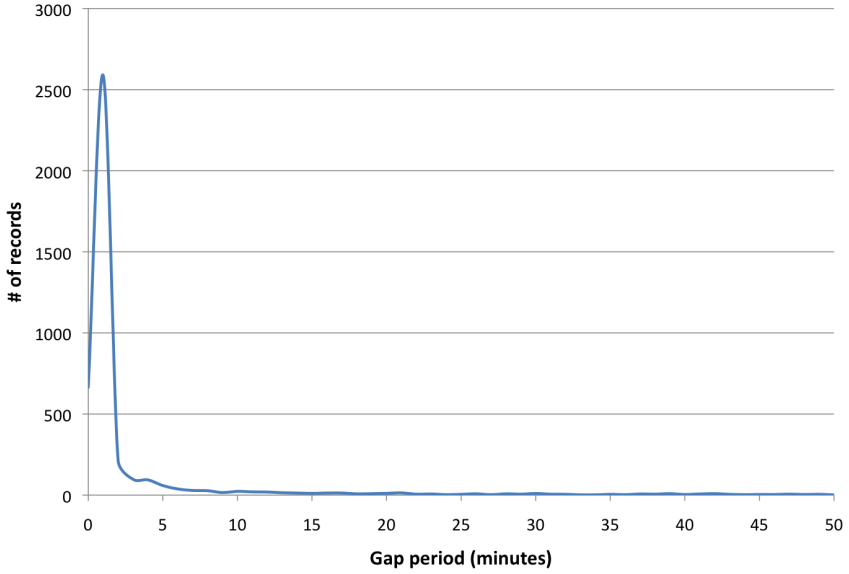


Figure 3.2: Frequency of each inactivity period between records. 80% of records have less than 10 minutes gap from their consecutive record. Beyond 50 minutes the frequency at each time point is less than 8 records.

3.3 Study Design

In order to characterize how people navigate and view online videos, we need to obtain information about users and their interactions with viewed videos. To do so, we developed an extension for Google Chrome, described in Section 3.3.1, that gathers and keeps track of users' interaction in YouTube such as play, pause or seek actions, videos being viewed, and intervals being watched from each video. We wanted to characterize users' watching patterns without influencing their behaviour by allowing users to watch videos, as they would normally do in their own time and pace while our extension recorded their interactions in the background. These records were then sent to our server to be used for analysis. Using this data, we investigated the following hypotheses and questions:

H1. People do not watch videos passively from start to end.

- Q1. Do people watch videos uninterrupted?
- Q2. Do people skip parts of video?
- Q3. Do people re-watch videos?
  - (i) in a single session or multiple sessions
  - (ii) entire video or intervals
- Q4. How often do they interact with videos while viewing?
- Q5. Is there any correlation between different behaviours (e.g. skip and drop-off)?

H2. Shorter and popular videos will have more users' engagement (i.e. interactions).

- Q1. How often were videos abandoned before the end (i.e. drop-off)?
  - (i) Where in-video does drop-off exist the most?
- Q2. Does video length have any correlation with the user's watching behaviour?
- Q3. Does popularity influence users' watching behaviour?

H3. Music videos will exhibit more replay actions while Education and How-to videos will have more re-watching actions.

Q1. Does the type of category affect users' watching behaviour?

Q2. Does replayed videos come from playlists?

3.3.1 Logging Users' Viewing Behaviour

To log users' viewing behaviour, we developed Video Viewing Behaviour extension (VVB), an extension to the Google Chrome browser that was built using HTML and JavaScript. This extension ran in the background when the user browses the web using Google Chrome. It starts recording once the user starts watching any video in YouTube. Each time the user watches or interacts (i.e. play, pause, seek) with the video being played, *VVB* keeps a record of the URL, video ID, video title, video duration, video category, number of views, start time of the interval being watched, end time of the interval, and the user's local time to identify when the interval was watched. The end time of the interval is being updated every 10 *ms* while the user is watching to avoid losing any information for any unexpected browser failure. It is also recorded when the user pauses, seeks, navigates to another video, navigates away from YouTube, or simply closes the browser. These records are kept in the browser's local storage on the user's machine, which are then sent along with the user's identifier (ID) to our server the next day the user opens the Google Chrome browser. These records are stored in our server for later analysis.

3.3.2 Participation Procedure

For participant recruitment, a webpage[2] was designed to describe the purpose of the study, how the extension works, how to install it, the ethics approval, and the extension files to be installed or verified by users who need to check the code before installation. Emails asking people to participate with a link to the extension's webpage were sent to friends, friends of friends, and student lists at the university. Once a person had installed the extension, they were asked to sign a consent

[2] http://www.openvl.org/MyView/VideoStudy.html

form electronically before the extension can start recording their behaviour. If they consented, they were asked to fill out a form to give us some details about their gender, age category, how often they watch online videos, and how many videos they watch per session, as shown in Section B.1. This data was then sent to our server, which in return sent an ID to the extension at the participant's browser to be used as a unique identification for future data. At this point *VVB* is ready to record the user's behaviour. Participants had the option to opt out from the study at any time by stopping the extension from recording or simply by deleting the extension using chrome://extensions/. This user study was approved by the University of British Columbia Behavioural Research Ethics Board [certificate #: H13-01589].

3.3.3 Participants

Nineteen online volunteers, 11 male and 8 female, participated in this study. Participants ranged in ages from 19 to 40, where nine participants were from 19-25, six were 26-30, and four were 31-40 years old. Participants reported watching videos either on a daily basis (eleven participants), 3-5 times a week (six participants), or once a week (two participants). Thirteen participants reported watching 1-3 videos per session, three participants watch 4-6 videos, one participant watches 7-10 videos, while two participants watch more than 10 videos per session. The duration of the experiment and the number of active days (i.e. actually watched videos in YouTube) were measured for each participant as shown in Table 3.1. The duration of the experiment is the time from when the user installed the extension till the time of the last record sent to our server, while the number of active days is measured by counting the days that a user had at least one record.

3.4 Collected Video Interaction Dataset

Our dataset consists of interaction logs from YouTube videos for nineteen users over the period from December 25th, 2013 to May 27th, 2014. Each participant had a different duration of the experiment, based on when they installed the extension and when they stopped or deleted it, as shown in Table 3.1. Each log entry consists of user ID, time of access, video URL, video ID, video title, video length, video category, number of views, start time of the interval being watched, and end time

Table 3.1: Demographics summary for participants in the video viewing behaviour study. Watching frequencies and videos per session are the values reported by participants. (Note: duration in days)

P	Gender	Age Group	Watching Frequency	Videos/ Session	Experiment Duration	Active Days
1	Female	19-25	3-5 / week	1-3	16.08	8
2	Female	31-40	Daily	1-3	41.96	6
3	Female	19-25	3-5 / week	1-3	6.04	4
4	Male	19-25	Daily	1-3	27.46	14
5	Female	26-30	Once a week	1-3	11.99	4
6	Male	26-30	Daily	> 10	79.14	29
7	Male	26-30	Daily	1-3	77.42	38
8	Male	19-25	Daily	1-3	10.40	7
9	Male	31-40	3-5 / week	4-6	59.22	10
10	Female	19-25	Daily	1-3	75.53	52
11	Male	19-25	3-5 / week	7-10	21.06	13
12	Male	19-25	Daily	1-3	22.15	3
13	Female	19-25	3-5 / week	4-6	1.04	2
14	Male	19-25	Daily	1-3	76.07	70
15	Female	31-40	3-5 / week	1-3	68.85	49
16	Male	31-40	Daily	1-3	69.50	20
17	Male	26-30	Daily	> 10	7.59	7
18	Male	26-30	Daily	1-3	15.24	4
19	Female	26-30	Once a week	4-6	19.79	6

of the interval. We gathered a total of 4,786 interaction logs or records from the nineteen participants, which came from 2,129 unique videos. The distribution of the number of unique videos among participants is shown in Figure 3.3. Figure 3.4 illustrates the number of collected records per participant, which shows a similar trend to the number of videos with the exception of participant 11 who seemed to be more active with fewer videos, in comparison to participants 14 and 15 who had a higher number of videos.

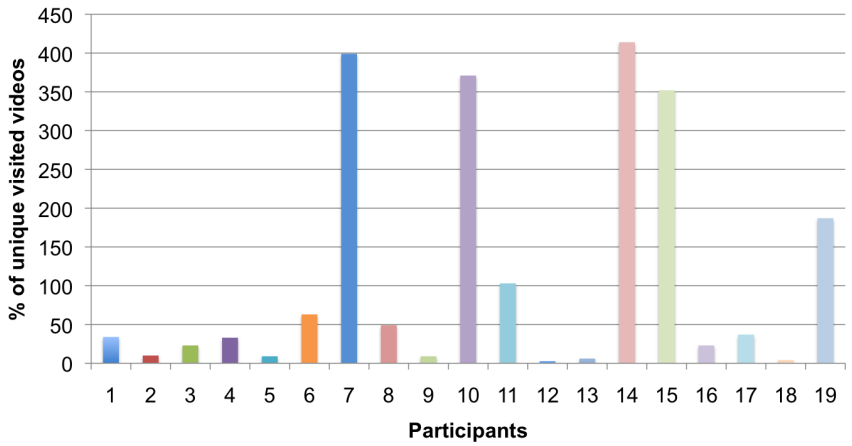


Figure 3.3: Number of unique visited videos per participant. Participants 7, 10, 14 and 15 had the most number of videos while participants 12, 13 and 18 had the least.

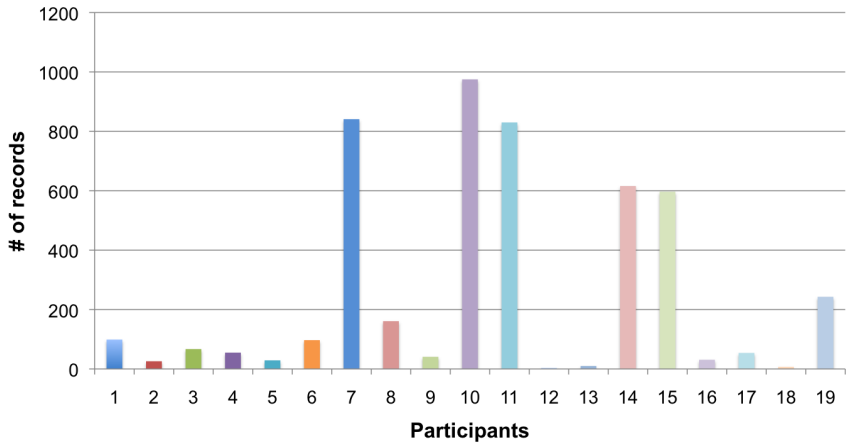


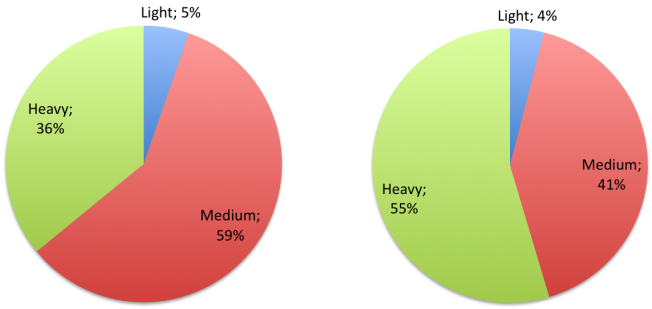
Figure 3.4: Number of collected records per participant. Participants 7, 10 and 11 had the most number of records while participants 12, 13 and 18 had the least.

3.4.1 Data Clustering

In order to look for differences between groups along with individual videos or participants, we had different clustering based on the activity level of participants and the length of videos.

Grouping Participants

To explore the difference between heavy viewers, medium and light users, we used the k-means clustering with $k = 3$ to divide participants into groups based on their total number of records, number of viewed videos and number of active days in the experiment. Based on this clustering, heavy viewers include participants 10 and 11, medium viewers are participants 7, 14 and 15, while the rest of the participants are light viewers. Heavy users were more active while viewing videos, as can be seen in Figure 3.5, where they watched on average fewer videos than medium viewers. However, they had more interactions on average indicated by the number of records.



Average no. of videos Average no. of records

Figure 3.5: Average number of visited videos and records per group. Medium viewers watched more videos on average, while heavy viewers had the most number of records, which indicates more activity among these users.

Video Duration Clustering

To examine the effect of video duration on different users' interaction, the k-means clustering with $k = 3$ was performed to divide videos into three categories based on their length: short, medium, and long. The clustering based on the collected data showed that short videos are videos within 193 seconds, medium videos have a duration between 193 seconds and 446 second, while long videos are longer than 449 seconds. Applying these groups, there were 774 short videos, 914 medium videos and 441 long videos in total.

3.5 Analysis 1: Watched Categories

The dataset was analyzed to examine each video's category and showed that the watched videos came from 16 different categories as defined by YouTube. These categories with the number of videos are: Music (681 videos), Science & Technology (289), Entertainment (238), Education (196), People & Blogs (163), Comedy (158), Gaming (112), Sports (98), Film & Animation (68), How-to & Style (59), News & Politics (25), Pets & Animals (19), Travel & Events (9), Nonprofits & Activism (8), Autos & Vehicles (5), and Trailers (1). Categories that had less than 2% of the total number of videos were grouped together in a category called Other. Figure 3.6 illustrates the distribution of videos among the new categories.

Similar to Sysomos Inc.[3] statistics on YouTube and [26, 42, 123], Music and Entertainment were the most watched categories among users. Even more, the percentage of Music videos is close to what is reported in Sysomos Inc. ($\sim$31%). However, in our data, Science & Technology was the second most watched category which was in 11th place according to Sysomos Inc. and was not mentioned in [26, 42, 123]. Education was highly watched among our participants and this was not the case in the previous research where it was not highly watched as it came 7th in the Sysomos Inc. analysis. This might be because many of our participants are students since we sent invitations for participation through the university mailing lists. Nevertheless, we cannot justify this claim as we have not collected any data about participant occupation or interest area. The other categories showed similar popularity among users to those presented in the previous work ([26, 42]).

[3] http://www.sysomos.com/reports/youtube/

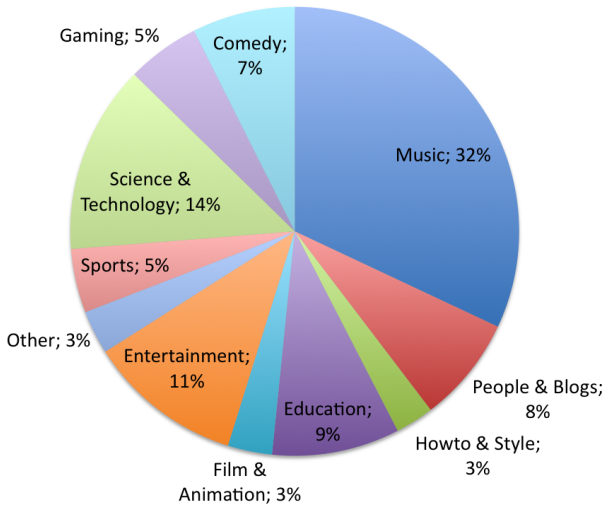


Figure 3.6: Number of visited videos per category. Music had the most number of videos 681 videos then came Science & Technology, which had 289 videos.

3.6 Analysis 2: Participants' Viewing Patterns

The goal of this study was to characterize each individual user's watching pattern and how often they occur to help define some design guidelines for a future video interface. Thus, we analyzed the dataset for each participant based on the six behaviours defined in Section 3.2. Participants showed large variations in the different behaviours. Moreover, each participant performed each behaviour with different frequencies. It indicated that individuals have a defined way of watching videos and this changes based on how much they actually watch. Thus, in this section we are going to categorize each participant's viewing pattern based on the behaviour they presented and how frequently these occurred. A detailed analysis of each behaviour is discussed in the following sections. Figure 3.7 shows the different shapes of participants' behaviours. The data for each behaviour was computed as the number of times that behaviour occurs divided by the number of videos that contained these behaviours. These values were then normalized based on the maximum behaviour the participant exhibited.

Based on the analysis, four main behavioural groupings appeared for most par-

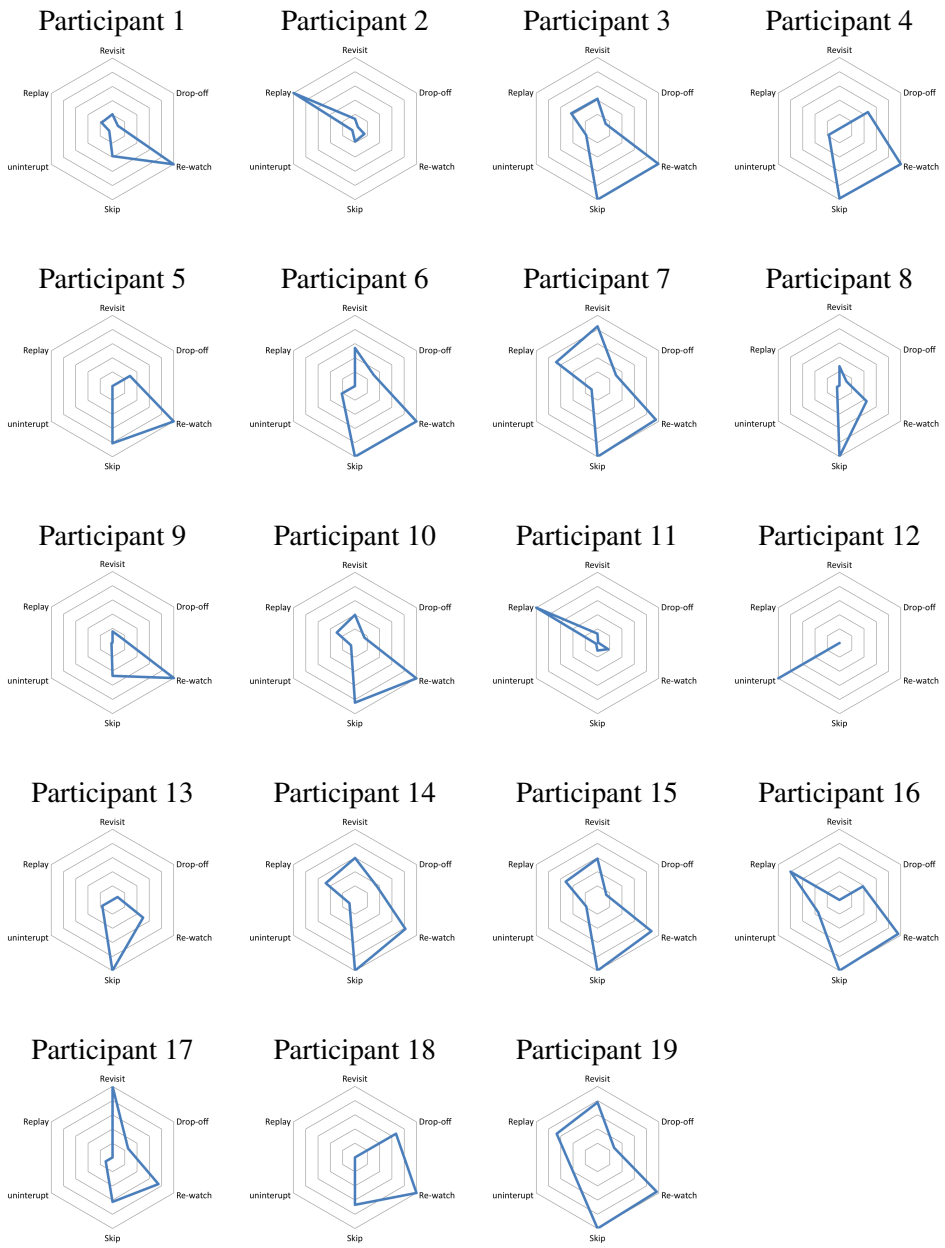


Figure 3.7: Participants' viewing patterns based on their own normalized behaviour and actions frequency.

ticipants which are skipper, re-watcher, replayer, and revisiter. The skipper category includes participants 3, 4, 5, 6, 7, 8, 10, 13, 14, 15, 16 and 19, the revisiter group has participants 6, 7, 14, 15, 17 and 19, while the replayer category contains participants 2, 7, 11, 15, 16 and 19. All participants except four (2, 8, 11, and 12) are re-watchers. Detailed participant viewing patterns are as follows:

Participant 1: She can be categorized as a re-watcher since it is the most frequent behaviour pattern for her.

Participant 2: She seems to be a replayer.

Participant 3: She is an active re-watcher and skipper.

Participant 4: He is a re-watcher and skipper similar to participant 3 but he tended to leave many videos before they finished.

Participant 5: She is another re-watcher and skipper.

Participant 6: He is similar to the previous participants but he is also a revisiter. Participant 6 seems to go back and watch the same videos in different sessions.

Participant 7: He appears to be very active while viewing videos. He is a re-watcher, skipper, revisiter, and replayer.

Participant 8: He is a skipper viewer with some re-watching and revisiting.

Participant 9: Similar to participant 1, he is a re-watcher with few skips.

Participant 10: She is a re-watcher and skipper with some replaying and revisiting similar to participant 3.

Participant 11: He is a heavy replayer similar to participant 2.

Participant 12: He is a passive viewer who never interacts with the videos while watching.

Participant 13: Another skipper with few re-watching.

Participant 14: Similar to participant 7, he is an active viewer who interacts so often while watching either by re-watching, skipping, replaying, or revisiting videos again.

Participant 15: She is another active viewer similar to 7 and 14 .

Participant 16: He shows a slightly different shape of viewing pattern. He is a heavy re-watcher, skipper, and replayer but never revisits videos.

Participant 17: He is the opposite of participant 16 in that he is a heavy revisiter but never replays videos in the same session.

Participant 18: He seems to get bored so easily with videos, which leads him to leave many videos before they finish, otherwise he is similar to participant 5.

Participant 19: Similar to participant 7, she is a heavy re-watcher, skipper, revisiter and replayer. However, participant 19 tends to watch more videos passively.

3.7 Analysis 3: Viewing Behaviour

To be able to categorize participants' behaviour described earlier, we needed to get a closer look at each behaviour. Thus a detailed analysis of each behaviour defined in Section 3.2 was performed using three different levels of analysis. First, to get a sense of the overall trend, the entire dataset was analyzed for each behaviour. Second, to explore how frequently this behaviour emerged in each category, the data was analyzed per category and lastly, for each participant, the data was analyzed separately to justify that these behaviours occur in individual data and not only collectively.[4]

3.7.1 Skip Behaviour

A user may skip parts of video to search for known items or portions of a video (e.g. specific step in instructional video), check whether it is the targeted video,

[4] Additional data is available in the Appendix in Section C.1.

check whether it is interesting to watch, or to pass over uninteresting portions [12]. This is also clear from the development of the Wadsworth constant[5] on YouTube where users can skip, for example, the first 30% of a video because it contains no worthwhile or interesting information. Users may have more specific information needs and selectively watch a video. Any time a user breaks the continuous viewing of a video by jumping to a part in the video that they have not seen before, then the number of skips for this video is increased by one. If a video contains at least one skip then it is marked to contain the skip behaviour.

Overall

The data revealed that 20% of the videos contained at least one skip action. Most of the skipped videos came from the Music category (29%), participants 7 (27%) and 10 (25%), and the medium viewers (52%). This indicates that participants do actually interact with a video to skip parts of it. This confirms our hypothesis (**H1** and **H1. Q2.**) and coincides with [46, 61, 70]. Looking at how often this behaviour exists on average per video, we found that when a video was skipped there were on average around 3 skip actions on that video. This indicates that it does not happen by chance and participants intentionally skipped parts of videos for different purposes. This was also found in [12]. It would be useful to relate these skips with users' intentions; however, as we mentioned earlier this might reveal participants identity. We could have asked participants to give feedback on each action but this would have been restrictive.

Per Category

The skip behaviour was significantly seen in videos from the Gaming category, as illustrated in Figure 3.8. This might be because users watch these videos looking for some tricks they can apply, which explains the selective watching behaviour where participants skip parts that they know. All other categories had results similar to the overall trend, in which around one-fifth of the videos in each category contained skips with the exception of the Science & Technology category, which was significantly less than the average with only 13%. Since most of the Science &

[5] http://en.wikipedia.org/wiki/User:Jorgenev/Wadsworth_constant

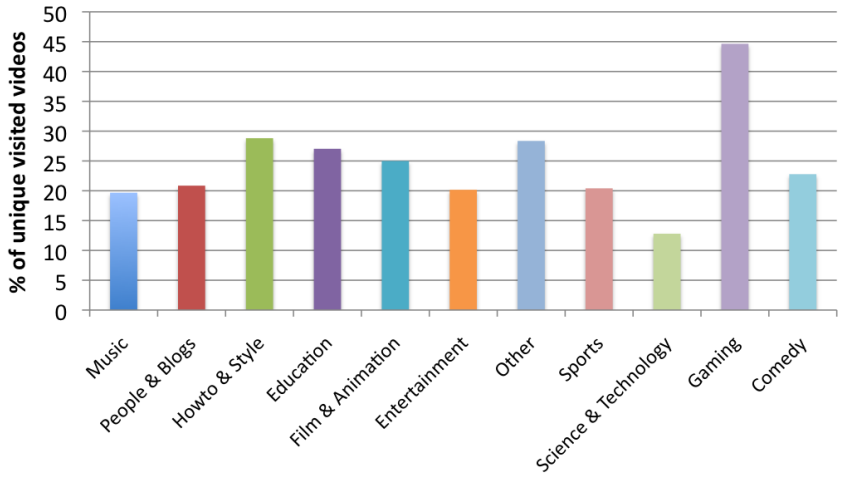


Figure 3.8: Percentage of skipped videos per category. Gaming (45%) showed a high percentage of videos being skipped while Science & Technology (13%) had a significantly lower percentage than the overall average.

Technology videos were short (i.e. less than a minute) participants might not need to skip any content, which can justify the low percentage of skipped videos in this category.

A skipped video in every category seems to have at least two skips on average per video. Again this justifies that these skips were purposely requested in those videos. A skipped educational video had the highest number of skip actions, with around 5 skips on average per video. It could be that participants are searching for specific information within the video, which causes a lot of skips as a result of misses. This was also seen in [70] where learners skipped often while watching lecture videos.

Per Participant

In terms of individual participants, six participants (4, 5, 7, 8, 9, and 10) had significantly more videos being skipped than average. More than 30% of their videos contained skip actions, as shown in Figure 3.9, while five participants (2, 12, 15, 17, and 19) had significantly fewer videos skipped than the overall average. Par-

ticipant 12 had no video with skip actions, which can be explained by the fact that we did not have enough data on him since he had only 3 records for 3 different videos on 3 different days. Participant 12 might had some privacy issues with *VVB* being on, which caused him to stop the extension, or he might have watched more videos on other devices that did not have the extension installed on them, since he reported watching videos on a daily basis.

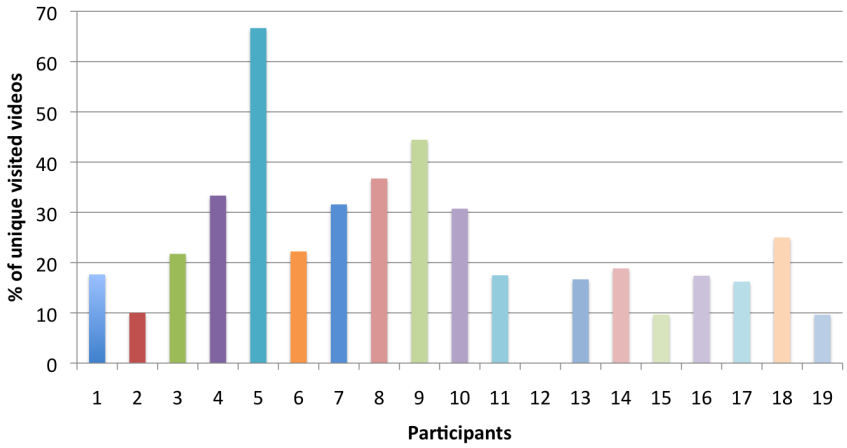


Figure 3.9: Percentage of skipped videos for each participant. Participants 4, 5, 7, 8, 9 and 10 had more than 30% of their number of viewed videos being skipped.

About the frequency of the skip behaviour per video, we found that apart from participant 12, who did not have any skipped videos, each participant had at least one skip action on average per skipped video. Participants 8 and 13 seemed to perform many skips when they skipped any video. They had 4 to 5 skip actions on average per skipped video. Most of participant 8's skips occurred in videos from the Entertainment category (8 skips on average per skipped video from Entertainment category), while participant 13 had most of the skips in the Science & Technology videos. An example of one of the videos viewed by participant 8 is shown in Figure 3.10 where it illustrates how he navigated and jumped around the video trying to find a part of interest and when it was found he watched that part for an extended period of time.

In summary, participants tend to skip videos intentionally to find specific infor-

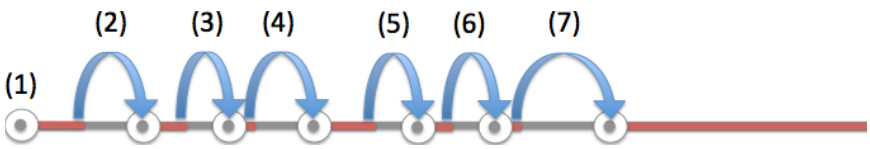


Figure 3.10: An illustration of how participant 8 viewed one of the videos. It is a clear example of how he skipped the video multiple times trying to find a position of interest and watch for an extended period of time.

mation within videos. More than one-fifth of the number of videos were skipped where the Gaming videos showed the highest percentage of videos being skipped by participants. As expected, heavy viewers skip videos more often in comparison to the other two groups (i.e. medium and light viewers). These results indicate that users tend to watch videos actively (i.e. interacting with the playback) when they have particular information to look for. These observations confirm that users do not watch videos passively (**H1**).

3.7.2 Re-watch Behaviour

A user may go back or rewind to watch some parts that they have seen before due to being interested in that part, to clarify any confusion, or to return to missed content [30, 70]. For example, in an instructional or a tutorial video a user may re-watch some steps to make sure that they follow the steps correctly, especially if they are applying them while watching. Any time a user watches any part of a video that has been seen previously, the number of re-watch actions is incremented and any video that has a number of re-watches larger than one is considered to contain a re-watch behaviour.

Overall

The data revealed that 25% of the collected videos had a re-watch action either implicit, where a person seeks to watch something he/she has not seen but accidentally while continue watching he/she watches a part that has been seen before, or explicit, where a person goes back in a video to deliberately watch a part that has been seen before. 24% of videos had explicit re-watch actions while only 1% had implicit re-watch behaviour. The overall observations show that participants do ac-

tually watch parts of videos again and this happens most often intentionally, which concurs with [17, 46]. Looking at the number of re-watch actions that occurred on each re-watched video revealed that there were on average 3 portions being re-watched per any re-watched video. These results answer **H1. Q3** and once more confirms that people do interact with videos while watching and they do not just sit and watch a video from beginning to end.

Per Category

Re-watch behaviour was significantly seen in How-to & Style, Film & Animation, Sports, and Gaming categories where more than 30% of the videos contained a re-watch behaviour (Figure 3.11). On the other hand, Comedy, Science & Technology, and People & Blogs had significantly lower percentages of videos that contained this behaviour. Film & Animation exhibited the highest percentage (∼37%) of videos being re-watched, which can be due to the fact that most animated movies contain a mix of comedy, actions and emotions, which are some of the key factors [100] prompting users to go back and watch the related parts again and again.

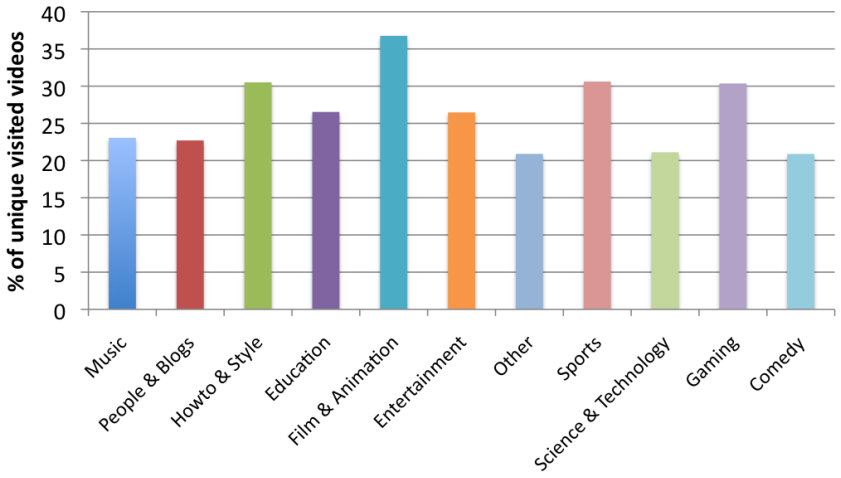


Figure 3.11: Percentage of re-watched videos per category. Each category had at least 20% of its videos encountering a re-watch behaviour. Film & Animation exhibited the highest percentage (37%) of videos containing a re-watch activity, which can be due to the mix of comedy and emotion these videos contain.

In terms of how frequent this happened per video, a re-watched educational video showed the highest average number of re-watch actions, containing seven re-watched portions on average which confirms the second part of hypothesis **H3**. This can be due to missing content, confusion, or in order to follow some steps within the video. Similar behaviour in educational videos was also found in [17, 70]. How-to & Style and Entertainment re-watched videos had three re-watch actions while all other categories exhibited two re-watch actions on average per any re-watched video. Having at least 20% of each category's videos containing a re-watch behaviour where each re-watched video contains at least two re-watch activity shows that this behaviour can emerge in any category, which contradicts with other researchers' [28, 44, 69] claim of the occurrence of this behaviour in Educational and How-to videos only.

Per Participant

In regard to individual analysis, participants 3, 5, 8, 9, and 18 had a significantly higher percentage of videos that were re-watched (more than 33%), as illustrated in Figure 3.12. However, these participants, aside from participant 8, watched few videos over the duration of the experiment. Around 37% of participant 8's videos contained re-watched portions. Looking at the frequency of this behaviour per participant showed that participant 9 tended to perform this action multiple times per video where he had eight actions on average per a re-watched video. Participants 1, 5, 10, and 11 re-watched four to five portions on average per any re-watched video. These results confirm that the re-watch behaviour not only occurred in the crowd data but even for each individual.

To summarize, the re-watch behaviour is common in any type of video and among participants. More than one-fourth of the viewed videos contained a re-watch behaviour, which confirm a user's high engagement with videos. Having this data provides a potential approach to detect which parts are appealing within videos that can be used to recommend clips for others to watch or to generate the video's abstraction and summarization.

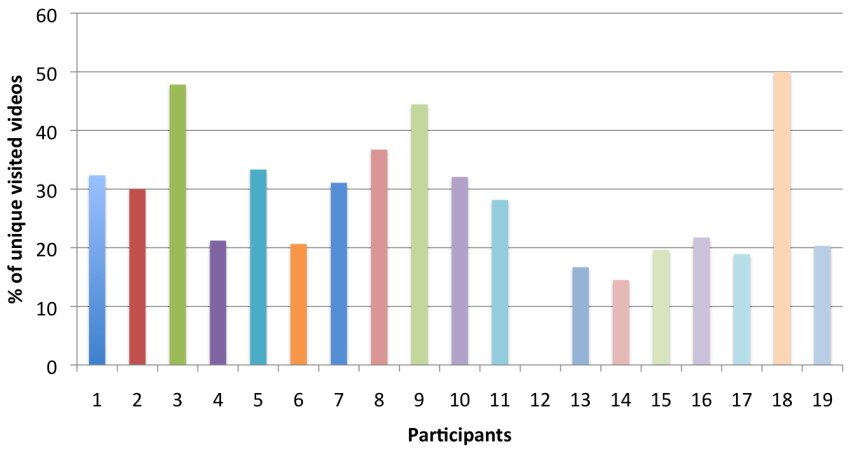


Figure 3.12: Percentage of re-watched videos for each participant. Participants 3, 5, 8, 9, and 18 had a significantly higher percentage of videos that exhibited re-watch actions (more than 33%).

3.7.3 Replay Behaviour

The data was analyzed to examine whether people watch videos in their entirety and multiple times in the same session and whether they do this explicitly per video or as part of repeating a playlist. If a video is played in its entirety for another time in the same session then the number of replay actions of this video is incremented. We anticipate that few videos will be replayed and most of these videos will be Music videos that are coming from playlists since users may play a playlist in the background while they are doing other tasks on their computers. We also predict that Comedy and animated short videos will be replayed, since they are funny, appealing and not too long to be re-watched entirely.

Overall

The results showed that only 4.5% of the videos were replayed entirely and only one-third of these were from a playlist, which answers our question **H3. Q2**. This indicates that when people replay a video, most of the time they deliberately repeat watching the entire video and not only because it is part of a repeated playlist, which contradicts with what we expected. As we expected, most of the replayed

videos were Music videos (77%) followed by videos from the People & Blogs category (9%). Most of the replayed videos from the People & Blogs category were short videos (75%), which can be the reason for re-watching the entire video rather than only small portions of it. Some videos that are categorized as People & Blogs on YouTube contain funny clips, which may justify the replay behaviour on these videos.

Investigating how frequently this behaviour occurred per video revealed that when a video is replayed, it has a high number of replay actions, 7 times on average. Going back to **H1. Q3**, people do re-watch the entire video in the same session multiple times but not so frequently.

Per Category

Based on video categories, shown in Figure 3.13, the percentage of the videos being replayed was very low. The Music category had significantly high percentage with $\sim$11% of the videos being replayed confirming the first half of our hypothesis **H3**. We think most of these replayed music videos were rn in the background while users were working on other tasks; however, our data cannot confirm this claim. In contrast, Sports and Gaming videos were never replayed in the same session, though, they had many re-watched parts as mentioned in the previous section (Section 3.7.2).

A result close to the overall frequency of actions per video was also seen in the replayed videos from Music category where a replayed music videos had around 9 replay actions. This high frequency of replay actions per video in a single session can justify the playback of the video in the background. Replayed videos in other categories apart from How-to & Style only had one replay action per replayed video, which is what we predicted. However, the How-to & Style category had only one replayed video, which had 4 replay actions.

Per Participant

Participants also had a similar variation in the percentages of videos that contained a replay action (Figure 3.14). For example, participants 2, 3, 11, and 15 had more than 10% of their videos being replayed while nine participants (4, 5, 6, 8, 9,

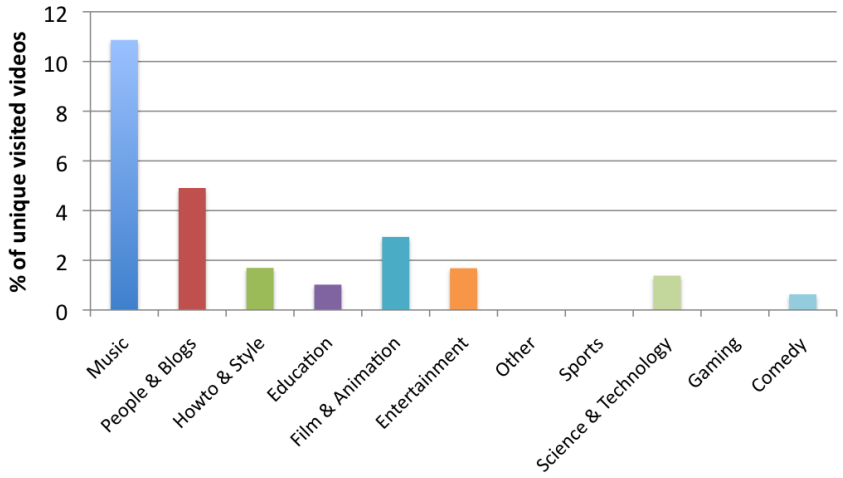


Figure 3.13: Percentage of replayed videos per category. Some categories had no replayed videos while the others had only few replayed videos. Music category had the highest percentage of videos being replayed ($\sim$11%) while Sports and Gaming videos were never replayed in the same session.

12, 13, 17, and 18) did not replay any video in the same session. Participant 11 had nineteen videos replayed out of his total 103 videos. The number of actions per replayed video for each participant showed that each participant had only one replay action per replayed video with the exception of participants 2 and 11 who had more than five replay actions on average per video. A replayed video for participant 11 had seven replay actions on average.

To sum up, the replay behaviour rarely existed where only 4.5% of videos were replayed and this was also observed between categories and participants. This infrequent behaviour while watching videos aligns with a passive viewing of a video since the replay behaviour does not require any interaction from the user while watching the video content. Users only need to interact with the video once it finishes playing in order to replay it, thus the viewing of this particular video was passive. Another justification for it being passive viewing is the replayed videos from playlists where a video is automatically replayed without any actions from the user. This tiny percentage of replayed videos again confirms that users do interact with videos while watching.

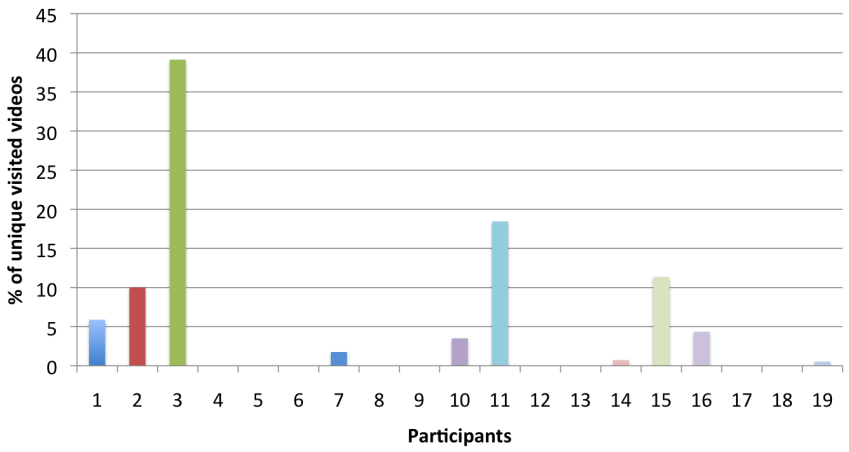


Figure 3.14: Percentages of replayed videos per participant. Participants 2, 3, 11, and 15 had more than 10% of their videos being replayed while nine participants (4, 5, 6, 8, 9, 12, 13, 17, and 18) did not replay any video in the same session.

3.7.4 Revisit Behaviour

We have seen that people re-watch parts of videos multiple times and they rarely re-watch the entire video in the same session. What about different sessions? Do people return to view a video they have seen before? And if they do, how often does this happen? In order to answer these questions the data was analyzed for a revisit behaviour where a video is accessed again but in a different session either to re-watch it entirely or just to find and enjoy some parts of its content. Each time a video is accessed again but in another session, the number of revisits for this video is incremented and the video is marked as a revisited video.

Overall

Around 13% of the total videos (269 videos) were revisited where around half of the re-watched behaviour occurred over multiple sessions. Thus, participants do re-watch portions of videos not only in the same session but even come back again to these videos to enjoy the content they liked over again or to refer back to some parts from videos. 54% of videos were revisited to be resumed while the rest were accessed to refer back to some content that had been seen previously. These

findings answer **H1. Q3** and verify that activity within a single video does not stop within one session but reoccur when a video is accessed for the next time. Having this data recorded and accessible to users can offer a potential for an easy way to find what a user has previously seen.

The majority of revisited videos were from Music (41%), followed by Science & Technology (16%), then Education (10%), which came third. Users may go back to educational videos to refer back to some content or simply to resume from where they left off [46, 61]. The data showed that most revisited educational videos were in the long videos group, which may confirm the resume action. Most revisited videos from the Music category were in the medium length group where 96% of them were revisited to re-watch and enjoy some of its content again. These results show that people resume watching videos in multiple sessions not only within a single session as shown earlier in Section 3.7.3. The data had also been examined for how often this behaviour occurred per video and the results revealed that on average a revisited video was accessed in two different sessions.

Per Category

Analyzing each category individually showed that Music, Education, Science & Technology and Sports categories had more than 14% of their videos being revisited (Figure 3.15). Most of the Music videos were revisited to be re-experienced again while the videos from the Science & Technology and Sports categories were accessed again mostly to be resumed from where they were left. Education videos were revisited for both resuming the playback and to refer to previous content. On the other hand, How-to & Style, Comedy and Entertainment categories had less than 7% of their videos revisited. This can be because these categories showed a high re-watching activity (Section 3.7.2) where participants enjoyed and experienced the content within a single session. Videos that were revisited from Comedy and Entertainment categories, they were resumed and re-experienced over multiple sessions, whilst How-to & Style videos were revisited to refer back to some previously seen content. Similar activity to the overall trend was seen for categories where a revisited video from each category had between one and three revisit actions on average.

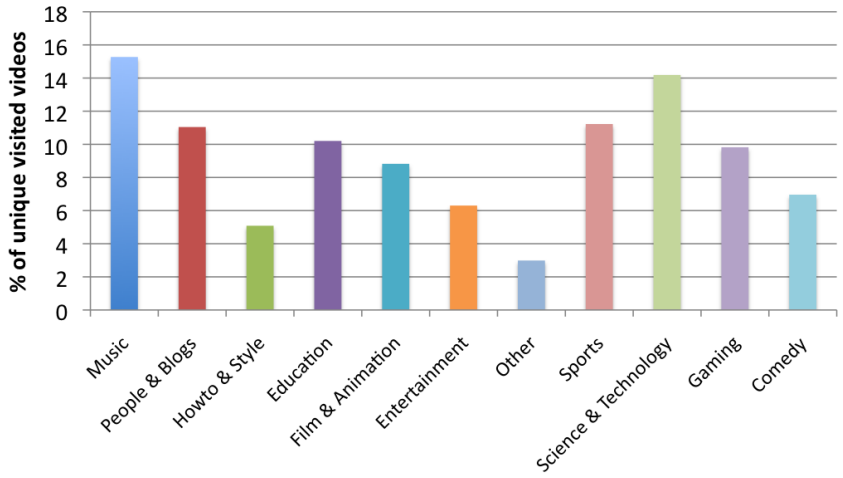


Figure 3.15: Percentage of revisited videos per category. More than 14% of the viewed videos in Music, Education, Science & Technology and Sports categories were revisited, whereas only 7% of videos from How-to & Style, Comedy and Entertainment categories were accessed again in multiple sessions.

Per Participant

With regard to individuals, illustrated in Figure 3.16, participants 2, 3, 11 and 19 had more than 20% of their videos revisited, while participants 4, 5, 12, 13, 16, and 18 did not revisit any video. Participants 2 and 3 viewed only a few videos, 10 and 23 videos respectively, and having one-fifth of these videos being accessed multiple time shows a high revisitation activity. Participant 3 mostly revisited Music videos to be enjoyed again. However, participant 19 went back to watch Science & Technology videos in different sessions where 75% of the time was to continue watching these videos from where they were left and re-experience some of their content. The lack of this behaviour for some participants could be explained by the small number of videos they watched over the duration of the experiment (e.g. participant 18 watched only 4 videos over 4 days).

The analysis of how frequently this happened per video for each participant showed that all participants who revisited a video, except participant 11, had at least one or two revisit actions on average for that video. When participant 11

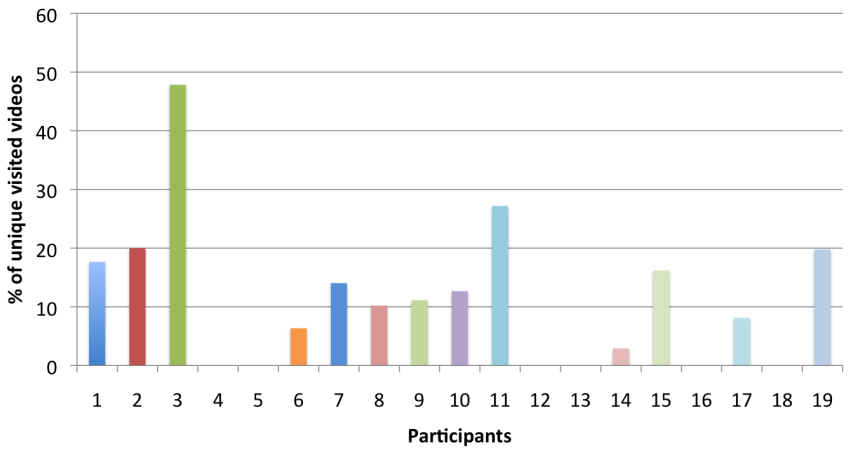


Figure 3.16: Percentages of revisited videos per participant. Participants 2, 3, 11 and 19 had 20% or more of their videos being accessed in multiple sessions while six participants (4, 5, 12, 13, 16, and 18) did not revisit any video in multiple sessions.

revisited a video, he tended to go back to it four times on average. Most of these came from Music videos where he re-experienced the same video again.

In short, the revisit behaviour appeared in 13% of the videos being watched by participants and across all categories where 46% of these were accessed multiple time to be re-experienced. Looking at this based on the number of videos uploaded to YouTube means a large number of videos are actually being accessed multiple times by users. Hence, having this data available to users would help them easily access these videos when needed. Furthermore, having access to every part being watched would be much appreciated since users do not have to look for the parts that they would like to refer to or to enjoy. The data showed that a revisited video is accessed twice on average, which indicates the necessity of having access to such information.

3.7.5 Drop-off Behaviour

One of the known issues with online videos is that viewers tend to abandon videos before reaching the end and that they do not watch the entire video [59, 61, 124, 126]. To investigate how often this happened in our dataset, we looked at any video

that was not watched entirely, any video that was left before the end, and when a video was left or abandoned. A viewed percentage of each video was computed based on the ratio of the watched content to the video duration. Those videos that had a ratio less than one were considered as abandoned videos. Some researchers (e.g. [61, 124]) found that the length of the video has an effect on the drop-off ratio. We would like to examine how this compared with our data.

Overall

The results revealed that around 60% of the videos were not watched entirely where the starting part was skipped, or abandoned before they reached the end. This confirms [59, 61, 124, 126] findings and answers our question in **H2. Q1**. This can be because participants found out that the video is not the targeted video or because they found what they were looking for, or just lost interest in the content. Most of the dropped videos came from the Music (28%) category where two-thirds of these videos were medium length videos. 5% of these music videos were shared videos where the starting part is skipped (i.e. shared from specific time feature on YouTube). This suggests a use case where recorded data can help users to easily share the parts they want from a video by simply using their own viewing history. 13% of the Music videos that were not entirely watched were skipped from the start. This signifies that users are applying the Wadsworth constant by themselves where they skip to the interesting information in the video.

To determine the percentage duration of a video being watched, we looked into when a video was abandoned and how much of it was actually watched. Figure 3.17 shows in seconds how long videos before were abandoned by illustrating how many videos were abandoned after t seconds from the start of the video. 50% of the abandoned videos were left within 2 minutes from the start while 10% were left within the first 10 seconds of the video which, according to YouTube's Creator Playbook[6], is the time needed to hook viewers into watching the rest of the video. Looking into each group of video durations (i.e. short, medium, and long), the data revealed that 50% of the short videos were only watched for 48 seconds, 50% of medium videos were left after 156 seconds, while 50% of the long videos were

[6] http://www.youtube.com/yt/creators/

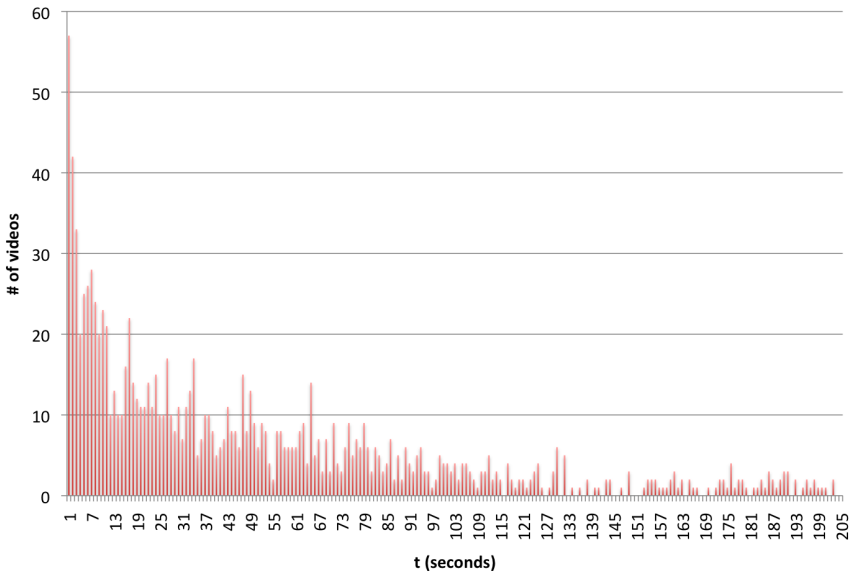


Figure 3.17: Number of videos being left before t seconds from the start of the video. 10% of abandoned videos were left within the first 10 seconds of the video.

abandoned after 237 seconds. Participants tended to watch long videos for much longer before being abandoned in contrast to the medium and short videos. This might be because short videos are more concise where a short shot of any short video can give a hint about its content and thus gets abandoned earlier than medium and long videos.

Examining the data on how much of a video is watched before stopping it, we found that 10% of the abandoned videos had just 3% of their content being watched, as illustrated in Figure 3.18. 47% of the videos were abandoned before reaching half of their content, which indicates that large portions of videos were wasted. Looking more closely into the duration groups, we found that 49% of short videos were abandoned before reaching 50% of the video length and only 6% of the short videos were watched entirely. In the medium length videos, 53% of the videos were left before reaching 50% of its duration and again only 6% were watched completely. However, 60% of long videos were abandoned before reaching half way through their content and 8% were watched completely. In com-

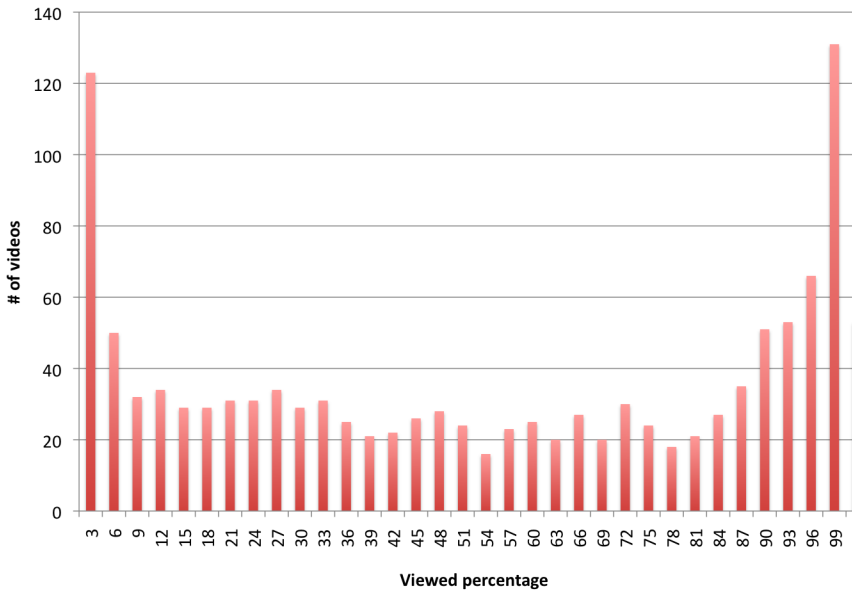


Figure 3.18: Number of videos per viewed percentage. 10% of abandoned videos had 3% of their content being watched.

parison to medium and short videos, shorter portions are watched from the majority of long videos, which matches [61, 124] results. Hwang et al. found around 55% of the long videos stopped before reaching 40% of the video length while shortest videos, only less than 20% of the videos dropped-off before viewing 40% of the video. These findings can be used for the benefit of the user where they can filter their viewing history based on how much of the video they have watched.

Per Category

Looking at the percentage of videos that were not completely watched within each category revealed that Gaming had a very high percentage of videos being left before the end, with around 78% of its videos abandoned or not watched entirely. Usually people watch these kind of videos looking for some tricks, and when found and mastered users stop playing the rest of the video. This can also be justified by the high percentage of videos in this category that were skipped so often as shown in Section 3.7.1, which confirms the search approach. Other categories had at least

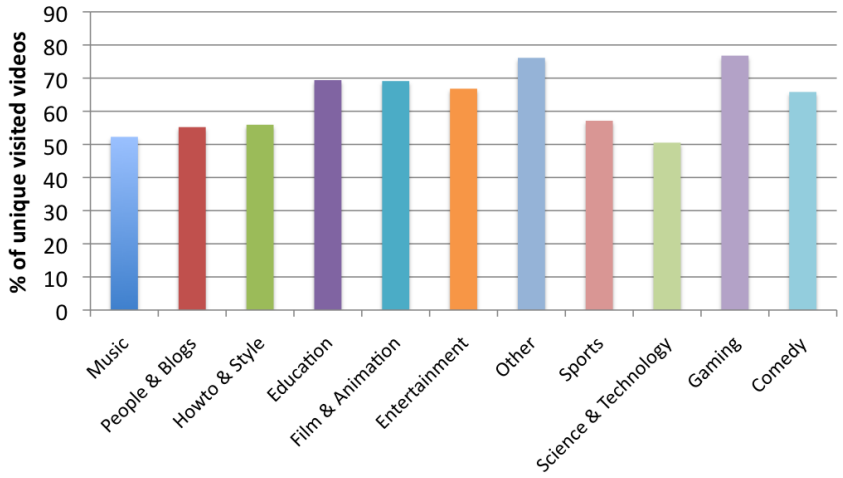


Figure 3.19: Percentage of abandoned videos per category. 78% of the viewed videos in Gaming were not watched entirely while around 50% of Science & Technology videos were abandoned before they finished playing.

half of their videos being not completely watched as shown in Figure 3.19. Putting this in the perspective of videos viewed on YouTube, around 5.5 billion videos (based on comScore.com[7] 11 billion unique videos were viewed on YouTube as of March 2014) are not viewed entirely which is a large number of videos. This data can be used to shorten videos where it could help to generate abstraction and summaries of videos.

Per Participant

With regard to analysis of individuals, six participants (4, 5, 7, 9, 14, and 18) left more than 70% of their videos before the end. Participants 5 and 18 actually did not watch any video till the end. It is worth mentioning here that these two participants had only a few videos in total; participant 5 had eight unique videos and participant 18 had only three videos. As shown in Figure 3.20, all other participants aside from participant 12 had at least 33% of their videos being dropped off before being

[7] https://www.comscore.com/Insights/Press-Releases/2014/4/comScore-Releases-March-2014-US-Online-Video-Rankings

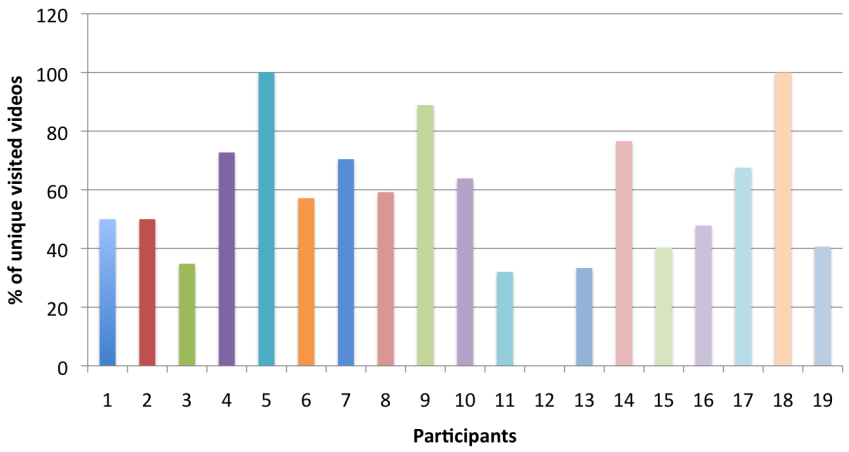


Figure 3.20: Percentage of videos that were not completely watched for each participant. Aside from participant 12, at least 33% of each participant-viewed videos were dropped-off. Participants 5 and 18 had all their videos abandoned.

played in their entirety. These results show the high tendency of abandoning videos before they finish among all participants.

To wrap up, abandonment of videos was highly seen across categories and participants where more than two thirds of viewed videos were not watched completely. The data showed that shorter videos were dropped-off earlier than longer ones; however, larger portions of short videos were watched in comparison to longer videos. At least half of the videos in each category were not entirely viewed and each participant abandoned at least one third of their viewed videos. Having this data available to users can offers a potential feature to filter videos and to generate video abstraction.

3.7.6 Interrupted Viewing

The results discussed earlier showed that people do interact with videos while viewing by applying the different actions (e.g. skip, re-watch, or drop-off the video). In this section, we would like to investigate how often these interruptions appear no matter what action is used to assess users' engagement while viewing. We achieve this by examining whether participants in a single run watch videos

entirely (i.e. from start to end) without any interruption and are passive viewers, or if they watch videos actively where they interrupt the video by skipping parts, re-watching other parts, or leaving the video before the end. We defined a video as uninterrupted if it has been viewed entirely without any actions from the user (e.g. skip, rewind in the middle, or leave the video), otherwise it is considered interrupted.

Overall

The results revealed a high percentage of interrupted videos where 1404 videos ($\sim$66%) were not watched passively. This suggests a clear proof that most of the time people do not watch videos uninterrupted, which confirms our hypothesis **H1**. 12% of these interrupted videos encountered all three actions (skip, re-watch, and drop-off), which indicates high engagement with their content. Our findings contradict with [59, 124] who found that 80% of their collected videos had no user interactions. This might be because Yin et al. [124] collected data during the 2008 Beijing Olympics and viewed videos were related to the Olympics Games and hence were from a more specific category. However, if we compare their results to the Sports category in our data, our data still shows a high interactivity where more than 60% of videos from this category had user interactions.

Per Category

To explore how often this behaviour emerges in each category, the data was analyzed per category as illustrated in Figure 3.21. Gaming, Film & Animation, Education, Comedy, Entertainment and How-to & Style categories had more than 70% of their videos being interrupted. This can be seen from the large number of skips, rewind, and drop-off behaviours presented on these videos as shown in previous sections. For example, in Gaming, How-to & Style, and Education videos, people may skip to specific steps within a video as needed or rewind some parts to understand or follow such parts [70]. Our result for the Education category agrees with [10] who found that 83% of the educational videos were actively viewed where they had between one and four interactions. All other categories had at least 50% of their videos being actively interacted with by participants. This high interactiv-

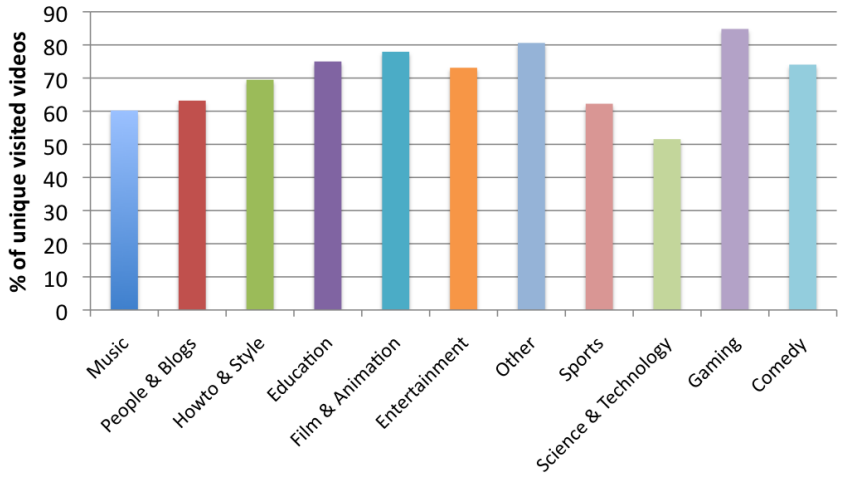


Figure 3.21: Percentage of interrupted videos per category. At least 50% of the videos in each category were actively viewed where Gaming (85%) showed the highest percentage of videos being interrupted while watching.

ity with videos provides evidence that users do not just sit down and watch a video from start to end without performing any action with the video while it is playing.

Analyzing the number of interactions (i.e. either skip or re-watch) per video in each category showed that each category had at least one interaction per video on average. Education showed the highest average number of interactions where it exhibited four interactions per video confirming [10] findings. These results demonstrate that users are most likely to interact with any type of videos, which contradicts Kim et al. [70] and Gkonela et al.'s [44] claim of the existence of interactivity in educational and instructional videos only.

Per Participant

In terms of participants, as demonstrated in Figure 3.22, each participant had at least 33% of the videos interrupted while watching except for participant 12 who watched all videos entirely and passively. In contrast to participant 12, participants 5 and 18 were active with every video they watched where they showed high percentages of skips and re-watch. Looking at the average number of interactions a

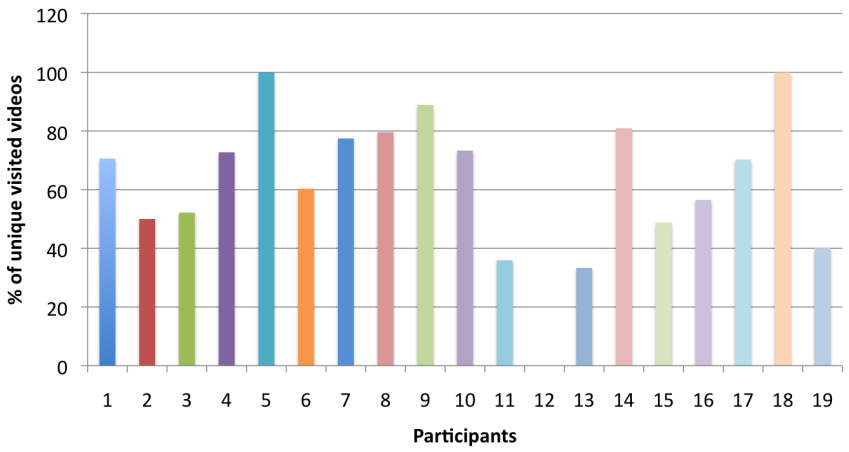


Figure 3.22: Percentage of videos that were interrupted for each participant. Aside from participant 12, at least 33% of each participant-viewed videos were actively watched. Participants 5 and 18 had all their videos been interrupted.

participant performed per video, we found that each participant interacted at least once on average per video with the exception of participant 12. Participants 9 and 11 showed high activity while watching videos where they performed five interactions (i.e. either skip or re-watch) on average per video (e.g. Figure 3.23). These results show that each participant is more likely to interact with videos they are viewing, which provides proof for those who think that people do not watch videos actively.

To conclude, participants are most likely to interact with any video they are playing and at least one-third of the videos a user watches are actively viewed or interrupted while playing them. Our data rejects Kim et al. [70] and Gkonela et al. [44] finding that this behaviour is only in educational and instructional videos. We showed that this activity occurred in any type of video and for each participant. Recording this interactivity and giving users access to it can offer different tools, for example, a search tool for previously seen clips from videos, a summarization and authoring tool, and a sharing tool.

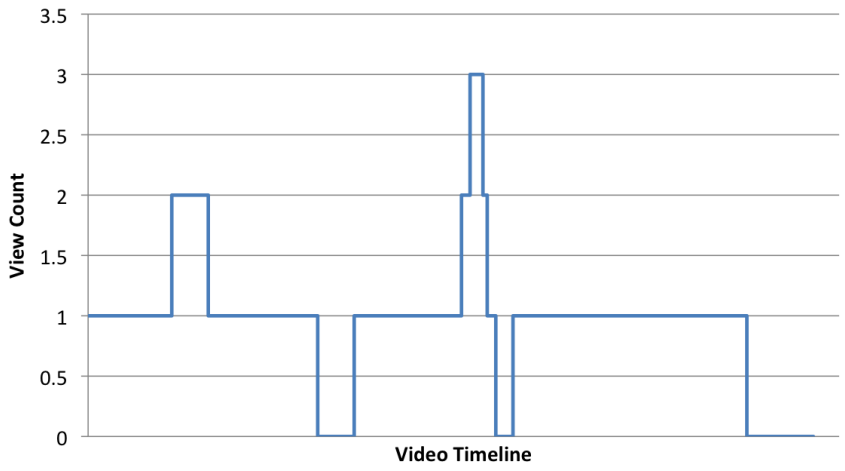


Figure 3.23: An example of one of participant 9 video viewing interactivity. As shown from the view count per second in the video, he performed 2 skips, 3 re-watch, and a drop-off at the end indicating a high user engagement while watching.

3.8 Analysis 4: Impact of Video's Popularity on Users' Behaviour

To investigate whether a video's popularity plays any role in users' watching behaviour, the data was analyzed to test if more actions appeared in the popular videos or if the popularity does not make any difference. Popularity was judged based on the number of views per video pulled from YouTube's metadata. A Pearson product-moment correlation coefficient was computed to assess the relationship between each behaviour and the video's popularity. There was no correlation between any of the behaviours and the number of views per video as shown in Table 3.2, which matches Hwang et al. [61] findings. This indicates that popularity does not affect the user's behaviour within a video, which rejects **H2. Q3**.

3.9 Analysis 5: Impact of Video's Length on Users' Behaviour

A video's length is known to be a key measure of a viewer's engagement. Wistia.com suggests that short videos will get more eyes than longer videos and the

Table 3.2: A Pearson product-moment correlation coefficient between each behaviour and the video popularity (i.e. No. of views obtained from YouTube) and video length. There is no correlation between any pair, (Note: * Correlation is significant at the 0.01 level (2-tailed).)

	No. of views	**Video length**
Skip	-0.025	0.196*
Re-watch	-0.013	0.103*
Replay	-0.007	-0.014
Revisit	-0.006	0.059*
Drop-off	0.004	0.094*
Interrupt	0.019	0.107*
Viewed Percentage	-0.041	-0.137*

longer the video, the less people remain engaged. To investigate whether this applies to our collected data, a Pearson product-moment correlation coefficient was computed to assess the relationship between each behaviour and the video length. The results revealed that there was no or a negligible relationship between any of the behaviours and the video length as presented in Table 3.2. This indicates that video length does not have any correlation with the user's behaviour within a video and this rejects **H2. Q3**. The findings of the viewed percentage of a video does not match Yin et al. [124], who found an inverse correlation between viewed percentage and video length. It seems like the content of the video, not the length, affects the activity within videos where videos that have great appeal are most likely be interacted with more.

We also analyzed the data based on the different duration groups, short, medium, and long, to check for any differences on the frequency of each behaviour per video when videos are grouped. Since the assumption of homogeneity of variance was not met for the data of each behaviour we ran the Welch test to investigate the difference between the three groups in terms of skip, re-watch, replay, and revisit. The results revealed that there was a significant difference between the three durations with respect to each behaviour, as shown in Table 3.3. Longer videos had a significantly larger number of skips than medium videos and shorter ones, which could be due to the presence of unimportant parts in these long videos and users not willing to spend a long time on them. They also had more re-watch actions than the

Table 3.3: Welsh Test comparing the average number of actions occurred per video in each duration group: short, medium and long. A significant difference was found between the three groups for each behaviour. Longer videos had significantly more skips and re-watches but less number of replays. (Notes: ( ) is Standard Deviation; $^{*}p < .001$)

Behaviour	Short	Medium	Long	Welch test
Skip	1.55 (0.95)	1.95 (1.82)	4.15 (5.47)	20.66^{*}
Re-watch	2.33 (5.71)	3.92 (10.03)	4.95 (7.27)	6.72^{*}
Replay	5.95 (2.61)	7.84 (5.08)	1 (.095)	7.65^{*}
Revisit	1.55 (1.72)	1.98 (2.11)	1.88 (1.64)	1.38^{*}

other two groups; however, it was only significant in comparison to shorter ones. On the other hand, longer videos had a significantly smaller number of replays in comparison to medium videos. This was expected since users will re-watch parts of longer videos rather than replaying the entire video again due to its duration.

3.10 Analysis 6: Correlation between Users' Behaviours

We anticipated that some behaviours will have some correlation with other behaviours. For instance, more revisited videos will have more re-watch actions. Accordingly, we examined for correlations between any pair of the behaviours using the Pearson product-moment correlation coefficient as presented in Table 3.4. As predicted there is a low positive correlation between re-watch and revisit, suggesting that people mostly access a video again to re-watch some parts of it rather than resuming the playback. There is a moderate positive correlation between re-watch and skip behaviours, which is explained by the number of implicit re-watches which, by its definition, contains a skip operation and then a re-watch action, as described in Section 3.2. This can also explain why most re-watcher participants were also skippers as discussed in Section 3.6. A high positive correlation also occurred between the drop-off and interruption behaviour and this is again because

Table 3.4: A Pearson product-moment correlation coefficient between each pair of behaviours. There is a moderate positive correlation between re-watch and skip behaviour, which was also seen in the individuals' behavioural grouping (Figure 3.7) where most re-watchers were also skippers. * Correlation is significant at the 0.001 level (2-tailed).

	Skip	**Re-watch**	**Replay**	**Revisit**	**Drop-off**	**Interrupt**
Skip	1	0.653*	-0.005	0.254*	0.100*	0.301*
Re-watch	0.653*	1	0.006*	0.365*	-0.028	0.149*
Replay	-0.005	0.006	1	0.620*	-0.184*	-0.135*
Revisit	0.254*	0.365*	0.620*	1	-0.074*	-0.031
Drop-off	0.100*	-0.028	-0.184*	-0.074*	1	0.854*
Interrupt	0.301*	0.149*	-0.135*	-0.031	0.854*	1

we considered any video that was dropped off as an interrupted video. Revisit and replay showed a moderate positive correlation, indicating that the more a video is replayed the more likely that it will be revisited again in the future sessions to be either played entirely or visit some parts of it only. We expected that videos that have more skips will more likely be dropped; however, the data showed very low positive correlation between them. A moderate correlation was found between skip and interruption and this was observed in Section 3.7.6 where most of the interruptions were caused by skip actions. All other pairs had little to no correlation.

3.11 Design Guidelines for a Video Interface

Based on the characterizations of the users' viewing and navigation patterns some guidelines are proposed for a video interface to account for each behaviour. These guidelines are:

- The interface should allow viewing videos similar to any typical video player where users can play a video and control its playback. This will allow users who like to watch their videos passively from the start to the end to use the interface as they would normally do.
- Having a replay functionality within the interface would be helpful for users who like to play their videos over again and again.

- Including a filmstrip in a video interface would allow users to easily get a glance of the video's content, which could help users to quickly decide if they are going to abandon the video or not. This would also aid users to navigate to specific content with videos that may require few skips from the user.

- The interface needs to provide users with access to previously seen videos if they wish to revisit and view them again. It should also provide access to each interval a user has watched from any video. This would allow users to easily go back to what they have seen and refer to or enjoy it again.

- Identifying previously seen intervals within a video interface would be helpful for skip, search and sharing, which could be accomplished by highlighting re-watched segments, known drop-off regions or mostly-skipped parts.

- Allowing the crowd-sourcing of user interactions with video could lead to more effective recommendations, automatic summary generation or social navigation tools.

3.12 Study Limitations

This section describes some limitations to our study and collected data. Some of these are addressed in Section 9.3. The first limitation is that our plug-in was developed for a desktop platform and more specifically worked only in the Google Chrome browser, which limited the size of the data collected. Thus, our results did not count for users who watch videos on platforms other than desktop and in browsers other than Chrome. Another limitation is the number of people that participated in our study and their demographics. We did not have any users younger than 19 years or older than 40 years. This limits the discovered results and needs to be furthered with more users covering the different age groups and user types. This limitation might be caused by our recruiting procedure since we only sent the invitation email for participation using university mailing lists and through our friends. Collaborating with YouTube would help to collect more data, which would cover most of these limitations, for example platform, number of users, age groups, and

users types. Nevertheless, even with such a narrow and small sample, we clearly saw a change in how these people watch videos. This is consistent with the literature on video viewing characteristics [28, 44, 61, 70, 126]. One of our main objectives was to confirm that there exist non-linear video viewing behaviours beyond linearly watching movies allowing us to build interfaces tailored to the behaviours we found. Thus it can also benefit those people who share similar viewing characteristics with our tested group. To establish viewing demographics and behaviours for different video genres would be facilitated by a partnership with an online video provider such as YouTube. We leave this to future work as it is outside the scope of this research.

3.13 Directions

Our analysis of users' viewing behaviour showed high interactivity while viewing videos, which has led to some insights on how this data can be useful and how it shows the potential for different tools that can be offered in any video viewer interface. However, as discussed in Section 3.12, our data is limited due to the platform used for data collection. This encourages us to develop a cross-platform plug-in to collect as much data as possible to reflect the general population. This will allow us to investigate how these behaviours and categories persist with a larger group of viewers. Moreover, it will enable us to explore other viewing patterns that can be translated into new features. These features then can be integrated into video interfaces to enrich users' experiences.

In future studies, we plan to connect these findings with users' intention by running a large-scale study where viewing history is recorded along with the purpose of such behaviour or action. This can be achieved by collecting explicit data from users through tags, questionnaires or interviews. It will allow us to better understand the purpose of each interaction to come up with meaningful viewing patterns and more defined behavioural groupings. Moreover, more actions can be also added to the captured behaviour, such as fast-forward and fast-rewind. These can introduce new behaviours or simply signify similar behaviour as skip where users skim the content in order to find something specific or just to check whether it is the intended video.

3.14 Summary

In this chapter, using traces collected in a 5-month period from YouTube, we presented a detailed investigation of users' video viewing behaviour. We have demonstrated that users are actively watching videos and they do re-watch different parts of a video no matter the type of video. This has changed the concept of a video being watched sequentially and passively to what is known as active viewing. Most importantly, the dataset revealed that when users revisit videos for the second time, they are mostly coming back to refer to some of its previously seen content or to enjoy these parts again, not just to resume the playback. Having many revisited videos with a high percentage of the re-watch behaviour indicates that users are accessing videos looking for more precise information than they have already seen before. This motivates us to explore the usability of such data when users have access to it. Thus, we plan to design an interface that is tailored to the behaviours we discovered by keeping track of users' viewing and navigation practices as well gives them access to this data. We aim to test the feasibility of having personal viewing history accessible to users and to explore different visualizations that can be applied to communicate the viewing history in a clear and comprehensible way to users. We want to check how these visualizations can help utilize and manage users' viewing behaviour.

Additionally, we think having access to portions of videos a user has previously seen can help finding and locating a specific content from certain watched videos. This encourages us to investigate whether having access to such data can speed up and improve the search for previously seen content. Furthermore, having more than one-fifth of the videos containing skip actions and one-fourth of the videos containing re-watch behaviour can be used to signify the interest or popularity of each part within a video. This information then can be used to easily navigate to the most or least trending content of the video to be referred back to, enjoyed again or simply shared with others to have similar experience. It also offers an effortless technique to generate video abstractions or summaries by, for example, stitching up the most popular portions of a video. Thus, using our interface, we intend to study these applications when users have access to their detailed viewing history and whether that can improve their performance.

Users' classification based on their viewing patterns helped in defining some design guidelines for a video interface. Through examining each participants' dataset in correspondence to their interactions, we were able to categorize and identify each user's viewing patterns. Four main behavioural groupings appeared for most participants: skipper, re-watcher, replayer, and revisiter. These findings introduce the potential for new approaches and techniques for a video viewer system and help in formulating some design guidelines for such a system and type of users. Thus, in the next chapter we look at how to develop an interface that records users' viewing history and provides them with access to this information. We want to investigate whether this data will be useful for these types of users and how it can be utilized.

Chapter 4

Recording and Utilizing Video Viewing Behaviour

The dramatic increase in the quantity of video now available, either in the online form from services such as YouTube and Vimeo or the more personal form of home video, provides new challenges such as finding specific videos (or intervals within video), authoring new videos from existing content (e.g. video summarization or home movie editing), sharing video playlists (or even intervals) of online video and annotating dynamic content. The increased volume and short average duration of online video has led to new use cases: users are generally free to skip ahead to find intervals which interest them; users may re-watch parts of the video they enjoyed; users may employ temporal video links to provide instant access to a specific time within a video, skipping irrelevant sections; playlists and temporal links provide opportunities for customization and sharing of video consumption among users. In short, users can easily view personally interesting intervals and are not required to view the entire video. These use cases are supported within current interfaces; however, there are no existing mechanisms to support new use cases that arise as a natural consequence, such as navigation of previously viewed intervals.

As found in Chapter 3 and [100], users often re-watch segments of video as part of the contemporary browsing and navigation experience. A historical record of video viewing would allow users (e.g. re-watchers) to quickly find, share and comment on popular intervals. Retrieving and reusing information from the past

is a common activity for users [33]. It has attracted researchers' attention in the domain of web browsers where they have introduced and developed different tools that keep records of users' browsing experiences for later use. This trend of research has recently started to get researchers' interest in the field of video where they found various applications of such information and metadata. In Chapter 2, we reviewed how researchers have used this information for various applications and tools. However, most of these tools do not provide users access to such metadata to understand how they are going to utilize it. Hence, in this chapter we explore how to get this data in a video interface and how to present it to users. We want to see if using users' viewing history facilitates better video navigation and mashup tools. To answer this question we developed a video interface and made sure that this interface looked like a normal interface leveraging the existing knowledge according to Don Norman [93].

In this chapter, we introduce a new video navigation interface, which captures users' viewing history (analogous to a web browser history) and offers users access to their viewing history, in order to investigate the significance of having access to such data and how it can improve users' tasks. Section 4.1 introduces video viewing history, how it is captured and a description of a history representation for video viewing or navigation. Section 4.2 presents the different usage applications for video viewing history. A description of the video navigation interface is presented in Section 4.5 along with the evaluation looking at the significance of the viewing history and users' impressions of using a personal video history to create mash-ups. Section 4.6 and Section 4.7 discuss the implications of the interface and directions for future refinement.

4.1 Video Viewing History

The video viewing history refers to the list of every video interval a user has watched associated with data such as video identifier (ID), video location, start and end times of the segment, and time of visit. It is captured to provide efficient access to what users have seen and allow them to go back to their previously viewed videos, and in particular, the intervals viewed within those videos without relying on their memory to remember what they have watched. This is accomplished by

recording a continuous video history for the user as they view a video.

4.1.1 How Viewing History Is Captured

Using the video navigation interface described in Section 4.4, navigation-level events are captured by recording high-level user actions such as seek, play, pause, and changing video. Internally, the user's video watching history is represented by a very simple data structure organized by linear user time. Whenever a segment of video is played, a new log is added to the history record. Each log consists of a sequence of basic elements (access time, video ID, start time, end time, favourite time, share time).

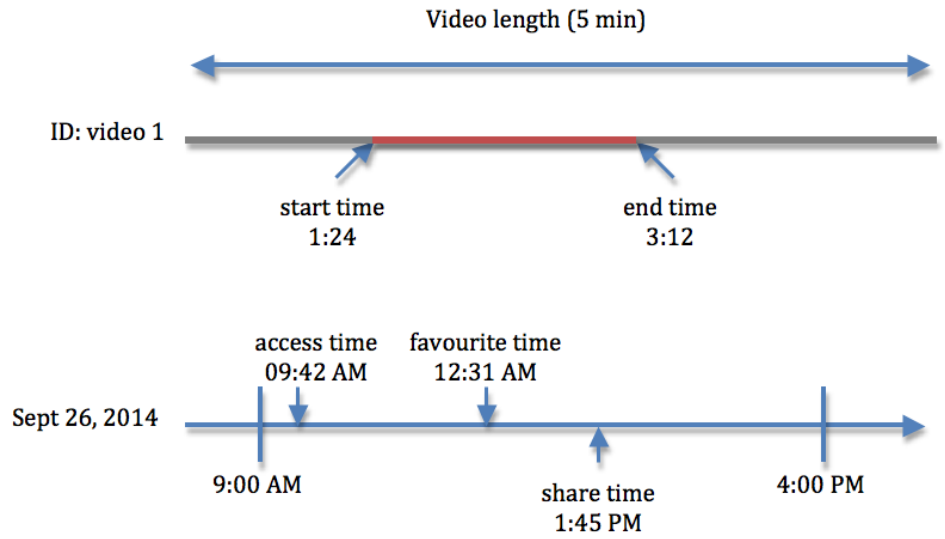


Figure 4.1: An illustration of a video history record (9:42 AM 26 Sept 2014, video 1, 1:24, 3:12, 12:31 AM 26 Sept 2014, 1:45 PM 26 Sept 2014).

Access Time: is the user's local time when they viewed this video interval.

Video ID: A unique ID for each video to identify the source of the content of the current interval.

Start Time: is the start timestamp of the current interval from the source video. This time is relative to the beginning of the video.

End Time: is the end timestamp of the current interval relative to the beginning of the video.

Favourite Time: The recent date and time when this interval was favourited. Nil if it has not been favourited yet.

Share Time: The recent date and time when this interval was shared. Nil if it has not been shared yet.

An accumulated view count is maintained for every instant of time in all viewed videos, which is derived from the intersection of all viewed intervals for each video. This provides the user with a better understanding of how a video was consumed and the importance of each interval based on their viewing frequencies. This can be used later to facilitate fast navigation, search, and video mash-up.

4.2 Use Cases of Video Viewing History

As mentioned in Section 2.2, researchers have identified different applications of the collective users' viewing history. In this section, we are going to present the main usages from an individual perspective. Video viewing histories and social navigation provide potential interaction techniques for fast navigation, event search, video mash-up, and segments and experience sharing.

4.2.1 Navigation

As seen in Chapter 3, people in real scenarios re-watch different parts of videos while they skip others. This behaviour may indicate that these parts are important, interesting, affective, or hard to understand [100]. This watching behaviour leaves digital footprints on the video frames, creating a non-flat video histogram emphasizing the interest of each part of the video. Visualizing this non-flat histogram lets users easily identify how this video was consumed and to quickly navigate to the different parts based on their level of importance [69, 82]. It provides a contextual map of how a video was viewed which subsequently defines a new quick way of navigating that video. Interestingly, using crowd-sourced data by applying the collective wisdom can be leveraged for the benefit of future viewers who have not

watched the corresponding video. First time viewers of a video can employ the collective wisdom to guide their navigation through this video and save their time of watching what others found interesting rather than watching the entire video. Thus, video viewing history can offer a faster navigation tool for skippers, re-watchers and droppers.

4.2.2 Scene Search

Looking for something that we have previously seen is sometimes crucial. We are not looking for just that video, but for that specific scene or segment from that video. Most online video sites (e.g. YouTube) provide users with a list of videos they have visited with no mechanism to find a specific scene aside from continuously scrubbing the video to find the scene you want. Having access to the detailed viewing history alleviates this problem by presenting not only the different videos but also the different scenes or segments within videos a user has watched. For the scenes a user has watched the most, applying the non-flat histogram in the visualization will make the search even faster. Providing filters which based on the viewing history can speed up the search task. For example, filters can identify the segments that were watched more than once, the segments that were watched from a specific video, and the segments that were watched at specific times. In Section 5.1.3, Section 5.2.3 and Section 6.5.7 we will show how having access to the viewing history significantly speeds up the search task.

4.2.3 Video Mashup

To produce some content that can be enjoyed or even shared with others, normally the source video is pre-existing and they need to be combined into a unified video. These source videos usually have a lot of content that is uninteresting or not worth including in the final production video. Thus, it becomes necessary to determine how to shorten these videos to emphasize the important content and reduce the time needed for viewing. It is one of the known issues that researchers (e.g. [95]) have tried to propose different approaches to ease the task varying from manual editing to fully automatic mechanisms. An extensive survey of these techniques is presented in [3, 86, 114]; however, in general, these methods do not take advantage

of any implicit information gathered as users consume video, which may be used to personalize the user experience.

As we mentioned earlier, viewing history can emphasize the interesting or important parts of each video, which can be used to filter the content of these videos and help deciding what to include in a video mash-up. It can be used to summarize a single video as shown in Section 5.2.3 and [108, 111, 125], or to combine segments from different videos in a single video as shown in Section 4.5 and [40]. Viewing history helps in reducing the time needed for creating video mash-ups and produces a good quality video summarization.

4.2.4 Sharing

Over the last few years, social media sites have become the most popular visited websites where 74% of all internet users are now active [1]. The key attraction of these social sites is sharing. People have become more interested in creating, publishing and sharing their own experiences with their families, friends, and fans. On average 68% of viewers share a video they have watched. Some social media sites (e.g. YouTube and Vimeo) have even tried to make it easier to share a video from a specific starting time rather than the beginning of the video. Video viewing history makes it even easier and faster where users can share a specific interval from a video rather than just specifying the starting time. It is just a simple click on the share button for that specific segment, as will be shown in Chapter 8. Moreover, as described in the previous section, a user can share video mash-ups or summaries that are generated using their own personal viewing history or the collective wisdom. Users can even share how they experienced a video by sharing their consumption history of that video. Having this ability allows users to compare their viewing behaviour with friends or even the collective wisdom.

4.3 Common Properties of Video Navigation Interfaces

In this section, we define the common visuals and their related properties that will be used throughout this dissertation.

[1] http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

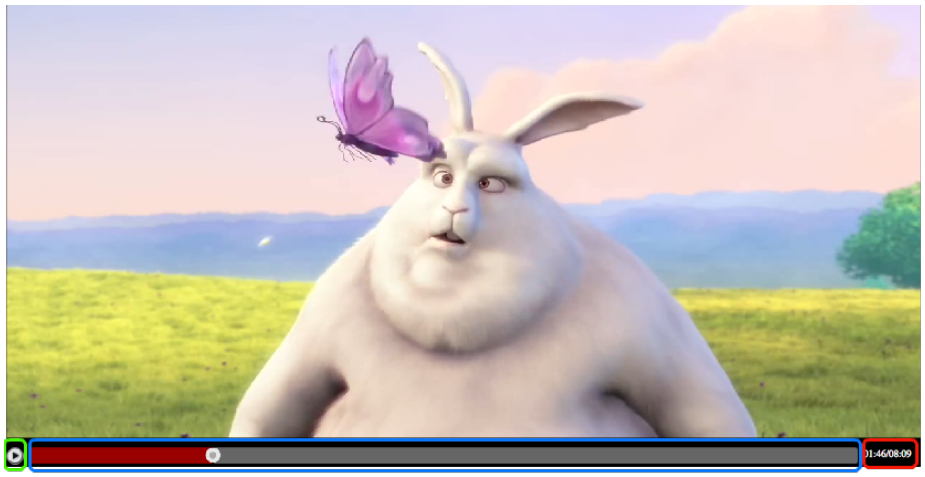


Figure 4.2: The common video player component used in all interfaces presented in this dissertation. It is used for a direct control of the selected video using the play/pause button (, highlighted in green) or via seeking using in the timeline (highlighted in blue). The current playhead time is highlighted in red.

Player: A familiar video player (similar to YouTube, QuickTime, etc.), shown in Figure 4.2, contains: play/pause, applied via a button (bottom left, highlighted in green) or by clicking directly on the video; seeking via playhead (white circle on timeline,) or mouse click on the timeline (red/grey bar, highlighted in blue); and a current playhead time label (white text, bottom right, highlighted in red). This component is used for direct control of the selected video.

Thumbnail: A frame preview that represents the starting frame of an interval watched from a video (Figure 4.3). We applied this design since thumbnails are an accepted form of preview in nearly all digital video retrieval interfaces. A *Thumbnail* can have as many properties from the following as needed.

Temporal Visualization: A gray timeline attached to the bottom of a *Thumbnail* with a red highlight to indicate the location of the interval within the entire video, as shown in Figure 4.3. It helps users spatially con-

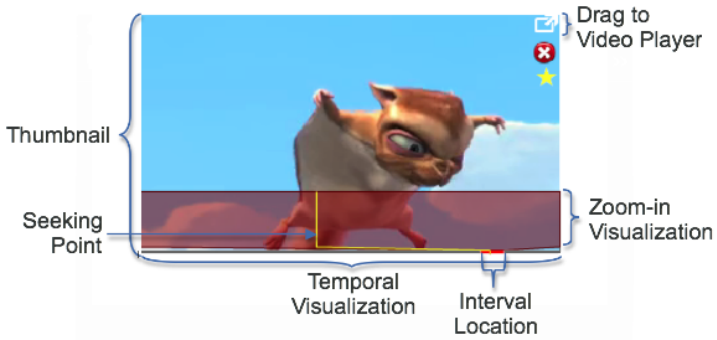


Figure 4.3: Thumbnail used to represent a video segment with its different properties: seek-able, drag-able, delete-able (⊗), and favourite-able (☆).

textualize the temporal location of intervals within the complete video where it came from.

Seek-able: Moving the cursor over the bottom third portion of a *Thumbnail* pops up a zoom-in visualization of the Temporal Visualization, as illustrated in Figure 4.3, which can be used to seek within the video interval using mouse motion as a cue. The seeking point is visualized by a yellow line and the *Thumbnail* updates to reflect the current seek position.

Drag-able: To play the represented interval of the *Thumbnail* in the main video player, users can drag the entire *Thumbnail* (or the seek location in the *Thumbnail*) to the main video player (or click the top right arrow (Figure 4.3): dragging the entire *Thumbnail* plays the corresponding interval from the beginning, while dragging the seek bar location plays from that specific time.

Favourite-able: The interval represented by the *Thumbnail* can be favourited/unfavourited by clicking on its corresponding ☆ button. If a *Thumbnail* is favourited then the corresponding interval will be marked which can be used for filtering video intervals.

Delete-able: Unwanted interval or *Thumbnail* can be removed from users' viewing history by clicking on its corresponding ⊗ button.

Play-able: The interval represented by a *Thumbnail* can be played within the *Thumbnail* itself (i.e. small player). On mouse over, a play/pause overlay is displayed over the *Thumbnail*, as shown in Figure 4.4. Clicking on the overlay plays the corresponding video interval within this small *Thumbnail* only.

Figure 4.4: A play-able Thumbnail: on mouse over, a play/pause overlay is displayed over the *Thumbnail*. Clicking on the overlay plays the corresponding video interval within this *Thumbnail*.

Size-able: *Thumbnail* has a variable size based on a weight factor that is based on how often an interval has been viewed and how long it is. It reflects the relative popularity of intervals.

View Count Visualization: A vertical red bar attached to the left of a *Thumbnail* visualizing how often the corresponding interval has been viewed in relation to other intervals in the video (Figure 4.5). The taller the bar, the higher the view count for that interval.

Figure 4.5: View count visualization attached to the left of a *Thumbnail* indicating how often it was watched in relation to other intervals in the video.

4.4 Video Navigation Interface

Given the volume of available video data, methods to effectively navigate this space are needed to provide users with more enjoyable video watching experiences. The principle goal of our interface is to provide efficient access to previously viewed video, and in particular, intervals within those videos. We accomplish this by recording a continuous video history for the user as they consume video content.

We chose video authoring and sharing as the context of our interface since they require bookmarking of intended events. We wanted to investigate whether users having their viewing history would ease the video authoring task and whether it would change users' authoring and viewing behaviour.

Figure 4.6: Our Video Navigation Interface introducing users' viewing history in a video interface. The video *Player* (yellow) is adjacent to *Video Grid* (green), which approximates the results of a search tool. The *History Timeline* (bottom in blue) provides the history based on the user's viewing time, in a scroll-able one-dimensional filmstrip interface. The vertical red bars on each thumbnail represent how often this particular segment has been viewed. The *Video Mash-up* (red) represents a video edit list by visualizing the videos and the intervals a user may combine into a summary.

4.4.1 Visualization Components

Our video navigation interface, shown in Figure 4.6, was developed in Flash CS4 using Action Script 3.0. It is based on five separate components, two of which provide the user with a familiar video setting (player and video selector), a third

visualizes the personal viewing history using a filmstrip metaphor; and two facilitate creating video summaries and edit lists, and sharing video histories.

Player

A familiar video player, as described in Section 4.3, facilitates direct control of the selected video. This player has an extra property, a frame preview (i.e. *Thumbnail*) triggered by the mouse over the timeline (including when seeking interactively i.e. with mouse button held down over time slider), as shown in Figure 4.7. This preview offers users a visual cue of the content at specific temporal location, hence, ease a search task.

Figure 4.7: The video player used for a direct control of the selected video. When the cursor is over the time bar a thumbnail for that instant is shown to help the user navigate the video.

Video Grid

The Video Grid, shown in Figure 4.8, is a very basic component, which displays the available videos to the user. We use this as an example of what could be returned from search results, video suggestions, or any other method of delivering videos to users. Users may select a video from this grid by clicking on it, and the video will start playing in the player component.

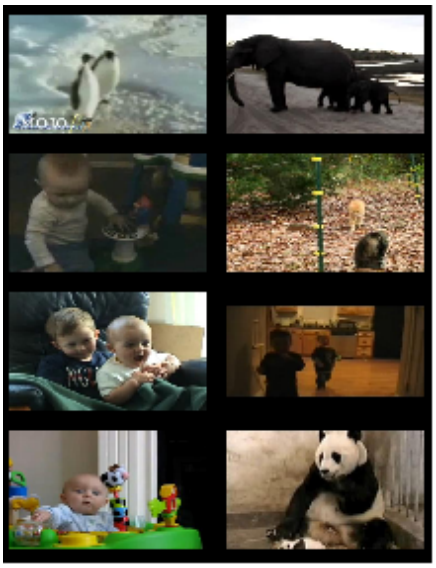


Figure 4.8: The video grid, which approximates the results of a search tool. Clicking on a video from a grid will start playing the corresponding video in the player component.

History Timeline

The history is the central component for visualizing a user's video viewing history. We use a filmstrip metaphor for visualization: the History Timeline consists of a set of thumbnails (similar to interactive storyboards) with view count visualization as shown in Figure 4.9. As the user consumes video, new thumbnails are added on the right as older items are pushed off the display to the left. The History Timeline is scroll-able, to provide navigation of the entire viewing history and allows the user to find a specific interval. Each interval is represented by a different rounded rectangle and each video has a different colour to make it easier for the user to differentiate the videos. Every time the user watches a video clip, a new rounded rectangle is added to the history, encoding the start and end of that video interval in the history. Thus, every seek or video change results in a new record added to the history.

The video history is similar conceptually to webpage bookmarks that are used to refer back to visited pages; however, it is more complicated, as each Video Seg-

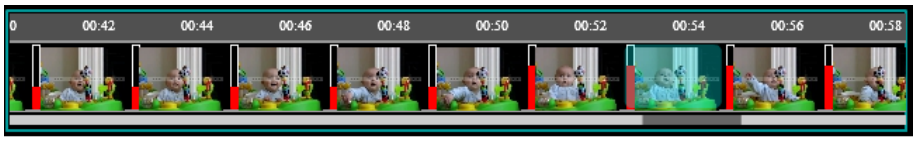


Figure 4.9: The History Timeline provides a visualization of users' viewing history based on their viewing time, in a scroll-able one-dimensional filmstrip interface. The vertical red bars on each thumbnail represent how often this particular segment within the video has been viewed.

ment in the history view is a video interval and also includes a representation of the time the user watched it. The history is unique in that it records in piecewise linear user-time, not video time (we only display the history of intervals watched and do not represent the complete video from where it came from in the History view.). The timecodes of user time are displayed above the thumbnails; each thumbnail represents a small portion of time (typically around two seconds).

The History Timeline also provides the user with the most-viewed intervals of each video (a temporally tracked frequency). A view count is maintained for all points of time for each viewed video. The view counts are summed and normalized for each video and shown in the interface as a vertical red bar to the left of each thumbnail. The height of the bar corresponds to the view count of that interval: the taller the bar, the higher the view count for that segment. Alternatively, these bars can be made to show different things, like letting users change the ranking based on time spent, liked, more actions made, more re-watched, etc.

As the user watches the video, the view count will dynamically update. This information propagates back to previously generated thumbnails. As the user reviews their personal video history, they can see the most popular intervals based on the view count. They can click on any thumbnail to begin playback at the corresponding time within the video; the video will play until the ending time of the last thumbnail in the selected interval.

Video Mash-Up

The Video Mash-up is used for short video composition such as summaries, previews, or trailers (analogous to video editing software e.g. iMovie, Movie Maker,

Figure 4.10: The Video Mash-up component illustrates a user edit list which consists of a user-defined Video Segments with the options of delete (X), re-order (☰), play (▶), and modify segment's start and end times.

etc). We included this component in our prototype to evaluate one of the use cases we identified that takes advantage of potential strength of having personal viewing history. The video mash-up, shown in Figure 4.10, represents a video edit list that shows the videos, and intervals within each video that a user may combine into a summary. In summary creation mode, when the user subsequently plays a video (either through opening a new video, a segment from history, or seeking within the current video) a new record is added to the edit list. If the video already exists in the edit list, a new interval is added (unless there is an existing identical interval). If the video does not exist, a new video record with the interval is added to the edit list. Thus, adding segments or intervals to the edit list is simply done by watching these intervals, in comparison to other video editing software where users need to add a video clip, watch it, and then trim it. Using this component makes video editing simpler for novice users as we discovered upon evaluation.

The video mash-up component allows the user to preview a specific interval by using the play button (▶) for that interval. Users can also modify their edit list by directly updating interval start or end times, delete an interval (X), delete an entire

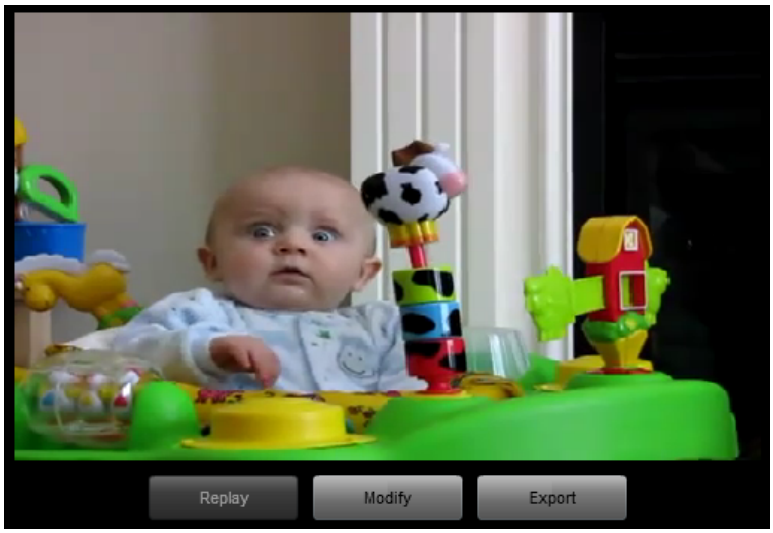


Figure 4.11: The Summary Preview allows users to watch the video summary they have created. Users can re-play, modify, or export their edit list.

video record (X), or reorder video records (). Previewing the entire summary is provided by clicking on the "Preview" button.

Summary Preview

After creating a video mash-up, the user can preview it using the Summary Preview mode, shown in Figure 4.11, where a new video player window pops up, and the rest of the interface is disabled and darkened. The summary preview takes the user's edit list as input and plays the intervals in order using this content. It allows users to replay, modify (back to the main interface to modify the edit list), or export it if they wish to save it - the edit list is exported as an XML file that can be shared or modified later.

Personal Navigation Footprints

The recorded video history offers additional benefits, such as within-video history and crowd-sourced user histories (neither of these were included in our study). The personal history or footprints offers users a visualization of the view count of every point of time in the video: the blue bar in Figure 4.12 becomes brighter for the

most-viewed intervals, and each one can be selected to play. It provides the user with information on the most-viewed intervals of each video (a temporally tracked frequency). As mentioned previously, a view count is maintained for all points of time for each viewed video. As the user views a video the view count will dynamically change, thus the count is continuously updated and the information propagated backwards through the history. The view counts are then summed and normalized within each video.

Figure 4.12: A Video Timeline is combined with navigation footprints visualization, where the more an interval within a video is watched the brighter that region becomes. Blue: a single user history (Personal Navigation Footprints), Red: a combined multiple users histories (Crowd-Sourced Navigation Footprints).

Crowd-Sourced Navigation Footprints

The personal histories from multiple users are combined to form the most popular intervals for a video (i.e. social navigation similar to these are shown in the red bar in Figure 4.12). This provides users who are watching a video for the first time with a mechanism to watch what others have found to be most interesting. This is an efficient navigation technique, predicated on trusting the crowd to be "correct" on interesting video content.

4.5 Investigating the Feasibility of Video Viewing History

In order to investigate the feasibility of video viewing history, we designed the video navigation interface, described earlier in Section 4.4, which keeps and visualizes users' personal navigation history. We wanted to evaluate the utility of the video viewing history using our interface, thus, we designed a formative study (approval certificate #: H08-03006) inspired by the studies performed by Grossman et al. [49] and Leftheriotis et al. [73]. The study evaluated the design aspects and performance of using personal video navigation histories, and investigated whether having a viewing history would ease and speed up the creation of a short trailer

from previously navigated videos. The experiment tasked participants to create a short trailer from a set of videos based on a predefined theme.

We hypothesized that:

H1: participants will create the short trailers faster when they have access to their personal viewing history.

H2: less trailer editing will be required when using viewing history to select clips for trailers.

- **H2.1:** less clip deletion will exist when clips are added from viewing history
- **H2.2:** less clip play or preview will exist for trailers created using viewing history
- **H2.3:** more timestamp modification will be needed for clips not added from viewing history

H3: large variation will exist between participants for the clips and videos used in each trailer

H4: participants will need time to understand viewing history: what the visualization means and how it works

4.5.1 Apparatus

The user study was conducted on a Toshiba laptop with a 2.10GHz Core2 Duo CPU with 2GB of RAM running Windows XP Professional. The laptop LCD display was used at a resolution of 1280×800 pixels at a refresh rate of 60Hz. A Microsoft optical wheel mouse was used as the input pointing device and the Adobe Air environment was set to 1260×430 pixels while running the Flash program.

4.5.2 Participants

Twelve paid university student volunteers, 6 female and 6 male, participated in the experiment. Participants ranged in ages from 23 to 35; all were experienced computer users and have either normal or corrected-to-normal vision. Participants reported watching videos either on a daily basis (ten participants) or 3-5 times a

week (two participants). Only six of the participants have previously used video editing software (e.g. iMovie, Movie Maker, Flash, etc.), and all stated "rarely used". Each participant worked on the task individually. Table 4.1 shows participants' demographics.

This project targets heavy video users, and according to Purcel[2], video is a growing trend especially amongst those aged 18 to 29. Thus, students are a primary target audience as they are representative of the current/near-future heavy users of video interfaces. We used students across the university to capture this demographic.

4.5.3 Design

Using our interface described in Section 4.4, participants were asked to create six trailers based on different themes using eight different videos where they tried three different modes. The different modes, trailer themes, and the set of videos used are described below.

Tested Modes

Three different interface strategies were tested:

- List selection (*Grid*) mode: participants were presented with the Video Player, Video Grid, and Video Mash-up components to construct a trailer. This mode represented the case where there is no access to the user's viewing history. This was the control method, which represented the case where a user has access only to the viewed videos similar to what they have in a social video website (e.g. YouTube).

- Personal Navigation History (*History*) mode: participants were presented with the Video Player, History Timeline, and Video Mash-up components to create a trailer. In this mode, users had access to the video intervals they had watched only. The History mode provided access to every interval or segment within videos a user had watched (i.e. detailed viewing) unlike

[2]http://www.pewinternet.org//media//Files/Reports/2010/PIP-The-State-of-Online-Video

Table 4.1: Demographics summary for participants in the investigating the feasibility of video viewing history study.

P	Gender	Age Group	Experienced Computer User	Watching Frequency	Have Used Video Editing Tool	Used Editing Tool	Usage Frequency
1	Male	19-24	Yes	Daily	Yes	Movie Maker, Flash	Rarely
2	Female	25-30	Yes	Daily	No	-	Never
3	Female	25-30	Yes	Daily	Yes	Movie Maker	Once a month
4	Male	31-35	Yes	Daily	Yes	Final Cut Pro, iMovie, Motion, Flash, Adobe	Rarely
5	Male	19-24	Yes	Daily	Yes	iMovie	Rarely
6	Male	25-30	Yes	Daily	No	-	Never
7	Male	31-35	Yes	Daily	No	-	Never
8	Female	25-30	Yes	Rarely	Yes	Movie Maker	Rarely
9	Male	25-30	Yes	Daily	Yes	Movie Maker	Rarely
10	Female	25-30	Yes	Once a week	No	-	Never
11	Female	25-30	Yes	Daily	No	-	Never
12	Female	25-30	Yes	Daily	No	-	Never

the Grid mode where users only have access to the videos. This mode was included to investigate the effect of history separately.

- A combination of the above (*Hybrid*) mode: Both the Video Grid and the History Timeline from the two previous modes were given to the user in this mode as shown in Figure 4.6 along with the Video Mash-up to construct a trailer. Users had access to their viewing histories and the entire videos were provided, i.e. they had access to watched and unwatched video intervals. This mode represented our proposed interface.

For all modes, participants used the Summary Preview component to preview their trailers before exporting.

Each participant tried each of the three modes (Grid, History and Hybrid) to create a 15-second trailer using four different clips for each trailer. Participants were divided into two groups of six where each group has a different order of the first two modes (Grid and History). We kept the Hybrid mode last for both groups since it is a combination of the other two modes.

Trailer Themes

Six different trailer themes were tested:

1. serene animals;
2. loud noises (sneeze, cry, laugh, etc);
3. funny parts;
4. two objects (either two humans or two animals);
5. the child or animal gets scared;
6. background laughter.

The six trailer themes (two for each mode) were tested and both groups kept the same order, but the order for the two contrasting modes was changed to eliminate the mode order effect. Thus, group 1 created the serene animals trailer using Grid mode while group 2 used History mode to create the same trailer.

Experimented videos

A set of eight different short videos of length between 14 seconds and one minute (sourced from YouTube, chosen based on total number of views) were used in this experiment. These videos are:

V1: Baby scared by his own fart![3]

V2: Baby elephant sneezes and scares himself[4]

V3: Funny baby reaction to light[5]

V4: Cat gets shocked by electric fence[6]

V5: The sneezing baby panda[7]

V6: Charlie bit my finger - again![8]

V7: Emerson - mommy's nose is scary![9]

V8: Evil penguin[10]

Once all the videos were viewed the participants were instructed to create a trailer based on a given theme. They created six different trailers (3 modes $\times$ 2 trailers per mode). The participants were asked to construct the trailers as quickly and accurately as possible. For each trailer, the navigation heuristics, the completion time, the final edit list of the trailer, the number of deletion tasks, and the number of modifications were recorded. The completion time was measured from when the participant clicked on the 'Start' button (after reading the task) until the moment the participant clicked on the 'Export' button after previewing the trailer. The participants were also asked to fill out the post-experiment questionnaire shown in Section B.2 to give their comments on the interface, feelings

[3] http://www.youtube.com/watch?v=zYV5BX0kcMc
[4] http://www.youtube.com/watch?v=gtlz1u8g1F0
[5] http://www.youtube.com/watch?v=xjMSPgw0QyI
[6] http://www.youtube.com/watch?v=Z8DqqPto9Iw
[7] http://www.youtube.com/watch?v=FzRH3iTQPrk
[8] http://www.youtube.com/watch?v=_OBlgSz8sSM
[9] http://www.youtube.com/watch?v=N9oxmRT2YWw
[10] http://www.youtube.com/watch?v=A2fntzGquZE

on usability, and suggestions for improvement. They were asked to rate different interface features and tools using a 7-point Likert scale, "I found it easy to use" and "I think it would be useful",(1 = strongly disagree, 7 = strongly agree). The experiment lasted approximately one hour per participant.

4.5.4 Procedure

The experiment proceeded as follows:

1. The researchers gave the participants a walk-through of the interface explaining the functionality of each feature and tool within the interface and their effects when applied to a video. The participants were shown how to use the features to create a short trailer. This stage took approximately five minutes on average. The participants were allowed to ask any question during this stage.

2. Participants created a trailer using the same theme and interface to become familiar and comfortable in the application of its features and tools. They were allowed to freely view, apply features and control the playback of eight different videos. These eight videos were not included in the actual experiment. Participants were also allowed to freely ask any questions during this familiarity stage, which took about 10 minutes.

3. The first step of the experiment started by asking the participant to completely watch eight different videos, described earlier in Section 4.5.3, and re-watch any parts they like. A single video was played upon the participant's selection (never more than one video playing simultaneously). Participants were allowed to freely select the order of the videos and re-watch any parts from videos. However, they could not start creating trailers unless they had watched the entire eight videos to make sure that they had created enough viewing history for each video, and all participants had seen equivalent video. Participants were given extra time (about three minutes) to play back different segments, seek to different parts within a video and create a condensed navigation history based on their interest. This navigation history was stored to be used for the History and Hybrid modes. At this stage of

the experiment, the participant did not know anything about the theme of the trailers, in order to create a blind navigation history.

4. Once all videos were viewed, the participants started the tasks by clicking on a button to create a trailer. The theme of the trailer and the mode used were provided to the user in a popup message, including a "Start" button. By clicking on the start button, the interface was modified to display components that are related to the task mode. The Video Mash-up component was shown to the user to start adding different clips. Upon the completion of the trailer (by clicking the export button after previewing the trailer), the participants advanced to the next trial to create another trailer using the same set of videos but with a different theme.

5. The experiment ended once the participants had submitted all the six trailers using the different modes. Finally, the participants were asked to fill out the post-experiment questionnaire shown in Section B.2.

4.5.5 Results and Discussions

The study showed positive results of the features of the interface. There were some overwhelming comments made by the participants about the interface. Most participants commented that they enjoyed using the interface and they could imagine seeing its features applied, especially in social networking websites like YouTube. They were impressed with the ease of history, finding clips from history and creating a trailer.

Task Performance

For the trailer creation, each participant was able to complete all the six trailers in under 6 minutes per trailer. On average, participants were faster in creating trailers when they used their viewing history compared to the Grid mode when they did not have access to the history as shown in Table 4.2. This result confirms our hypothesis **H1**. Providing participants with both modes (i.e. Hybrid) significantly outperformed the two modes. The mean completion time for each trailer theme using the different modes is illustrated in Table 4.3. For 'Serene animals' and

Table 4.2: Means (and standard deviations) for number of times video Grid used, number of times History Timeline used, completion time in minutes, number of clip previews, number of clip deletions, number of clip modifications and number of trailer plays for the different modes. When given the hybrid mode, participants created trailers significantly faster than using the two modes separately, which illustrates the utility of navigation history in video authoring. ($^{*}p < 0.005$)

	Grid	**History**	**Hybrid**	**F-test**
Grid Usage	5.45 (3.06)	- -	1.96 (2.05)	-
History Usage	- -	5.33 (2.18)	2.79 (2.45)	-
Completion Time	4.08 (1.39)	3.44 (1.44)	2.67 (1.27)	6.44*
Previews	1.88 (1.03)	1.63 (0.82)	1.71 (0.95)	0.44
Deletion	4.13 (2.94)	2.33 (3.57)	2.46 (2.34)	2.68
Modifications	17.83 (9.81)	15.63 (11.03)	11.13 (7.73)	3.03
Plays	6.96 (4.79)	5.58 (5.51)	3.71 (4.82)	2.51

'Two objects', participants took more time creating them using the Grid mode. As can be seen, when the Grid mode was faster, it was only few seconds faster than the History mode; however, when it was slower it was at least one minute slower. This shows that History was much better than Grid, which indicates that viewing history sped up video mash-up.

In terms of trailers' editing, as shown in Table 4.2, there was no significant difference between the modes in term of number of previews, deletion, modifications, and plays. The means per trailer and mode is illustrated in Table 4.3. The insignificance in the performance may be due to the small number of tested themes

Table 4.3: Means of number of times video Grid used, number of times History Timeline used, completion time in minutes, number of clip previews, number of clip deletions, number of clip modifications and number of trailer plays for themes per mode. For 'Serene animals' and 'Two objects' trailers, participants took more time creating them using the Grid mode.

Theme	Mode	Grid usage	Timeline usage	Time	Previews	Deletion	Modifications	Plays
serene animals	Grid	6	-	5.47	2	3.83	23.17	6.5
	History	-	6	3.83	2	0.83	17.83	10
loud noises	Grid	4.17	-	3.84	2.33	3.83	14	4.83
	History	-	6.17	3.91	1.5	4.5	17.33	6.67
funny parts	Grid	4.83	-	3.58	1.5	5.33	23	10
	History	-	4.67	3.74	2	0.8	17.17	4.33
two objects	Grid	6.83	-	3.44	1.67	3.5	11.17	6.5
	History	-	4.5	2.26	1.17	0.5	10.17	1.33
gets scared	Hybrid	1.5	3	3.13	2	2.17	11.83	5.08
background laugh	Hybrid	2.42	2.58	2.21	1.42	2.75	10.42	2.33

where we had only two per mode and the unfamiliarity with the viewing history and its usages. Moreover, even though participants were asked to create these trailers as quickly as they could, they felt that it was a subjective task and they wanted to spend more time on it to get almost a perfect trailer out of the provided videos. This could also be seen from the significant positive correlation between the completion time and the time the participants spent in previewing the trailers several times, playing each interval, deleting intervals, and tuning intervals to an accuracy of 0.1 seconds. One of the participants tried to extract clips from almost all the videos that are related to the theme, not just taking the four clips as asked.

Nevertheless, these results show that having access to the personal viewing history positively impacts the video mash-up task. The access to viewing history also changed users' authoring behaviour, where participants used History Timeline to select clips for the trailers in the Hybrid mode more than accessing each video and scrubbing it to select the intended clips. This shows the promise of providing the viewing history and its tendency of changing the way users view and author videos.

Modes' Ranking

We asked participants to rank the different modes for ease and speed in creating trailers. They ranked the Hybrid mode as the easiest while Grid and History had almost the same ranking (both came second). This might be due to the fact that the Hybrid mode gave participants access to the components of the other two modes. This gave them two options to find clips: either by navigating the entire video or searching from their personal history. When it was easier to remember the location of the clip within the video, participants preferred to navigate the video (i.e. using the Grid mode's components), whereas otherwise they used the history. When asked which mode for creating trailers was the fastest, they ranked both Hybrid and History as the fastest modes with Grid coming last, which coincides with the quantitative results, illustrated in Table 4.2, where a significant difference ($F(2,69) = 6.443, p < 0.005$, effect size $= 0.157$) of the completion time existed between modes. A post-hoc pairwise comparison showed that Hybrid and History were faster than Grid. This was because the time needed for participants to find

Table 4.4: Clips agreement percentage per trailer theme. Background laughter trailer showed the highest agreement while 'two objects' trailer showed the lowest.

	Clip1	**Clip2**	**Clip3**	**Clip4**
Serene animals	66.67	50	91.67	66.67
Loud noises	75	75	66.67	66.67
Funny parts	58.33	66.67	75	50
Two objects	41.67	50	41.67	75
Gets scared	66.67	66.67	66.67	91.67
Background laughter	91.67	100	100	100

clips using these two modes was faster than navigating the entire video since they bookmarked the events. It could be also because most participants used the history component more frequently than the grid to select clips in the Hybrid mode as shown in Table 4.2. Participants commented that they liked the interface when both modes are available (Hybrid), to have more freedom to use any feature they like.

Trailers' Agreement

The generated trailers were analyzed to check the agreement of the individual included clips between participants and how often each video was used. Table 4.4 shows the percentage of agreement between participants for each clip used in each trailer, while Table 4.5 illustrates the percentage of agreement for each video used in each trailer. Surprisingly, a high percentage of agreement existed between participants, which contradicts with hypothesis **H3**.'Background laughter' and 'Serene animals' trailers had 100% agreement of videos used, even though there are 6 videos that contain background laughter. 'Background laughter' trailer had also 100% agreement for clips used except for clip1, as shown in Figure 4.13, which was a result of one participant's interval added from V1 that did not match other participants' intervals. 'Funny parts' and 'Gets scared' trailers showed a large variations in the video used which can be explained by the variation in personal perception of events to be funny or scary. The lowest clips agreement was found in the 'two objects' trailer which was due to the fact that videos that contain this

Table 4.5: Videos agreement percentage per trailer theme. Background laughter and serene animals trailers showed 100% agreement of videos among participants.

	V1	V2	V3	V4	V5	V6	V7	V8
Serene animals	0	100	0	100	100	0	0	100
Loud noises	75	75	75	25	41.67	66.67	50	0
Funny parts	25	58.33	66.67	16.67	75	41.67	58.33	66.67
Two objects	8.33	41.67	58.33	33.33	58.33	91.67	8.33	100
Gets scared	66.67	66.67	16.67	66.67	91.67	8.33	83.33	0
Background laugh	100	100	100	100	0	0	0	0

theme have the two objects appear in several parts of the video which caused the variation of the clips chosen.

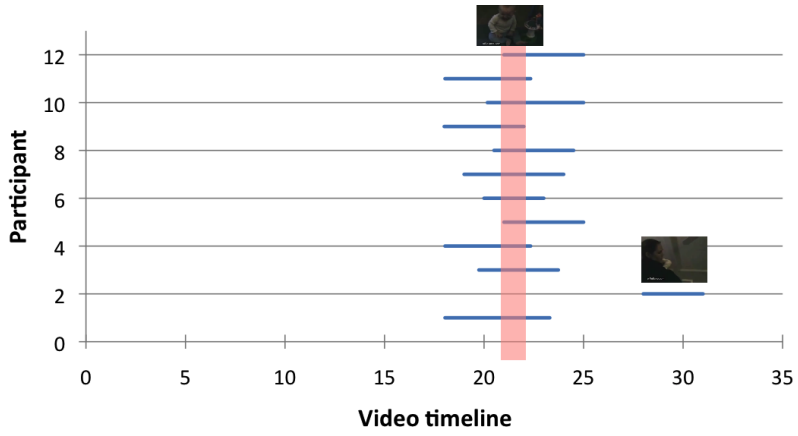


Figure 4.13: Participants' clip used in the background laughter trailer. All participants used the same clip with some tolerance at the start and end times of the clip except participant 2 who used a different interval.

This high agreement between participants in the clips and videos they used

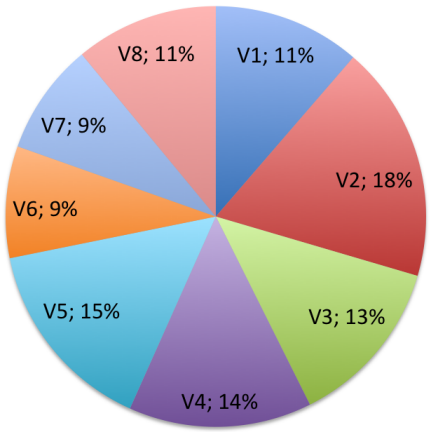


Figure 4.14: Percentage of clips extracted from each video. Most of the clips in the trailers came from V2 (Baby elephant sneezes and scares himself) while V6 (Charlie bit my finger) and V7 (Emerson - Mommy's Nose is Scary) were the least used in the trailers.

shows a promise of using the collective wisdom to generate video mash-ups. It will help reduce the time needed for someone to create them where it can be automatically generated using some smoothing algorithm for the collective data.

Videos' Usage

Figure 4.14 illustrates how often each video was used to create the 6 trailers. Most of the clips used in the trailers came from V2 (Baby elephant sneezes and scares himself) since it has scenes that match all themes used for the trailers. While V6 (Charlie bit my finger) and V7 (Emerson - Mommy's Nose is Scary) were the least used in the trailers because they only match specific themes. For example, V6 can be used for 'two objects', 'loud noise', and 'funny parts' trailers, while V7 can be only used for 'funny parts' and 'gets scared' trailers. Moreover, the variations in personal taste and perception plays another factor on which clips and videos being used for each trailer.

Features' Ranking

We found positive results from the questionnaire in relation to the ease and usefulness of the interface components and features. The average ranking across all components and features was 6.2 for 'would be useful', and 5.8 for 'easy to use' (all scores are out of 7).

Least useful features: Two features scored less than 6; all others were graded as highly useful (more than 6 out of 7). The two which scored lowest were: the use of different colours to differentiate between videos in the history timeline (5.1 out of 7) and the different rectangles used for each segment in the history timeline (5.5 out of 7). This might be due to the clear differentiation between the tested videos, which made these features less useful.

Most useful features: In the Video Mash-up component: modify clips timecode (6.84) and remove clips (6.5). In the Summary Preview component: replay the trailer (6.84) and modify trailer (6.5). In the History Timeline component: play specific segment (6.4) and create trailer (6.3). In the Video Grid component: select a video to play (6.6). Finally, in the Video Player component: play/pause (7), seek (7) and frame preview (6.9). For easiness, almost all the features were easy to use except the editing of the clips' timecodes in the trailer edit list component (4.5) since participants had to type these timecodes.

Participants' Suggestions

Participants suggested that dragging intervals from the History Timeline to the Video Mash-up to add clips might be easier for them instead of re-watching the clips to be added to the trailer. They also liked to have the dragging interaction to order clips rather than using the corresponding clip's up/down buttons (). Participants found it frustrating when new clips were added to the trailer edit list every time they used seek in the video. One of the participants said, "I would like to have control of when these clips are added rather than going back every time and delete the clips which I had not originally added". Allowing for enabling and disabling the trailer edit list component can mitigate the problem of automatic addition of

clips to the list.

For the history, at the beginning when viewing and navigating videos, it was difficult for participants to understand how the history is constructed and when thumbnails are added. Some participants found the addition of new segments to the history strip every time they used seek in the video a bit confusing and impacted the understanding of their history. However, as reported in the results by the time when participants experimented with the hybrid of both modes, they tended to use the history more often. Moreover, some participants stated that the history is fun and easy to use once you get familiar with it. This confirms our hypothesis **H4** and we anticipated that, since it is a new concept and users have not had any earlier exposure to it.

In terms of the history visualization, the size of the thumbnails was criticized as too small for some participants, which hampered content recognition. Most of the participants suggested using larger interface components. Some participants found having a horizontal component for the history confusing, since it represents a different timeline from video. To separate the concepts, we need to apply a design guideline such as using a vertical visualization for the history and horizontal for video. This may also make scrolling faster since it is the norm in most long lists. Lastly, participants suggested having control over the vertical red view count bar attached to each thumbnail in the history. They would like to decide on the metrics used for these bars instead of using just the frequency. This will enhance the ability to filter the different thumbnails. Some metrics that may be used are: time spent, liked segments, number of actions, users ranked, or mostly watched.

Participants' Comments

There were some overwhelming comments made by the participants about the interface. One participant commented that, "I definitely see how this would be really helpful for long videos because I will not have to waste my time watching the whole video again to get to the important stuff. I could directly use my previous history to navigate to these intervals." Others said, "It is really cool and easy to create different trailers that I could share with my friends"; "I need to have this. Could we have it in YouTube?"; and finally "I really do like this interface and I

would love to have it to create wonderful clips from my home videos."

Based on the results and the participants' valuable comments and suggestions, we believe using implicit bookmarking through a personal viewing history is helpful for navigation of video spaces. It can be applied to different applications such as video highlights, video summaries, authoring using multiple videos, sharing interesting clips, quickly navigating and skimming previously watched videos, and watching new videos' interesting parts using crowd-sourced viewing histories. Participants foresee that our interface would be valuable in social networking contexts such as online video sites.

4.6 Lessons Learned

We have learned a great deal from this user study. Primarily, the introduction of viewing history to users changes their behaviour in viewing and video mash-ups creation. More exposure to the concept will introduce a new way of watching, navigating, and enjoying videos from which future video viewing behaviour will emerge. Many other lessons have been learned about the design of the interface, which we employed to formulate the design guidelines presented in Section 4.6.1. For example, visual orientations can introduce confusion for some users; for instance, the horizontal user's timeline (i.e. history timeline) was confused with the video timeline. Also, when the content of videos is distinguishable, then there is no need to introduce extra cues to differentiate between videos because it is considered as clutter. Size of thumbnails plays an important rule in recognizing their content and users tend to like them bigger. However, there is a trade-off between size, organization of components and interface clutter, which can be used to assess an acceptable size. In terms of the interactions, they need to be simple and effortless. Examples of simplifying include drag-and-drop interaction to add clips to users' edit lists instead of the necessity to watch these clips, and for ordering clips instead of using specific buttons, as well as using '+' and '-' buttons for editing users' clips to control timecodes for the clips instead of just being able to type them. Users want more control over how the data is filtered. For example, they would like to decide on the metrics used for the view count visualization instead of using just the frequency. In relation to the user study design, to illustrate the

difference between methods, more themes need to be tested, more tasks need to be performed and more participants need to be recruited.

4.6.1 Design Guidelines

In this section we list some design guidelines that are formulated based on the results and feedback from the user study presented earlier in Section 4.5.

- Use horizontal orientation to indicate video timeline and vertical for user's timeline (i.e. history).
- Following the previous guideline, apply the component's width to those that are representative of videos to indicate the video length or the representative interval length. For example, the width of a thumbnail represents the length of the interval being represented; thus, seeking that thumbnail will only seek within the representative interval.
- Use considerable thumbnail size to recognize its content.
- Never play more than one video at a time to avoid distraction and to avoid overwhelming users by the interface.
- Use drag-and-drop interaction to move thumbnails or clips.
- Use interactive thumbnail instead of a set inactive images to convey more information using a smaller space.

4.7 Directions

Our preliminary results point to a few promising directions. Since using history to create videos' mash-up or abstraction showed encouraging results, we plan to investigate its performance when compared to other authoring tools. This motivates us to explore various tools that can be proposed using history to improve task performance. Thus, we are going to look at the different interactions that can be offered to easily access, manipulate and reuse the viewing history.

4.8 Summary

To summarize, we have developed a video-viewing interface to provide users access to their viewing history as well as a platform to navigate, play, and generate videos. Previously proposed interfaces did not offer users accessibility to what they have seen before, aside from just showing users their footprints in a video. Moreover, the interfaces that utilized video viewing history kept this data from their users. Our contribution of offering transparent access to what users have viewed is the ability to assess how it can change the way users view videos and what kind of applications can emerge based on users appropriation of this data.

The interface defines a new way to navigate a video space using a user's own personal history, which provides a new mechanism for consuming media content. In this chapter we have presented how to structure a video navigation history to facilitate later reuse and sharing. An interface was described with different visualization components that use the history representation structure. Using the interface and the structure, we provided four different applications where viewing history could be efficiently applied. However, there are different aspects that need to be taken into consideration in relation to the interface. The interface applied the interactive filmstrip metaphor, which we know sacrifices screen space as it gets longer and condenser. Visualizing large sets of data is a well-known problem that has received a significant body of work in the literature. We need to explore visualizing the history by applying some of these mechanisms against the usability of the history. The uses of the history are not limited to the provided applications. For instance, it could also be used to analyze users' navigation behaviour, and extract some entries of the history to be re-executed.

Evaluating our interface showed positive results and we received highly positive comments from participants. This encouraged us to look at how to improve the way a viewing history is presented and in which application it excels. In the next chapter, we describe different visualization designs to represent a single video viewing history, illustrate additional interactions, and investigate how they perform.

Chapter 5

Single Video Viewing History Visualizations

Consuming video online, on mobile devices or on home computers is now a well-accepted form of communication and entertainment, shown by the rapid growth of various providers such as YouTube, Vimeo and Dailymotion. Despite the volume of video available, methods for efficient navigation and sharing of in-video content have not provided users with the ease of use or level of personalization required to accommodate their needs. Constraints such as limiting the length of videos (e.g. six seconds on Vine and 15 seconds on Instagram) can simplify the problem; however, these do not address the challenges with unconstrained video. Part of the problem is that the spatio-temporal representation of video complicates relatively simple actions such as search or selection. Video search often taxes human memory by requiring memorization of a large quantity of previously seen content. In particular, finding and selecting interesting parts has poor navigation and search support. We propose that the addition of a single-video visualization mechanism using viewing statistics will overcome some of these difficulties.

We investigate the usefulness of visualizing prior viewing by either single or multiple users to support fast navigation (to popular or unseen parts), search and directly previewing content, without interrupting normal playback. We envision users will watch videos differently when they have a visualization of their personal navigation: they can implicitly tag segments of video by re-viewing (thereby in-

creasing the view count); it would also capture their natural behaviour, such as watching a funny section multiple times in a lengthy video. This non-linear viewing behaviour is already evident as reported in Chapter 3 and as in YouTube audience retention graphs[1]: videos have peaks in the graphs, implying users watch different content and likely seek to find interesting parts (unfortunately these graphs are not generally public, and require voluntary publication by video owners). Viewing graphs often show a shallow negative exponential curve (i.e cold-start problem) from crowd-sourced data, which can be very simply filtered to highlight the most popular content. Likewise, viewing statistics can be used to filter out videos where only the first few seconds are watched.

In this chapter, we focus on the design of visualizations for a single video viewing history that support fast in-video navigation and quick scene search. This chapter introduces two approaches for these visualizations: (1) using a list of every viewed record (presented in Section 5.1), and (2) using the viewing heuristics as a summary of how the video has been consumed (described in Section 5.2). Section 5.1.1 describes how to visualize each viewed interval in a list of thumbnails, whereas, Section 5.2.2 explains how viewing heuristics can be presented as a summarized Filmstrip. The evaluation of each approach looking at how each visualization performs in comparison to the state-of-the-art is presented in Sections 5.1.3 and 5.2.3. These user studies and their associated methods were approved by the University of British Columbia Behavioural Research Ethics Board [certificate #: H08-03006]. Finally, Section 5.3 discusses the implications of each approach and directions for future refinement.

5.1 Visualize History Using a List of Records

One of the earliest approaches that have been used to represent a browsing history was a linear scroll-able list of the user's traversed components with or without screenshots of the components (i.e. thumbnails) where the most recently visited component is at the top of list. Most web browsers, YouTube, and Netflix use this approach to visualize the history of users' visited content as shown in Figure 2.5, where clicking on one of the thumbnails navigates to the corresponding content.

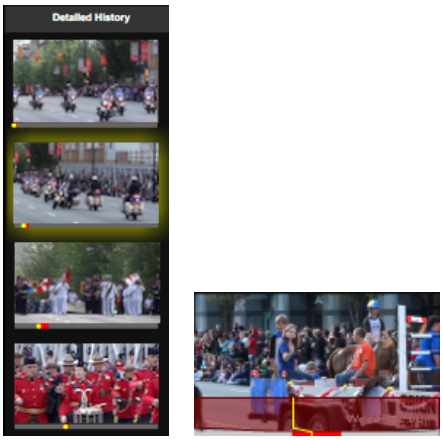
[1] https://support.google.com/youtube/answer/1715160

Thus, due to the familiarity of this visualization among users, we are going to employ it for visualizing a detailed in-video viewing history where every viewed interval from the video is presented as a separate thumbnail.

5.1.1 History Timeline as a Vertical List of Thumbnails

History Timeline uses a Filmstrip metaphor for visualization where it consists of a list of thumbnails (similar to interactive storyboards) as shown in Figure 5.1. These thumbnails represent video intervals that the user watched. We represent this as a *Video Segment* component as discussed in the next section. Every time the user seeks to a new temporal location, a new Video Segment is added to the end of the user's history list and represents the interval a user watches in the main video player. These video segments are ordered in chronological order so that any new video segments are always added to the end of the visualization. We applied this approach instead of the reverse order commonly used in web browsers and YouTube because History Timeline is always visible along with the video player (as shown in Figure 5.2) and it automatically updates as the user watches or interacts with the video player. This was applied to avoid any confusion that could be caused by continuously changing the position of the previously seen segments as the user starts watching any new segment. The History Timeline is a scroll-able box, to provide navigation of the entire viewing history and allow the user to find a specific interval.

The video history is conceptually similar to bookmarks that are used to refer back to visited pages; however, it is more complicated, as each Video Segment in the history view is a video interval and also includes a representation of the time the user watched it. Thus, representing both of these quantities requires additional mechanisms for visualization and interaction. Moreover, as seen in Chapter 4, there may be confusion with the currently viewed timeline and with the history views since two timelines exist: when/what the user watched and the video's timeline. We address this by showing the History Timeline vertically, while the video timeline remains horizontal, consistent with the user's mental model of the video. The history is unique in that it records in piece-wise linear user-time, not video time (we only display the history of intervals watched and do not represent the complete



(a) (b)

Figure 5.1: The *History Timeline* represented as a *video history* (a), made up of individually search-able *video segments* (b)

video from where it came from in the History view). Each Video Segment is a small interactive video widget as described below.

Video Segment

Each interval a user watches is represented as a seek-able, play-able, and drag-able thumbnail with a temporal visualization as described in Section 4.3. Thus, each thumbnail contains the viewed interval only, visualized by the starting frame of that interval. The temporal visualization shows the location of the interval within the entire video to help users spatially contextualize the temporal location of intervals within the complete video where it came from. Each thumbnail shown in Figure 5.1(b) is seek-able and play-able to allow users to easily search within the intervals and minimize the time needed to search for a previously viewed scene. On mouse over, a play/pause overlay is displayed over the widget. Clicking on the overlay plays the corresponding video interval within this small widget only (Note: while technically, this is a re-watching of video, we do not add this activity in the History View.) Moving the cursor over the bottom third portion of the widget pops up a zoom-in visualization of the interval timeline, which can be used to seek

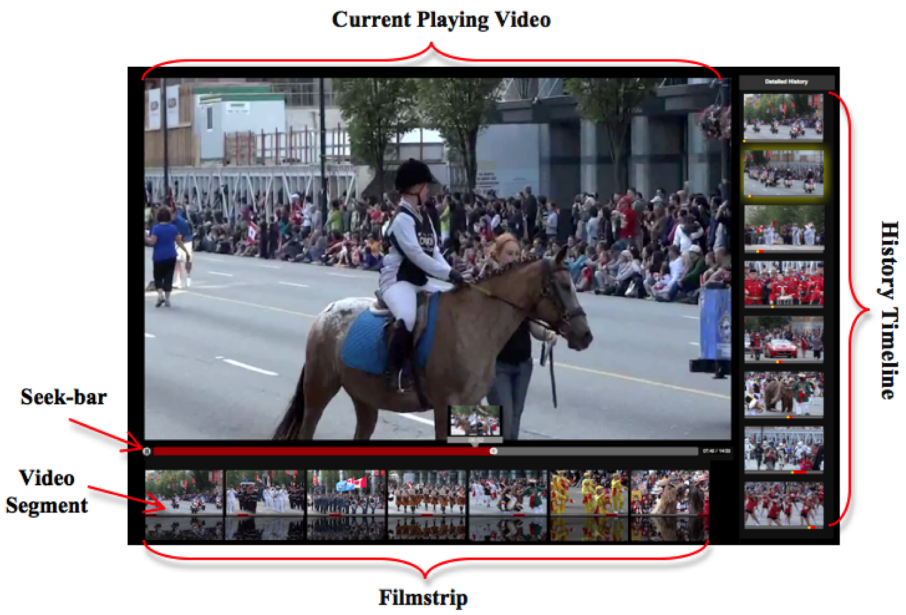


Figure 5.2: Our video navigation interface: the majority of space is devoted to the currently playing video (top left) with a seek-bar preview; below is a horizontal array of Video Segments arranged by video-time (the *Filmstrip*), and a vertical array of Video Segments to the right (the *History Timeline*) ordered top-down by user-time i.e. the order in the intervals were viewed.

within the video interval using the mouse motion as a cue. The seeking point is shown by a yellow line and the thumbnail updates to reflect the current seek position. To play a Video Segment in the main video player, users can drag the entire widget (or the seek location in the widget) to the main video player: dragging the entire widget plays the corresponding interval from the beginning, while dragging the seek bar location plays from that specific time. These interactions can be easily transferred to a touch-screen device supporting today's most common online video consumption platforms.

5.1.2 List of Thumbnails in a Video Viewing Interface

We designed our interface in a way that allows users to play, view and navigate videos similar to any video player as we envision our history management to be

Figure 5.3: The Filmstrip component visualizes the entire video into *n* equal length segments.

used to augment it. As described in Section 4.4, since our interface records the user's navigation history, *History Timeline* was introduced to offer the users the flexibility of using this history for fast navigation. However, based on the results of our previous study explained in Chapter 4, a few modifications have been applied to this component as described earlier in Section 5.1.1.

Our video viewing interface, shown in Figure 5.2, is based on three separate components: a *Player*, which provides the user with a familiar video setting, a *Filmstrip*, which provides a navigation tool for the entire video content, and *History Timeline*, which visualizes the personal viewing history using a vertical list of thumbnails (described in Section 5.1.1). The main interface components are the *Player* and the *History Timeline*.

Player

A familiar video player as described in Section 4.3. This component is used for direct control of the selected video. To watch a specific video from disk, users click on the "Open" button and choose their video file.

Filmstrip

The Filmstrip component, shown in Figure 5.3, is the state-of-the-art video navigation tool. It provides a visualization aid to different parts of a video for faster navigation and supports access to the entire video content. The Filmstrip simply consists of a fixed number of thumbnails (n) from the playing video. The entire video is divided into n equal length intervals, where each interval is represented by a seek-able, play-able, and drag-able thumbnail with a temporal visualization described in Section 4.3. These intervals are created systematically based on the length of the video and the number of segments to be visualized. We applied this

design since thumbnails are an accepted form of preview in nearly all digital video retrieval interfaces. Moreover, the Filmstrip metaphor is commonly used to present content of video as a navigation device, and is considered effective on desktop systems [31], while also providing a quick summary. As such, we chose to employ the Filmstrip metaphor to aid video navigation.

User Viewing History Timeline (History Timeline)

This is the central component for visualizing the personal viewing history of the user. As described in Section 5.1.1, every time the user seeks a video, a new record is generated and hence, a new Video Segment is added to the end of the user's history list to represent the interval a user has watched. The History Timeline (Figure 5.1) updates its content as a user interacts with the video player illustrating when/what parts of the video a user has watched.

5.1.3 Investigating the List of Thumbnails Visualization

An extensive set of pilot studies were performed (12 participants total) to investigate a user's preferred scenario for history, and to triangulate the use cases in which a history is beneficial for video viewing. Throughout this triangulation process it became more apparent to us how complex and intricate the video-viewing task can be. For tasks such as seeking to a specific time or finding a particular event from a previously watched video, using a history was found to be as good as using a Filmstrip methodology. Under more modern viewing patterns (e.g. non-linear viewing behavior) where users view only parts of the video (e.g. trailers, summaries, playlists or direct temporal links) using a history to find events and seek to them was found to be more efficient.

Using this result from the pilot studies, a full comparative user study was performed to evaluate the design and performance of our interface and to demonstrate the utility of personal video viewing histories. We developed an evaluation protocol that satisfies the use cases previously defined (without biasing our design or the control) and provides the user with sufficient viewing history while keeping the experiment relatively short. For fair comparison, we ensured our interface mimics currently adopted approaches with logical extensions.

Using our protocol, we investigated whether visualizing and using a video navigation history would make searching for previously seen events more efficient. We conducted a user study comparing the performance of tasks employing a personal viewing history (History Timeline) against those with the state-of-the-art navigation method (Filmstrip) to find events within a previously seen video (i.e. finding answers to questions similar to [44, 69, 92, 125]).

Both methods have similar layouts and functionality, differing on the process of segment creation described earlier in the description of *Filmstrip* and *History Timeline*. In Filmstrip, the intervals are created systematically, while in the History Timeline, segments are constructed based on a personal navigation history. Each participant tried both methods, on different videos. In Filmstrip tasks, the participants were only presented with the main Video Player and a horizontal Filmstrip component. They only used these components and their features to find the answer to each question. In History Timeline tasks, the participants were presented with the main Video Player and a vertical History Timeline component.

Our experiment was divided into two phases: phase 1 was conducted to compare the two methods (Filmstrip and History), and phase 2 was performed to qualitatively analyze the entire proposed interface. In phase 1, participants used the two components (Filmstrip, and History Timeline) of the interface separately, while in the second phase they were exposed to all components and features.

Apparatus

The experiment application was developed in Flash CS4 and ActionScript 3.0. The experiment ran on an Intel dual-processor dual-core 3 GHz Mac Pro desktop with 8GB RAM and equipped with a 24" Dell LCD monitor with a resolution of 1920×1200 pixels at a refresh rate of 60Hz. A Microsoft optical wheel mouse was used as the input pointing device with default settings and the Adobe Air environment was set to 1920×900 pixels while running the Flash program.

Participants

Twelve paid university students (different from those in the pilot studies), 6 female and 6 male, participated in the experiment. Participants ranged in age from 19 to

30, all were experienced computer users and have either normal or corrected-to-normal vision. Participants reported watching videos either on a daily basis (ten participants) or 3-5 times a week (two participants). Each participant worked on the task individually. Table 5.1 shows participants' details.

Table 5.1: Demographics summary for participants in the investigation of the list of thumbnails visualization study.

P	Gender	Age Group	Major	Watching Frequency
1	Male	26-30	Engineering	Daily
2	Female	26-30	Forestry	Daily
3	Male	26-30	Computer science	Daily
4	Female	19-25	Business / management	Daily
5	Female	26-30	Computer science	Daily
6	Male	26-30	Computer science	Daily
7	Male	19-25	Engineering	3-5 times a week
8	Female	19-25	Science	3-5 times a week
9	Female	26-30	Natural sciences / medicine	Daily
10	Male	26-30	Engineering	Daily
11	Female	19-25	Natural sciences / medicine	Daily
12	Male	26-30	Education	Daily

Design

To evaluate the efficiency of using the list of thumbnails as an aid for efficient search within a previously seen video, we gave participants a predefined history, which they had to watch and subsequently answer questions based on the content. We followed this procedure rather than allowing participants to create their own history. This decision was made based on the pilot studies where participants found creating a history based on an unknown list of questions was confusing. Some participants said, "What interests me might not be what you are looking for." Thus, they tried to create many history segments at each point where they thought there was potential for a question. The result was a long list of short segments that required significant scrolling when searching. In short, the uncertainty of what should be included on the list led to a large number of segments and created some confusion that affected the performance while using the method.

In order to tackle this problem, the participants could be allowed to create their own history based on their interest after which they can be asked questions that exist within these segments. However, the variation between participants' interests would make the comparison difficult. Thus, instead of asking participants to create their own history, we decided to give them all the same history from which to answer the list of questions (analogous to watching a playlist created by a friend or a video summary). All participants watched the same clips from the video and were asked questions only from these clips. Thus, we ensured they had experienced the same clips and eliminated the personalization factor within the history. Our method was compared with the state-of-the-art method (Filmstrip). The History Timeline contained only the segments of the predefined history while Filmstrip contained seven evenly divided intervals over the entire video.

To run this evaluation, a set of videos needed to be chosen so that they meet certain criteria for our study. We looked for videos that:

- were salient enough for everybody so that they would pay attention and would not get bored, which could affect their performance;
- contained enough events of interest on these videos to accumulate history very quickly;
- were relatively short to fit a laboratory controlled study;
- had a narrative structure, which should favor and be fair to Filmstrip.

Accordingly, in this experiment, we used five different short videos that meet the above criteria (**V1**: Toy Story 3 trailer [2], **V2**: Geri's Game[3], **V3**: Partly Cloudy[4], **V4**: Alma[5], **V5**: For the Birds[6]). **V1** was used to explain the interface's components and features, and demonstrate how to use them. The other four videos were used for the actual experiment where each participant used a single method for each video. Each participant experienced each method in a different video to eliminate

[2] http://www.youtube.com/watch?v=fjIj_j4hIc4
[3] http://www.youtube.com/watch?v=dnGBuhu6Txc
[4] http://www.youtube.com/watch?v=HAbiJPGHeV0
[5] http://www.youtube.com/watch?v=irbFBgI0jhM
[6] http://www.youtube.com/watch?v=yJzQiemCIuY

the learning effect. The number of questions per video were: **V1:** 20; **V2:** 38; **V3:** 25; **V4:** 38; and **V5:** 32.

The participants were divided into two groups, **A** and **B**. Both groups experienced the videos in the same sequence but with different method sequences. They both started the familiarity phase where they tried a hybrid of the two methods to answer example questions and to become familiar with both methods. Group **A** was given Filmstrip for **V2** and **V4** and History Timeline for **V3** and **V5**, while group **B** used the opposite. Thus, the results were compared between the two groups for each video.

To quantitatively compare the results of the two methods, we measured six variables for each video:

1. The percentage of questions answered for each video within a specified time range. The total time given is 15 seconds multiplied by the total number of questions (chosen based on the pilot studies);

2. The time needed to answer each question - this is the time measured between two subsequent correct submission clicks;

3. The number of wrong submissions (errors) for each question, which is measured by counting the number of submission clicks that did not result in a correct answer;

4. The number of seeks within a video for each question;

5. The number of previews performed in a video for each question, which is measured by counting the number of seeks and playbacks within thumbnails (either in Filmstrip or History Timeline);

6. The accuracy of the submitted answer, which is measured by comparing the submitted interval with the ground truth interval. It is measured using a weighted factor where the accuracy at the edges of the interval is given 20% while the inclusion of the interval in given 80%. The calculation is as follow:

$$accuracy = 20 \times edges\ accuracy + 80 \times interval\ intersection$$

$$edges\ accuracy = \frac{(g_e - g_s) - abs\,(g_s - t_s) - abs\,(g_e - t_e)}{g_e - g_s}$$

$$interval\ intersection = \frac{g \cap t}{g_e - g_s}$$

where g is the ground truth interval, t is the submitted interval, s is the start time of the interval, e is the end time of the interval, and $g \cap t$ is the intersected duration between g and t.

These counts were tracked to analyze the user's behaviour and to investigate the difference between the two methods.

The participants were also asked to fill out questionnaires (Section B.3) and give their comments on the entire interface, and suggestions for improvement. This was conducted to qualitatively compare the two methods and to provide an indication of the importance of correcting each particular aspect of the proposed interface. Participants rated different interface features and tools using a 7-point Likert scale. The experiment lasted approximately two hours per participant.

Procedure

Participants were given a description of the procedures to be employed in the study, informed of the goals and objectives of the study, and informed consent was requested to participate in the study. Two methods were evaluated in the study: Filmstrip and History Timeline.

The researcher started by showing the interface (shown in Figure 5.4), describing its components, explaining the functionality of the tools within the interface and how to answer a question in order to complete each task. Participants were then asked to try the same interface with the first video (**V1**) and task. They were given the freedom to play with the interface and ask any questions. This familiarity stage was intended to allow the participant to try all the interface components and become familiar with the available functionality. The participants started watching the video provided without knowing the questions they will be asked.

For each video, the task started by playing the segments from the predefined history in the main video player. Participants were asked to watch the playback

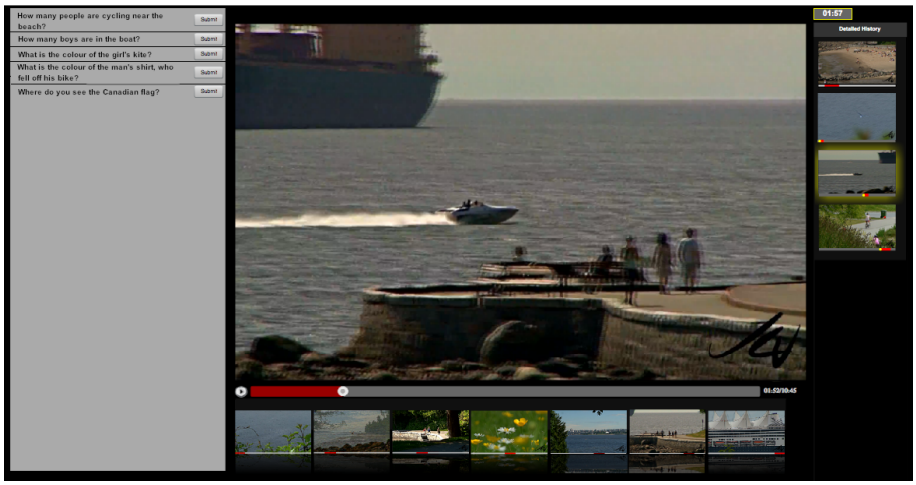


Figure 5.4: The Experiment Interface illustrates the familiarity phase.

and pay attention to the video content in order to answer questions later. Once all segments from the history were viewed, the video paused and a list of questions (compiled by researchers) was displayed to the left of the main video player, as shown in Figure 5.4. Participants were asked to read these questions before clicking on the "Start" button, to be able to ask for any clarification before starting the actual task. Once the "Start" button is clicked, the timer starts, the method's corresponding components were only shown and participants could begin providing answers. Participants were advised to answer as many questions as they could within the given time by watching the clip that contains the answer.

The questions were randomly ordered (i.e. they were not temporally ordered according to their occurrence in the video). Each participant had the freedom to choose the order that they would like to follow to answer these questions (e.g. temporally, linearly, difficulty, memorability, etc.). The participants started answering these questions by clicking on a "Start" button after reading the list of questions. In the familiarity phase, the participant could use either the Filmstrip or History Timeline to navigate the video in an attempt to find an answer. To answer a question, the participant needed to re-watch the clip that contains the answer in the main video player. They navigated Video Segments in the provided component (Filmstrip or History Timeline) to find a clip or a video frame that contains the

answer. Once they found the frame, they dragged the Video Segment of a specific video frame using the seek thumb of the Video Segment to the main video player to start playing that clip. After watching the interval that contains an answer for a single question, they clicked on a submit button beside the corresponding question to proceed to the next question.

Each submitted clip was automatically evaluated: it was considered correct if it contains at least one frame from the ground truth answer. If the answer was correct, the question was faded out from the list and a message was displayed to participants (e.g. "Great. You have 2:14 to answer the rest. Hurry up.") to encourage them and maintain time pressure. However, if the answer was wrong, a message is shown which asks the participant to try again "Sorry. Try again". Participants repeated the same actions until they answered all questions or the time elapsed. A new list of questions was displayed and the participants continued applying the same procedure to answer these questions. The questions for each video were divided into two lists to avoid overloading participants with too many questions. A short break was provided between question sets.

After completing the second list of questions, the familiarity stage ended and the participants were advanced to the experiment where they watched a new video in similar settings to the watching phase of the familiarity stage. Upon the completion of the watching phase, a new list of questions was given based on the content of this video. In this stage, the user was provided with only one method (Filmstrip or History Timeline) to be used in finding the answers. Once all questions were answered or the time given elapsed, the participants proceeded to the next video where they repeated the previous stage but with a different video and using a different method than the one used for the previous video. Participants continued tasks until they completed all four videos.

During the experiment, participants were asked to fill out three questionnaires, shown in Section B.3. These questionnaires were designed using the standard Questionnaire for User Interface Satisfaction (QUIS version 6.0 [27]), modified to reflect the functionality and tools applied and the usability of our interface. Some sections were removed (e.g. Terminology and system information, and System capabilities), some questions were modified, and others were added. The first questionnaire was after **V4** in which participants evaluated the method they used for

that video. The second questionnaire was after **V5** where participants rated the usability of the second method they used. After completing the second questionnaire, participants were given the full interface where they have the freedom to create their own history and explore use cases that might be applicable to them. Every performed seek within the video or any watched interval in the main video player was recorded and visualized as a Video Segment in the History Timeline. This helped the participants to understand how history is created for later usage. Once they were satisfied playing with the interface, they were asked to fill out the last QUIS questionnaire, which evaluates the entire interface.

Results and Discussions

The study showed positive and promising results on the list of thumbnails visualization and its features. It showed a significant performance over the state-of-the-art method and participants preferred it to the Filmstrip.

Method's Performance: T-test analysis results, illustrated in Table 5.2, showed that our method (History Timeline) was significantly faster than the state-of-the-art (Filmstrip), which allowed participants to answer more questions within the same time. Additionally, History Timeline had significantly fewer thumbnail previews compared to Filmstrip. Participants were significantly more precise in finding the answers using History Timeline. However, both methods demonstrated similar behaviour in terms of average number of seeks, and errors.

For each video, some of the measured variables showed a significant difference between the two methods, illustrated in Figure 5.5. There was a significant difference in the percentage of questions answered for History Timeline and Filmstrip in V2 (i.e. Geri's Game); $t(10)= 14.1$, $p < 0.001$, V3 (i.e. Partly Cloudy); $t(10)= 20.41$, $p < 0.001$, and V4 (i.e. Alma); $t(10)= 18.4$, $p < 0.001$. In terms of the average time needed to answer a question, with History Timeline, participants took significantly less time than Filmstrip in V2; $t(1592)= 4.81$, $p < 0.001$, and V4; $t(1592)= 4.06$, $p < 0.001$. V2;$t(1592)= 5.78$, $p < 0.001$, and V4; $t(1592)= 5.65$, $p < 0.001$, showed also significant difference between the two methods in terms of the average number of previews existed per question. With History Time-

Table 5.2: Results of the comparative study between list of thumbnails and Filmstrip for the answer search task, showing a significant advantage using History Timeline in % of answered questions, time needed to answer a question, previews, and interval accuracy. Note: SD = standard deviation; ns = not significant; average time per question measured in seconds. * $p < 0.025$

	Filmstrip	**History**	
	Mean (SD)	Mean (SD)	t-test
Total % of answered questions	64.54	83.21	-
Mean % of answered questions	62.12 (15.20)	73.26 (18.87)	11.23*
Average time per question	21.18 (14.28)	19.39 (12.35)	2.27*
Average no. of seeks	1.55 (1.29)	1.49 (1.14)	ns
Average no. of previews	27.79 (26.67)	24.00 (24.89)	2.49*
Average no. of errors	0.23 (0.63)	0.18 (0.56)	ns
Average accuracy	70.34 (23.70)	76.07 (20.09)	4.40*

line participants were significantly more accurate in selecting the required interval for all videos (V2: $t(1592)= 4.07$, $p < 0.001$, V3: $t(1592)= 2.84$, $p < 0.005$, V4: $t(1592)= 6.35$, $p < 0.001$, and V5: $t(1592)= 4.75$, $p < 0.001$). And lastly in terms of the average number of seeks, a significant difference between the methods only occurred in V4 ($t(1592)= 5.11$, $p < 0.001$) where History Timeline had fewer of seeks. Participants tended to make more errors when using Filmstrip.

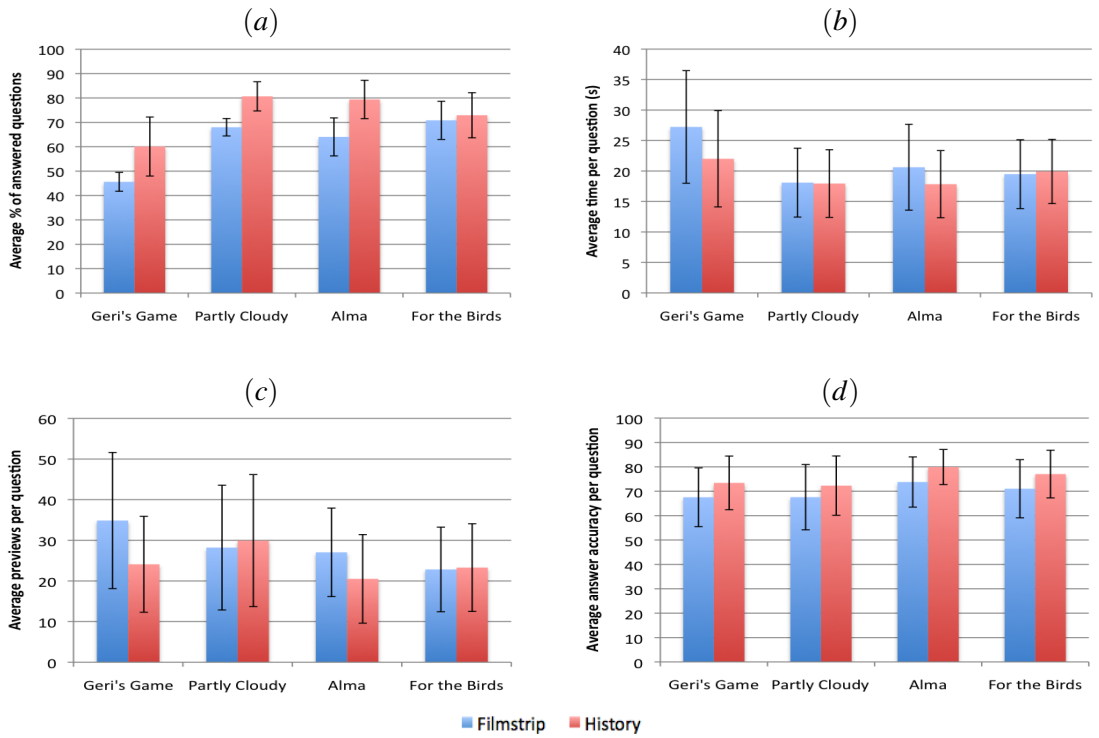


Figure 5.5: The performance of List of thumbnails and Filmstrip in terms of: (a) Average percentage of answered questions, (b) average time per question, (c) average number of previews, (d) average answer accuracy per question, for each tested video. List of thumbnails had significantly more questions answered, less time to answer each question, less number of previews, and more accurate answers.

Method's Features Ratings: Each participant answered two identical questionnaires (using questions 8-28 from Section B.3), one for each method, to compare the overall reaction, the time it takes to learn and general impressions. Each questionnaire contained thirty-seven 7-point Likert scale questions. The data collected for the two methods were analyzed and compared in terms of these 37 questions. A Wilcoxon Signed Rank Test revealed that the overall rating of the Filmstrip was significantly lower than the History Timeline, $z = -1.73$, $p < .05$. The History Timeline showed no significant difference from the Filmstrip in all questions except for the question "Learning to use the method features" where (1:Difficult, 7:Easy) History Timeline ($Md = 6$), and Filmstrip ($Md = 5$), $z = -1.82$, $p < 0.04$. This also coincided with participants' preference where nine out of twelve participants stated that the History Timeline was faster and easier in finding the answers for the questions. Participants reported that the History Timeline was faster because it is based on a personal mental context map created for the video. It was easier for them to refer back to the corresponding segments when needed, which was not the case for Filmstrip. Since Filmstrip segments were created systematically, participants needed at least to navigate one segment. If the answer was within the segment, they submitted it, otherwise they needed to navigate the preceding or subsequent segment. This was also demonstrated by the participants' quantitative results where they did not need to preview thumbnails so often and were able to answer more questions using History Timeline. All questions for both methods were rated above 5 out of 7 except for one question for the Filmstrip: "Using the method is effortless" ($Md = 4.5$).

Interface's features Ratings: After acquiring some familiarity with the overall interface in phase 2, participants were asked to complete a third questionnaire (Section B.3) to rate the entire interface. Responses were compiled for each of the 12 participants in the study, along with any written comments that the participants had. The overall mean rating of all sections of the QUIS was 5.73, on a 7-point scale, and all questions were rated above 5.

For the "Overall Reaction" section, the "Ease of use" overall rating was the only factor that was rated significantly different from the mean response (M = 6, SD = 0.48) where it was higher, indicating that users can easily learn and use

the interface. The others seven ratings, Impressiveness (M = 5.58, SD = 0.54), Satisfaction (M = 5.17, SD = 0.51), Stimulation (M = 5.92, SD = 0.48), Perceived "powerful" (M = 5.83, SD = 0.56), Flexibility (M = 5.5, SD = 0.69), Helpfulness (M = 5.92, SD = 0.58), and Usability (M = 6.08, SD = 0.50) were not significantly different from the mean user response level. From these "Overall Reactions" to the interface, we can conclude that users found our interface easy to use, helpful, useful, and flexible. In the "Learning" section, the overall rating of all items were not significantly different from the mean response except for one item "Steps to complete a task follow a logical sequence" (1:Never, 7:Always) (M = 5.83, SD = 0.42), which was rated significantly more than the mean response. However, for the general impressions section, the overall ratings for "Screens are aesthetically pleasing" (M = 6, SD = 0.43), "Screen designs and layout are attractive" (M = 5.92, SD = 0.33), "Interface is impressive" (M = 5.75, SD = 0.61), "Interface can do a great deal" (M = 5.83, SD = 0.42), and "Interface is fun to use" (M = 5.75, SD = 0.57), were rated significantly better than the mean response. This indicates that users found our interface aesthetically pleasing and fun to use. Nevertheless, "Interface maintains one's interest" was rated significantly lower than the mean response (M = 5.67, SD = 0.44), which might be due to the limited videos used in the experiment. The remaining items of this section were not significantly different from the mean response. For the "Satisfaction" section, all items were also not significantly different from the mean response.

Interface's Negative & Positive Aspects: In addition to the QUIS item findings, participants were also asked to list the three most negative and positive aspects of the interface. Some participants reported that "I found it weird to have a vertical list of video pieces", "Vertical scrolling", "Not being able to delete segments from history", and "Not being able to favorite some segments from history." The comment about the vertical scrolling coincided with their responses to the location of the component where the mean rating for the location of History Timeline ($M = 5.1, SD = 0.79$) was significantly lower than the location of the Filmstrip ($M = 6, SD = 0.67$), $t(22) = 2.49$, $p = 0.021$. However, some participants found having two different orientations for the different components (i.e. horizontal Filmstrip and vertical History Timeline) made it easier for them to differentiate. We de-

signed the component in this layout to eliminate the confusion between the History Timeline and the video timeline; this may need additional investigation. About favouriting and deleting items from history, we are considering adding these features to the interface to help users manage their history and being able to filter it using the favourite items. Participants also commented, "The main framework could be better if the vertical and horizontal list can be chosen to be disappeared". We think having the History Timeline in a different mode (similar to web browsers and YouTube history) might help since it will reduce the number of thumbnails presented at the same time in the interface, as well as prevent the dynamic change in the History Timeline as the user interacts with a video.

Participants found the interface helpful and impressive in being able to dynamically create points or bookmarks, which would allow them to skip to a favourite clip. This was also seen in participants' response to how segments were created, where "Having a control over the creation of the segments" in the History Timeline (M = 5.92, SD = 0.79) was rated significantly higher than "Systematically created intervals" in Filmstrip (M = 5.01, SD = 0.79). Participants appreciated having control of what they are watching, being able to go back to seen intervals, and the ability to create video segments. One participant said, "If I create my own bookmarks in the video, I may skip to a favourite song e.g. in a concert." Most participants stated that the interface is fun to use once you get familiar with it.

Participants' Vision of the Interface: In order to explore whether users foresee potential for the interface we asked them "Where and how do they think video navigation history can be useful?" Participants provided us with valuable responses that gave us some insight as to how this interface can be further modified and tested. They foresee that it would be useful in the educational environment as well for home usage. Some participants indicated its usefulness for sharing where it can be used to pinpoint movies and clips for sharing with friends. Other participants pointed out its application for monitoring video consumption at home and at the office; "Home: parents can either set video clips that are allowed for their children or to monitor what they have watched. Office: employer can monitor if employees are watching videos during office hours." Some indicated that it would be useful for chaptering long videos and emphasize important parts. Others mentioned its

application for educational lectures and tutorials, for example, "pinpoint important clips within educational videos" and "create a summary of an educational lecture." Some participants foresee using it to create tutorials/demos for a technical or an educational presentation or creating song lists.

Participant's Comments: There were also some very positive comments made by the participants about the interface. One participant commented, "I definitely see how this would save my time because I will never need to watch the entire video again to find good stuff." Others said, "Pretty impressive framework and dynamic response"; "The interface is impressive and it has potential usage in comparison to other video player interfaces"; and finally, "I would love to see this implemented within social websites. I could see how I would use it and definitely I will have more fun."

Based on the participants' valuable comments and suggestions, we believe using personal navigation history is helpful in navigating a video space. This could work for different applications such as highlights or summary of videos, a movie using home videos, sharing interesting clips, quickly navigating and skimming previously watched videos, and finally watching new videos' interesting parts using crowd-sourced viewing histories. Getting this positive feedback from the participants and how they welcomed the idea of using their personal navigation history motivated us to investigate other ways to visualize viewing history and how to visualize multiple-video history. In the next section, we are going to present another approach to visualize this viewing history and explore the usefulness of crowd-sourced metadata.

5.2 Visualize History Using Consumption Frequencies

Online video viewing has seen explosive growth, yet simple tools to facilitate navigation and sharing of the large video space have not kept pace. Our objective is to design a visualization that supports fast in-video navigation (play most popular or unseen parts), search (seek within intervals with prior knowledge e.g. 'seen the event before' or 'never seen the event'), preview, and instant sharing (share a single interval directly). To accomplish this, we propose the use of single-video view-

ing statistics generated from an individual's video consumption, or crowd-sourced from many people watching the same video; as the basis for visualizing video's footprints. Whenever a segment of video is played, the video ID and timestamps for the interval's start and end are recorded. An accumulated view count is maintained for the video at a given resolution (e.g. 15 samples per second of video). This data can then be used to visualize how a video was consumed, create a summary of a video viewing history, and to easily navigate through its clips.

There are two different approaches to how this data can be visualized: (1) using colour or heatmap, or (2) using representative variable-sized thumbnails.

5.2.1 Footprints Visualization Using Colour Intensity

The viewing history, or footprints, offers users a visualization of the view count of every point of time in the video or how often each part was viewed. This is similar in concept to timeline footprints [82], as shown in Figure 2.2. The blue bar becomes brighter for the most-viewed intervals, and each one can be selected to play. It provides the user with information on the most-viewed intervals of that video (a temporally tracked frequency). Using the user's viewing statistics, a view count is maintained for all points of time for each viewed video. As the user views a video, the view count will dynamically change, thus the count is continuously updated and the information propagated backwards through the history. The view counts are then summed and normalized within each video. This visualization can be also used to visualize the viewing statistics from multiple users (i.e. crowd-sourced), which can be combined to form the most popular intervals for a video. This provides users who are watching a video for the first time with a mechanism to watch what others have found to be most interesting. This is a fast navigation technique, predicated on trusting the crowd to be "correct" on interesting video content.

Limitations of a Coloured Timeline

Applying colour intensity to visualize users' footprints in a video faces many problems such as:

1. There is no in-place preview, which makes it impossible to determine the

content from the footprints. For example, users cannot tell what the content is of the most viewed part.

2. It requires users to navigate to any part to reveal its content.

3. There is no direct interval sharing.

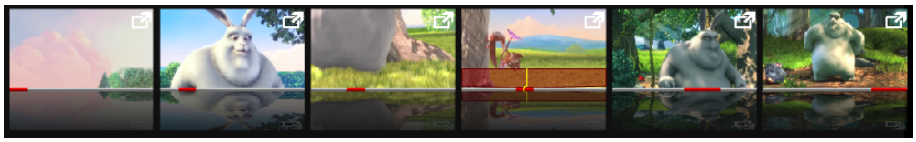
4. Having multiple variations of the colour to indicate the intensity makes it hard for users to distinguish these frequencies.

Figure 5.6: A Coloured Timeline visualizes a user's navigation footprints over a video Timeline using colour intensity. The more an interval within a video is watched the brighter that region becomes.

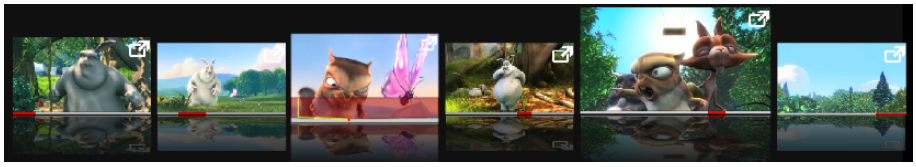
5.2.2 Footprints Visualization Using Variable-sized Thumbnails

This approach also uses viewing statistics (personal or crowd-sourced) as the basis for modification to the well-known Filmstrip visualization [31], to create our proposed visualization, which we call *View Count Record* (VCR). The *VCR* employs a Filmstrip metaphor since thumbnails are accepted as a standard on which nearly all digital video retrieval interfaces are built upon. Moreover, the Filmstrip metaphor is commonly used to present content of video as a navigation device. It is a presentation scheme for abstracting information in a digital video segment, which is considered effective on desktop systems for video retrieval [31]. Furthermore, the Filmstrip communicates shot information all at once in a static form, which is a quick method for showing video content as opposed to playing it.

The *VCR*, shown in Figure 5.7, uses a variable thumbnail size (and variable interval length) to reflect the relative popularity of intervals. We used size instead of colour since we are representing intervals using thumbnails where colour discrimination may be confused with the thumbnail content and would be difficult to differentiate for some videos. We also piloted the visualization using a coloured frequency bar attached to each thumbnail to indicate the popularity (Chapter 4); while the information was welcomed, the visualization was reported as cluttered.



(a) Filmstrip (equivalent to VCR without visualization of viewing statistics)



(b) View Count Record (VCR)

Figure 5.7: The View Count Record (VCR) component visualizes the video viewing statistics. When no viewing history is available, the VCR presents a familiar Filmstrip (a). When a history is available, our view count manipulation algorithms can be applied to visualize popular intervals within the video, leading to fast personal navigation and social navigation applications. Each thumbnail can be searched via seeking in the popup red time bar when hovering the cursor over the preview image.

The *VCR* applies a histogram visualization similar to [69] but it uses thumbnails instead of a time-series graph, where the height of these thumbnails indicates the height of the histogram bars. It consists of a fixed number of size-able *video segments*[7] (described in Section 5.1.1). The duration and size of each segment is based on how often its corresponding interval has been viewed. If no viewing statistics are available, the *VCR* appears as a normal Filmstrip as shown in Figure 5.7(a). As the user consumes video, the VCR updates its segments accordingly.

VCR Construction

The construction of the *VCR*, illustrated in Algorithm 1, starts by gathering intervals of time in which consecutive frames have equal view counts. While there are intervals less than a set threshold, the algorithm attempts to merge these intervals with one of their neighbouring intervals. The neighbour to merge with is deter-

[7] In our interface, we used 6 segments based on the width of the interface and the maximum width of a single *VCR* segment.

mined by two criteria: first by the difference in view counts, and if the difference of view counts are equal, then by the duration of the neighbouring intervals. The merging process chooses the smallest difference in view counts, or the smallest interval duration. This process repeats until all intervals' duration are greater than the preset threshold.

Algorithm 1 VCR construction: every VCR segment duration and view count (VC) is measured based on other existing segments.

```
Retrieve segments S = {S_1, S_2,..., S_n} using Algorithm 2
if n > no. of thumbnails in the VCR then
  repeat
    Retrieve peak segments P = {P_1, P_2,..., P_k}
    if k > no. of thumbnails in the VCR ÷2 then
      p ← P(index(min(P))).sIndex
    else
      p ← index(min(S.VC))
    end if
    if S_{p-1}.VC = S_{p+1}.VC then
      m ← index(min(S_{p-1}.dur, S_{p+1}.dur)
    else
      m ← index(min(abs(S_p.VC − S_{p-1}.VC), abs(S_p.VC − S_{p+1}.VC)))
    end if
    S_m.VC ← (S_m.VC × S_m.dur + S_p.VC × S_p.dur) ÷ (S_m.dur + S_p.dur)
    S_m.dur ← S_m.dur + S_p.dur
    remove S_p
  until S.length ≤ no. of thumbnails in the VCR
else
  while S.length < no. of thumbnails in the VCR do
    p ← index(max(S.dur))
    insert new segment s at p + 1
    S_{p+1}.VC ← S_p.VC
    S_p.dur ← S_p.dur ÷ 2; S_{p+1}.dur ← S_p.dur
  end while
end if
draw {S_1, S_2,..., S_{S.length}}
```

Upon the completion of the merging process, the *VCR* contains a set of intervals with durations that are greater than the preset threshold. However, since the

Algorithm 2 Segments Merging: every segment with a duration less than a predefined threshold is merged with its closest neighboring segment.

```
Retrieve segments S = {S_1, S_2,..., S_n} using Algorithm 3
for each s ∈ S do
  if S_s.dur < threshold then
    if S_{s-1}.VC = S_{s+1}.VC then
      m ← index(min(S_{s-1}.dur), S_{s+1}.dur)
    else
      m ← index(min(abs(S_s.VC − S_{s-1}.VC), abs(S_s.VC − S_{s+1}.VC)))
    end if
    S_m.VC ← (S_m.VC × S_m.dur + S_s.VC × S_s.dur) ÷ (S_m.dur + S_s.dur)
    S_m.dur ← S_m.dur + S_s.dur
    remove S_s
  end if
end for
```

Algorithm 3 Frames Gathering: every consecutive frames with equal view count are gathered into one segment.

```
Retrieve frequencies of video frames from user's heuristics F = {F_1, F_2,..., F_n}

S = {}
index ← 1; S_1.VC ← F_1; S_1.dur ← 0
for each f ∈ F do
  if F_f ≠ S_index.VC then
    index ← index + 1
    S_index.VC ← F_f; S_index.dur ← 1;
  else
    S_index.dur ← S_index.dur + 1
  end if
end for
```

number of items in the visualization component is limited, we must reduce the set of intervals to match. Thus, we look at the peaks of the view count graph, keep the highest peaks, and merge the other intervals until we get to the desired resolution. Conversely, if we do not have enough intervals, we linearly sample and split intervals until we have enough.

We then create our visualization by using a size-able Video Segment compo-

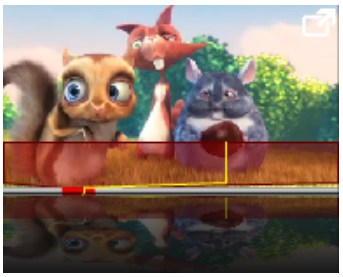


Figure 5.8: Each video thumbnail in the VCR is visualized as small video segments. Each segment is seek-able and play-able on mouse events. The red/gray portion at the bottom of the widget indicates the temporal location of its interval within the complete video. The yellow line illustrates the current seeking point within the thumbnail, within the zoomed interval for higher-resolution seeking.

nent (described in Section 5.1.1), shown in Figure 5.8, for each interval. The size of each segment is based on a ratio of the current segment's view count to the maximum view count for the video. The *VCR* updates automatically when the video is paused, based on the latest viewing statistics (it does not update while viewing so as not to distract from the main video). It illustrates to the user how they consumed the video and which parts were viewed the most/least. This provides a simple mechanism to find or navigate back to these segments when needed.

VCR Scalability

The construction and visualization of the *VCR* is based on the video used, the number of peaks, and the interface size, which are independent of the platform used. It is not affected by the length or duration of the video being visualized since the algorithm as described in Section 5.2.2, merges (or linearly samples and splits intervals) until the *VCR* gets to the desired resolution (i.e the required number of segments). However, due to the limited space and the fixed number of video segments, some medium-height peaks may be diminished and not easily viewed in the *VCR*. To alleviate this problem, the interface supports a zoom feature (via mouse wheel) where the selected video segment expands and is represented by its own *VCR* with the same number of segments. When a segment is zoomed-in, the

VCR updates to visualize segments within that zoomed segment only and hides any other segments. Thus the *VCR* always uses the same number of video segments.

Personal and Crowd-Sourced VCR

The *VCR* can be used to present either personal viewing statistics or crowd-sourced metadata. It uses the same algorithm described in Section 5.2.2. In a crowd-sourced VCR, the viewing count is gathered from multiple users and then used in the algorithm to construct the *VCR*. Having both options will allow users to compare their viewing behaviour with the crowd and for unseen videos the crowd-sourced VCR will give them suggestions on what to watch from the questioned video.

Advantages over Coloured Timeline

Applying size-able thumbnails to visualize how users viewed a certain video have resolved the issues presented by a coloured Timeline. It allows users to easily indicate the content of most popular parts of the video. Overlapping peaks can be easily distinguished by just scrolling the mouse wheel over the intended segment, which reduces the need of navigating or seeking to each peak. Thus, having distinctive segments allows fast sharing of these segments. Similar to the coloured timeline, *VCR* provides quick navigation to different parts of the video, but, *VCR* also provides fast search of content without the burden of continuous seek or clicks on different peaks.

5.2.3 Investigating View Count Record (VCR) Visualization

We designed a comparative study to evaluate our navigation tool and see how the history would help people to access their previously seen clips to share one of their liked segments. Our aim was to investigate 1) if our visualization of video navigation provides faster search for user-specific affective intervals, and if users prefer our visualization for this task; 2) if crowd-sourced histories provide good summaries of video. Participants were asked to find and share their favourite intervals using either the *VCR* or *Filmstrip* visualizations. In this study we went with a personalized approach for the search task to get a better insight on how this would

work for real scenarios. We compared against the Filmstrip design instead of the footprints' coloured timeline [82] for several reasons: footprints does not easily let a user directly select or share a full interval; video cannot be previewed inside the footprints visualization (*VCR* and *Filmstrip* can both directly preview without seeking); *VCR* and footprints could be used together, so we believe a comparison against *Filmstrip* is more informative.

In this study, we hypothesized that users would perform better when using *VCR* in terms of time needed to find their previously seen liked segments, and less navigation would be required. In terms of the agreement with the crowd-sourced history, we anticipated that users would appreciate having *VCR* as a recommendation tool for unseen videos. However, in terms of participant agreement with each single segment from the crowd-sourced history, we expected a big variation between participants and videos. For method preference, we expected that participants would like having a *VCR* with history available over when no-history is visualized.

Apparatus

The experiment application was developed in Flash CS4 and ActionScript 3.0. The experiment ran on an Intel dual-processor dual-core 3 GHz Mac Pro desktop with 8GB RAM and equipped with a 24" Dell LCD monitor with a resolution of 1920×1200 pixels at a refresh rate of 60Hz. A Microsoft optical wheel mouse was used as the input pointing device with default settings and the Adobe Air environment was set to 1920×900 pixels while running the Flash program.

Participants

Ten paid volunteers, 6 female and 4 male, participated in the experiment. Participants ranged in ages from 19 to 40. Three of the participants were undergraduate students while the rest (i.e. seven participants) were from the general public (non-academic). Each participant worked on the task individually. All participants were experienced computer users and have normal or corrected to normal vision. Seven participants watch online videos on a daily basis, two watch videos 3-5 times a week, and one watches once a week. Five of the participants watch 1-3 videos on

average per day, while three watch 4-6 videos per day and two watch more than 10 videos per day on average. Table 5.3 presents participants' details.

Design

Two different navigation modes were tested: no-history (*Filmstrip*) and with-history (*VCR*). The case where no-history was available was represented by the state-of-the-art *Filmstrip*, shown in Figure 5.7(a). Each participant tried both modes to navigate and share their preferred parts of the video. Participants were divided equally into two groups where Group 1 used *Filmstrip* first and *VCR* second, and Group 2 had the order reversed. Participants freely watched a set of 5 different videos (Disney short animations) between 3 and 5 minutes long. These videos were chosen based on the same criteria listed in previous experiment, described in Section 5.1.3. The selected videos were: V1: One Man Band[8], V2: Partly Cloudy[9], V3: Day & Night [10], V4: For The Birds[11], and V5: Presto[12]. Video length does not affect the *VCR* as mentioned in Section 5.2.2; however, due to the time constraints of the experiment, short videos were tested.

Each participant performed a total of 14 search tasks (2 modes $\times$ 7 segments per mode); they were asked to perform as quickly as possible. For each task, the completion time, the number of previews, and the number of zoom events were recorded. The completion time was measured from when the participant clicked on a 'Find' button until the moment they found the correct interval (confirmed by the researcher). The navigation behaviour and statistics were recorded during the viewing phase. The participants were also asked to rank the best mode based on speed, ease and preference.

Upon the completion of the sharing tasks, participants started the second task where they were shown a short version of each video, automatically created from crowd-sourced histories (described in Section 5.2.3). Participants were asked whether the shortened version was a good summarization and whether each segment in the shortened video matched their own affective segments. The final task was to fill out

[8]http://www.youtube.com/watch?v=454nNoD6-TI
[9]http://www.youtube.com/watch?v=HAbiJPGHeV0
[10]http://www.youtube.com/watch?v=bi7ybKxymao
[11]http://www.youtube.com/watch?v=yJzQiemCIuY
[12]http://www.youtube.com/watch?v=a0DqCxYfeT0

Table 5.3: Demographics summary for participants in the investigation of the *VCR* visualization study. (Note: WMP = Windows Media Player, QT = QuickTime, VLC = VLC Media Player, RP = Real Player, KMP = KMP Player, Gom = Gom player, M = Mplayer, YT= YouTube

P	Gender	Age Group	Vision	Watching Frequency	Videos per Day	Familiar Players	Frequent Player
1	Male	31-40	Normal	Daily	1-3 videos	iTunes, QT, RP, VLC, WMP	QT
2	Male	26-30	Corrected	Daily	more than 10 videos	iTunes, M, QT, RP, VLC, WMP	YT, KMP
3	Male	19-25	Normal	Daily	4-6 videos	iTunes, M, QT, RP, VLC, WMP	iTunes
4	Female	19-25	Corrected	Daily	4-6 videos	QT, VLC, WMP	WMP
5	Female	19-25	Corrected	3-5 times a week	1-3 videos	RP, VLC, WMP	VLC, WMP
6	Female	19-25	Normal	3-5 times a week	1-3 videos	iTunes, VLC, WMP	VLC
7	Female	26-30	Normal	Once a week	1-3 videos	RP, VLC	RP
8	Female	19-25	Normal	Daily	1-3 videos	iTunes, RP, VLC, WMP	VLC
9	Male	31-40	Normal	Daily	more than 10 videos	RP	YT, RP
10	Female	19-25	Normal	Daily	4-6 videos	VLC, WMP, Gom	WMP

a questionnaire to rank the modes, and provide feedback on the interface (attached in Section B.4). The experiment lasted approximately one hour per participant.

Procedure

The experiment proceeded as follows:

1. The researchers gave the participants a walkthrough of the interface explaining the functionality of each feature and tool, and their effects when applied to a video. This stage took approximately five minutes. The participants were allowed to try the interface and ask any questions during this stage.

2. Participants were then asked to completely watch the five different videos and re-watch any parts they wanted. Only one video was played at a time. Participants' navigation behaviour was stored to be used for the searching task. To ensure all participants had seen an equivalent video content, the tasks began after all videos were viewed.

3. Once each video was viewed, participants were asked to list five different intervals they would like to share; these were recorded by the researcher. Once the participants named the segments, they were advanced to the next video.

4. When all videos were completely viewed, the researcher chose an event from those provided by the participant that they had to find, one event at a time. Events were chosen from different videos so that no consecutive search tasks came from the same video. The search task began by clicking 'Find', choosing a video from a grid of thumbnails, and then the navigation layout for the current mode was displayed - the participant used this to find an interval representing the event. The interval is submitted for consideration by playing it in the main player: if approved by the researcher as correct, the task is complete. Participants proceeded to find and share the next event. Upon the completion of finding the seven segments, the participants were advanced to the next mode to find another seven segments.

5. The first stage of the experiment ended once the participant had experimented in two modes where they found seven segments for each mode.

After which, a short version of mostly viewed segments of each previously watched video based on the crowd-sourced statistics was played. The crowd-sourced heuristics was an aggregation of a navigation history of 6 people, described in Section 5.2.3, who are different from the experiment's participants.

6. Once a shortened video stopped playing, participants were asked if they thought the shortened version was a good summarization and whether each segment in the shortened version matched their own affective segments.

7. The experiment ended when participants had ranked the five different shortened videos. Finally, the participants were asked to fill out a questionnaire where they ranked the modes and gave their feedback and comments about the interface, its features and their experience.

Crowd-Sourced Data Collection

Six graduate students (2 female, 4 male, aged 24 to 37), completely separate from participants in this study, voluntarily participated in the crowd-sourced data collection. Participants were invited prior to the experiment to freely watch and navigate the same set of videos while their viewing statistics were recorded. Their data was then aggregated and visualized using the *VCR*. At least nine peaks existed for each video. However, due to the experiment time constraints (one hour), we decided to use only the highest four peaks of each video in the shortened videos that were tested. Figure 5.9 illustrates the crowd-sourced data for "One Man Band" video, highlighting the segments chosen for the shortened video, and the *VCR* of this data is shown on top of the graph.

Results and Discussions

Most participants commented that they enjoyed their time using the interface and they foresaw its applicability as a navigation aid for unwatched videos where social navigation can be leveraged for the benefit of future viewers, as well as a summarization tool for their own videos. Participants were impressed by how closely the crowd-sourced popular intervals matched their own preferences for favourite

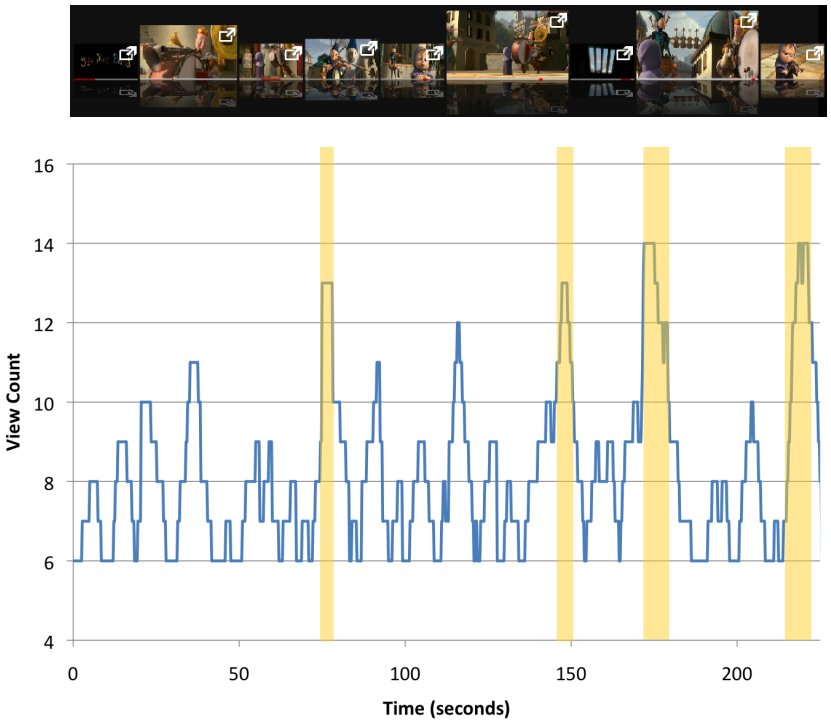


Figure 5.9: The crowd-sourced data of the "One Man Band" video. The highest four peaks used for the shortened video are highlighted in yellow. The View Count Record (VCR) visualizing this crowd-sourced viewing statistics is illustrated on top of the graph.

intervals, confirming that in most cases this would provide an effective tool for navigating new video.

Search Task: The main task in the experiment was to search for previously seen preferred intervals: each participant was able to complete each search task in less than one minute (for all 14 trials). A paired-samples t-test analysis determined the significance of the results in terms of the average completion time per search and the average number of previews per search. The analysis, shown in Table 5.4, demonstrated that the search task using *Filmstrip* took significantly more time than with the *VCR*. Participants were asked to rank the different modes for preference, ease and speed: they ranked the *VCR* as the most liked, easiest and fastest mode,

Table 5.4: Results of the comparative study for the interval retrieval task, showing a significant advantage using our method (VCR) in terms of completion time. Note: SD = standard deviation; completion time measured in seconds. * $p < 0.03$

	Filmstrip		**VCR**		
	Mean	SD	Mean	SD	t-test
Completion Time	24.31	10.42	21.11	5.38	2.28*
No. of Previews	40.39	35.41	35.53	27.56	0.88

which coincides with the quantitative results. This indicates that having access to the user's personal navigation record is useful for finding previously-seen content within video, and that our visualization cues (e.g. size) of the mostly watched segments helped users to quickly and easily navigate to the correct intervals.

In terms of the average number of previews, the results revealed no significant difference between the two modes, which we did not anticipate. This may be due to the fact that many view count peaks can exist within a single video segment of the *VCR*, and that some segments ended up much smaller in size, which made it harder to navigate. When analyzing the participants' navigation history, we found that participants created 11 history segments on average per video. This means that when using heuristics some *VCR* segments had more than one peak since there are only 6 segments in the *VCR*. However, as we mentioned in Section 5.2.2, we added the zoom functionality to mitigate this. Participants rarely used the zoom feature and preferred to navigate through these segments instead, which explains the large number of previews.

Agreement With Crowd-Sourced VCR: All participants agreed that the shortened video (created automatically using the crowd-sourced data) was an effective summary of the video content. Before using the interface, participants were asked whether they would use others' recommendations as a tool for navigating unseen videos; we were most interested in discovering if participants' views would change after using our interface. Most (9 out of 10) said they would not use recommendations; however after using the interface and viewing the shortened video they

expressed surprise at the quality of the summary. Participants mentioned that having the crowd-sourced *VCR* would save time, especially for long videos, since they can decide whether to watch the entire video or just the summary, or even just parts of the summary.

For each video, participants were asked to rank each segment derived from the crowd-sourced data. At least eight out of ten participants agreed that each segment represented something they liked or illustrated an affective clip. Out of a total of 20 segments, 7 segments were liked by eight participants, 8 segments were liked by nine subjects, while the remaining 5 were liked by all participants. Some of the segments that were not liked by some participants were either due to religious beliefs or perceived as violent content, while other participants considered these segments to be funny. Looking at participants' viewing heuristics for each video, as shown in Figure 5.10, also revealed high agreement with the crowd-sourced segments, which matches and justifies participants' rankings. We expected the variation between participants; however, we did not predict the generally high level of agreement. This suggests implicit tagging of video from many users may serve as a valuable navigation tool for online video.

Participants' Viewing Heuristics: Participants' viewing heuristics were gathered for each video to be compared with the crowd-sourced metadata and explore the feasibility of using viewing heuristics as a recommendation tool. The new cumulative view count for each video showed a similar trend to the crowd-sourced metadata but with less variations. Figure 5.11 illustrates the new collective heuristics for "One Man Band" along with the crowd-sourced view count of the six volunteers used in this experiment. They have similar trends with a cleaner graph for the new collected view counts. Having more users viewing the same video will remove most of the noise in a graph leading to a more representative summary of the video. New viewers can then use these summaries to judge whether or not to watch the questioned video.

Looking at each participant's viewing heuristic compared to the crowd, it showed similar behaviour between the video portions that participants watched more than once. Individual peaks almost aligned with the highest peaks of the crowd-sourced metadata. For example, for the video "Partly Cloudy", all participants had more

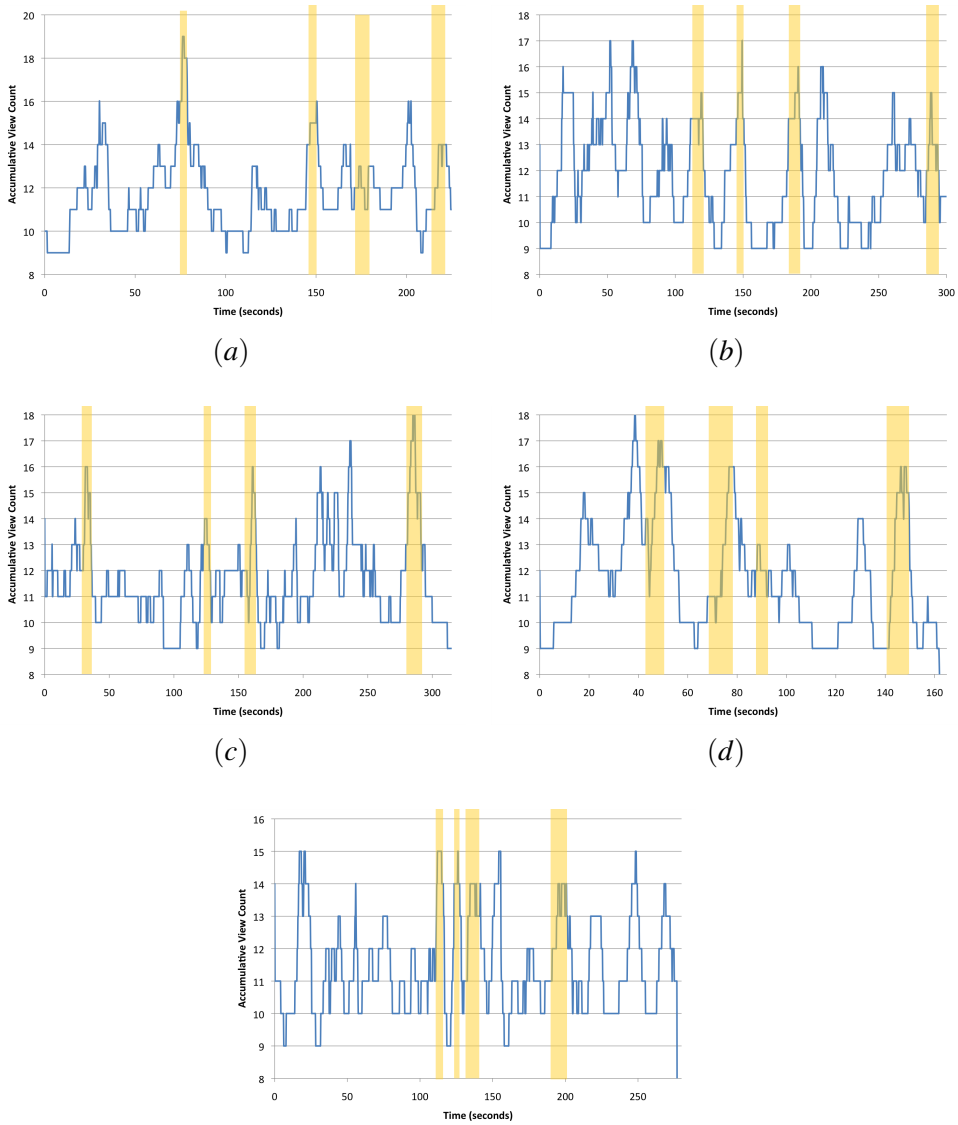


Figure 5.10: The accumulative view count for each video: (a) One Man Band, (b) Partly Cloudy, (c) Day & Night, (d) For The Birds, and (e) Presto. The segments used for each shortened video are highlighted in yellow. There is a high agreement between view count peaks and selected crowd-sourced segments.

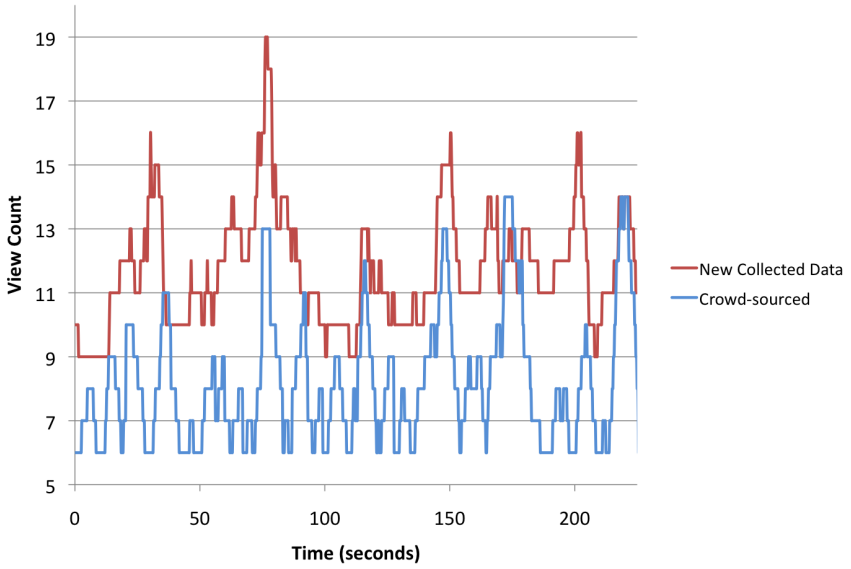


Figure 5.11: The crowd-sourced data of the "One Man Band" video along with the new cumulative view count. A similar trend appears in both graphs but the new collected data shows cleaner data with distinctive peaks.

than one peak aligned with crowd-sourced highest peaks with the exception of participants 4 and 6 who had only one (Figure 5.12). This high alignment justifies the high agreement for the clips used in each shortened video. Moreover, comparing each individual viewing behaviour with the collective behaviour from all participants showed at least five matched re-watched segments per video for all participants as shown in Figure 5.13. This coincides with what we saw in the crowd-sourced data, which proves that crowd-sourced data can provide a potential video summaization tool.

We also looked at the list of the events participants named for each video that they would possibly share with others. The total number of different named events per video is illustrated in Table 5.5 with the number of participants per event in each video. There were at least four (out of 5) events that were listed by more than five participants, which indicates around 80% agreement per video on events to be shared. Looking at individual listings showed that almost all of the events the

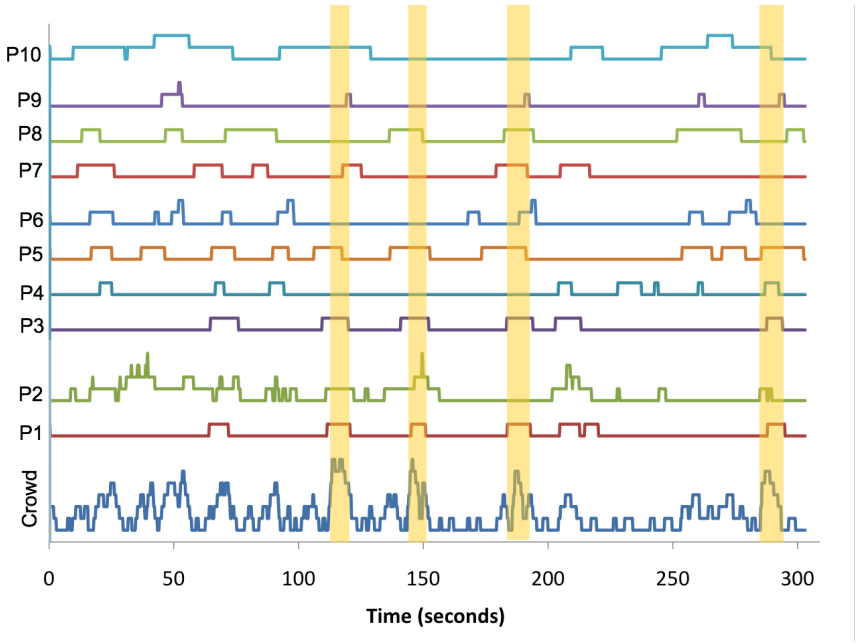


Figure 5.12: Difference between individual behaviours and crowd-sourced data for "Partly Cloudy". All participants had more than one peak aligned with crowd-sourced highest peaks with the exception of participants 4 and 6 who had only one. Highest peaks are highlighted in yellow.

participants listed match the peaks in their personal viewing history. This proves that the re-watching behaviour coincides with affective parts in a video (similar to [12]), which can offer a simple tool for clips recommendations within a video for sharing.

Features' Ranking: From the aggregated results of the questionnaire (measuring ease-of-use and usefulness), the average ranking across all components and features was 5.82 out of 7. All features were ranked above 5 except for three items: getting started (($M = 4.5$), remembering how to use the interface ($M = 4.6$), and using the zoom ($M = 4.3$). The zoom scored slightly lower due to the mouse wheel sensitivity being reported as too high, which led to some participants becoming confused or frustrated. This could also explain the low usage of this feature while

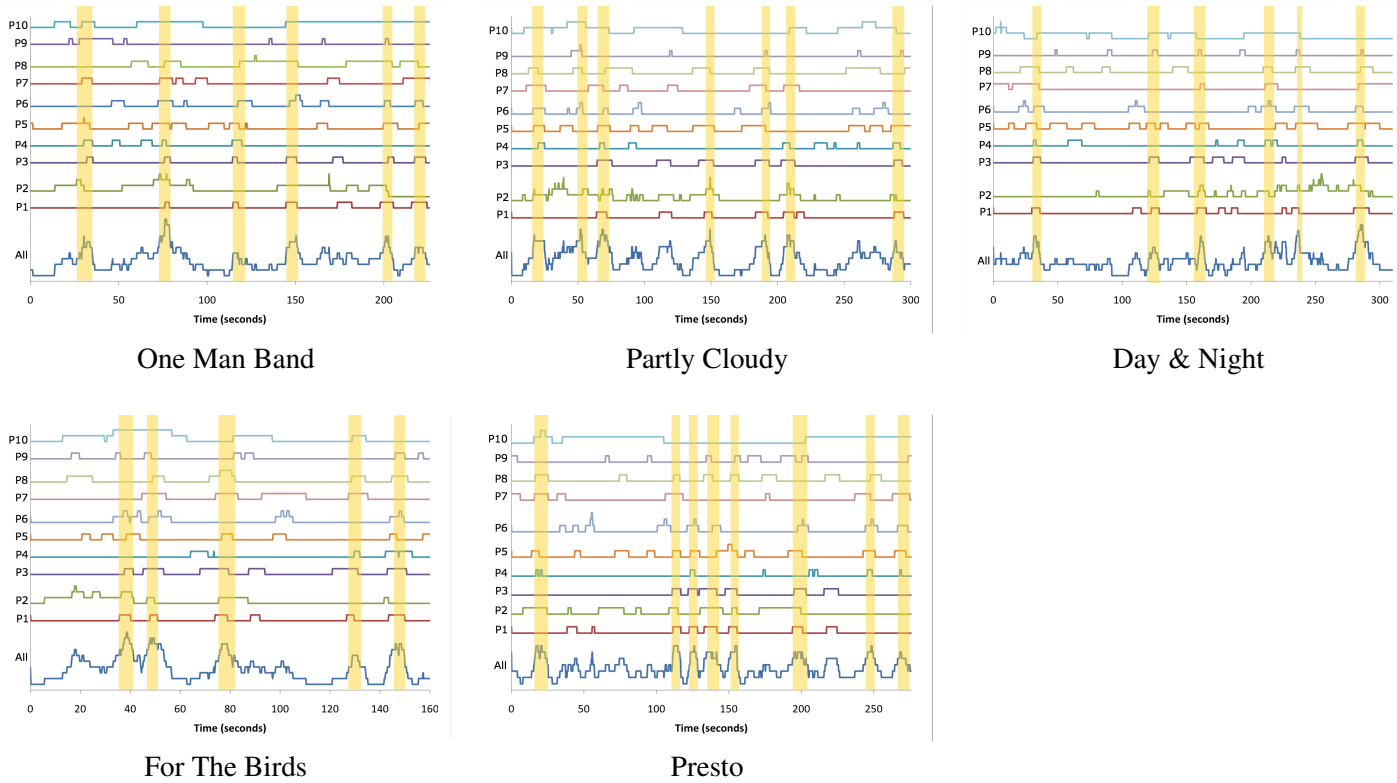


Figure 5.13: Participants' viewing heuristics for each tested video. Intervals that are re-watched by more than five participants are highlighted in yellow. A high agreement between participants re-watched segments can be seen for each video where there are at least five matched segments.

Table 5.5: Agreement between events participants listed for each video. There are at least 4 (out of 5) events that were listed by at least 50% of the participants. Note: **V2:** One Man Band, **V3:** Partly Cloudy, **V4:** Day & Night, **V5:** For The Birds, and **V6:** Presto

	# of Events	# of Participants per Event																		
V2	19	8	7	6	5	5	3	3	3	2	2	1	1	1	1	1	1	1	1	1
V3	15	6	6	6	5	4	4	4	3	3	3	3	3	2	1	1				
V4	19	7	7	6	5	5	3	3	3	3	2	2	2	2	1	1	1	1	1	1
V5	12	8	7	6	5	5	4	4	4	3	2	2	2							
V6	19	8	7	5	5	4	4	3	3	2	2	2	1	1	1	1	1	1	1	1

performing the tasks where only two participants used it for 4 tasks out of 140 tasks (10 participants $\times$ 2 modes $\times$ 7 tasks) when searching for events. This has been taken into account for future versions of the interface. Overall participants appreciated the zoom since it enabled them to get a more detailed view of the video's content.

Participants' Feedback: There were some positive impressions and comments made by the participants about the interface. One participant commented, "YouTube statistics has already a feature that shows you how others viewed your video. Why don't you employ your tool there? It will really help me decide what to watch." Others said, "Is this available in any online videos? Can we try it in YouTube or Vimeo?"; "It is really cool and easy to use. When are you going to apply this to online video websites?"; and finally "I didn't expect others' history would be useful, but, you showed me it is."

5.3 Directions

Our current visualizations were limited to a single video viewing history and they were tested in a controlled laboratory experiment for a short period of time. Thus, we aim to deploy a field study to check the validity and the scalability of these visualizations and navigation mechanism. We intend to explore how users respond to these mechanisms, in conjunction with the presented *VCR*, via a field study

utilizing online video. Extensive data will help determine general users' current viewing behaviour for all types of video, and how it changes when given a *VCR* and other methods based on viewing statistics. We plan to investigate the validity and acceptance of crowd-sourced data as a basis for video navigation, summarization and teasers generation.

5.4 Summary

Viewing heuristics were generated from individual video consumption, or crowd-sourced from many people watching the same video; both provide quick tools for navigating, searching and generating video summaries. In this chapter, we have presented list of thumbnails and *VCR*, two different approaches to visualize a single-video viewing heuristics that provide simple navigation, search, preview and sharing of video intervals. They establish a new way to navigate and view a video space using a personal or crowd-sourced video history. We performed user studies testing these approaches that found positive significant results and highly positive affect and comments from participants. Through these studies, we have demonstrated that applying users' viewing history significantly improved the search and navigation through videos. Moreover, using crowd-sourced data as a tool for recommending segments within videos (i.e. social navigation) was found to be appreciated, and we confirmed that the summaries generated from crowd-popular segments were effective at communicating the content of video. The *VCR* and list of thumbnails were rated highly by users who recommend integrating these mechanism into online video websites.

In the next chapter, we look at how to visualize multiple-video history and how to extend the list of thumbnails approach, presented in this chapter, taking into consideration participants' comments and feedback.

Chapter 6

Multiple-Videos History Visualizations

Video navigation histories are a simple archive that a person can use to easily find a previously viewed video interval. They can navigate to the exact location within the original video by simply clicking on the references within their history. This provides the user with a record for historical navigation and removes much of the burden of relying on memory. However, finding previously viewed content in a navigation history is often a difficult task due to the design, organization, and volume of information to visualize. In the previous chapter, we proposed and evaluated two different approaches to visualize users' viewing history that showed better performance over the state-of-the-art methods. However, those visualizations are designed for a single-video viewing history that lacks the capability of visualizing and managing users' entire viewing history of a video space. Thus, the goal of this chapter is to extend the work presented in Chapter 5 for multiple-video history by testing different design layouts that support user-centred management of history, and to evaluate the benefits this brings.

In this chapter, we describe a Video History System (VHS) framework in Section 6.2.1 that offers users a platform to track and manage their video viewing and navigation history. Section 6.3 presents our proposed visualizations of a detailed multiple-video navigation history: *Video Tiles* and *Video Timeline*. These are both part of the Video History System (VHS) framework, and utilize the same

underlying representation. Section 6.5 reports the results of evaluating the history visualizations against the state-of-the-art method. Finally Section 6.6 addresses the limitations, refinements and directions for future research.

6.1 History Visualization Considerations

Video viewing history visualization as described in previous chapters, is more complex than web browsers' history. Based on the results of our studies, and on participants' comments and feedback, we have defined different design requirements that need to be taken into consideration for history visualization. Below is a list of some of the design requirements that we think are important for a video history visualization. These are:

- A distinction between video timeline and user timeline (i.e. history). As mentioned in Chapters 4 and 5, a vertical timeline is used to represent user timeline while a horizontal is used for video timeline.
- Thumbnail sizes need to be large enough for users to recognize its content, since this can hamper the search task as criticized in Section 4.5.
- Users need to be able to easily navigate and skim the content of each history record (i.e. thumbnail) without the need to play it again in the main player. This was taken into consideration when designing thumbnails as described in Section 4.3.
- History visualizations need to offer access to the list of videos accessed (i.e. similar to YouTube) and the detailed viewing history of each video where users can manipulate each history record.
- History visualizations need to automatically update as a user views and interacts with videos.
- History visualizations need to be easily accessed from the same framework where users view and navigate videos.
- History visualizations should not distract a video viewing action as criticized in Section 5.1.3. Thus, having a different mode to access history can be an

option.

- History visualizations need to allow users to manage their history, for example, to favourite and delete records from history as suggested in Chapter 4 and Chapter 5.
- History visualizations should offer the ability to sort and filter users' history, for instance, sort by mostly viewed or filter by favourite records.
- History visualizations need to have a fast response to users' interactions with its features.

Some of these considerations have already been applied into the design as has been seen in the previous chapters, while some will be applied in the new proposed designs to offer users more accessibility and options for their history.

6.2 Multiple-Video Navigation Interface

Our goal is to provide users with intuitive access and management of their personal video viewing history. This requires an interface that can capture and visualize the detailed navigation history of the videos viewed by a user. To be able to test our proposed visualizations, we need an interface to support realistic video watching experiences comparable to commercial systems so that the visualizations are integrated with the same fidelity. Thus, due to the lack of such interfaces, we developed a prototype viewer, which we call Video History System (VHS), for evaluating history visualizations. This interface is a progressive work of what we presented in Chapter 4 and Chapter 5. In this section, we give a detailed description of this interface, the different visualizations designed for the history, and the scalability of the visualizations.

6.2.1 Video History System (VHS)

The *VHS* is based on three modes: the *video library*, the *viewer* and the *history*. Users can easily switch between modes using the navigation controller.

Video Library

The videos library (shown in Figure 6.1) provides a user with a grid of available videos or a user can simply load a video from their disk. It allows the user to select a video he/she would like to view. Once they choose a video to view, the interface transitions to the viewer mode.

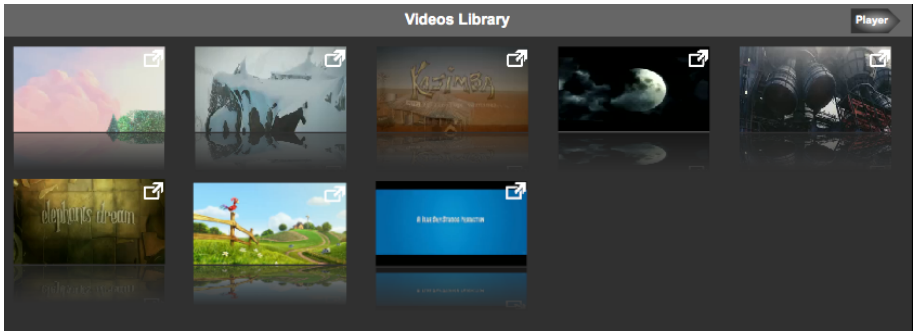


Figure 6.1: Videos Library Mode from which users select or open a video they would like to view and navigate. Each video is represented as a small video segment that can be dragged to its top right corner (white square) to start playing in the viewer mode.

Viewer

The viewer represents the core component of the interface, allowing users to watch video while their history is captured. The viewer consists of a video player and a filmstrip, as presented in Section 5.1.2. The viewer's video player (described in Section 4.3) and shown in Figure 6.2, allows a direct control of the selected video. The viewer also contains a filmstrip, shown under the player, which is described in Section 5.1.2 to allow fast navigation of the current video. Each video segment in the filmstrip can be individually searched (via seek) and played to minimize the time needed to search for a specific event. The filmstrip supports level-of-detail manipulation via zooming using mouse wheel gestures with the cursor over the target segment.

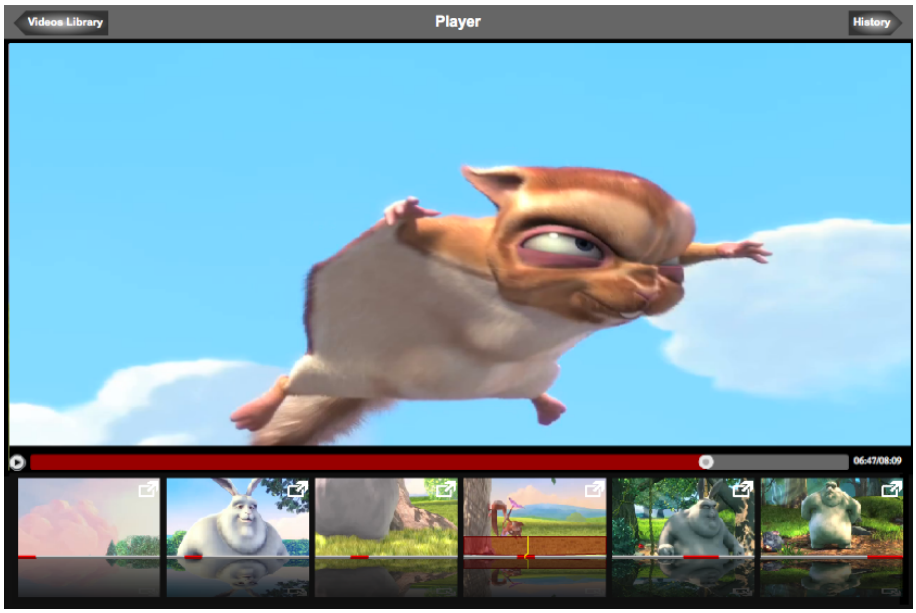


Figure 6.2: Video playback is performed and controlled within the viewer. The video player occupies the majority of the space; the video can be played/paused using the dedicated button below the player (on the left) or by clicking on the video itself; seeking is controlled via the white circle playhead or simply by clicking/dragging on the red/gray video timeline. The filmstrip below the player provides real-time previews based on the cursor position, allowing faster navigation of the video.

History

Every time a user seeks or plays a different video/segment within the main player, a new record is added to their history. The user can switch to the history mode shown in Figure 6.3 by clicking on the 'History' button. Each record in the history is visualized as a variable-sized video segment (described in Section 5.1) with the size indicating how often that interval was viewed. Previous work applied varying colour intensities to indicate the importance of each segment [82]. We used size, since we are representing intervals using thumbnails where colour discrimination would be confused with the thumbnail content and would be difficult to differentiate for some videos.

Video Segment size is determined based on a weight factor, which is derived

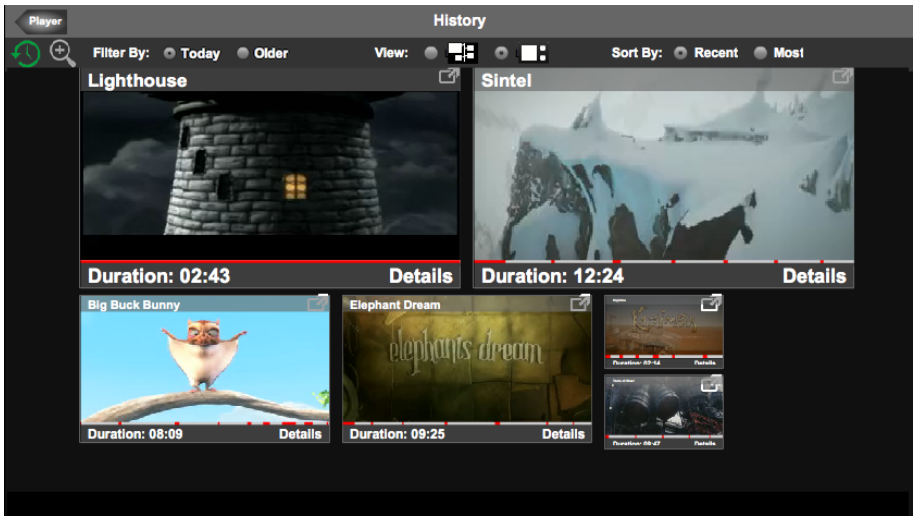


Figure 6.3: Each history visualization (presented in the History mode) displays the user's navigation history, and provides top-level access to all previously viewed videos from which the user may zoom into any history entry for more detail. The history can be filtered by date, sorted by time or popularity, and the type of visualization used can be chosen. The thumbnails' size is based on the view count within each video.

from how often the segment has been viewed as well as its duration. For our study, we used a weight factor of 75% of the peak view count plus 25% of the duration in minutes (normalized duration is not used to ensure short segments are not biased to a larger size). We believe that when a user views the same part multiple times then he/she has more interest in that part than just watching a longer segment once. We define three different sizes for segments within our interface based on this weight factor: small, medium and large. A segment is considered small when its viewing weight is less than or equal to one-third of the maximum viewing weight among all the displayed segments, medium when the viewing weight is between one third and two-thirds of the maximum viewing weight, and large when it is larger than two-thirds of the maximum viewing weight. Thumbnail sizes are proportional to each other such that they fit the tile template patterns, shown in Figure 6.7. For our interface, we use 320×180 (large), 210×118 (medium) and 100×56 (small), measured in pixels on a one-to-one PPI display. The organization of these segments

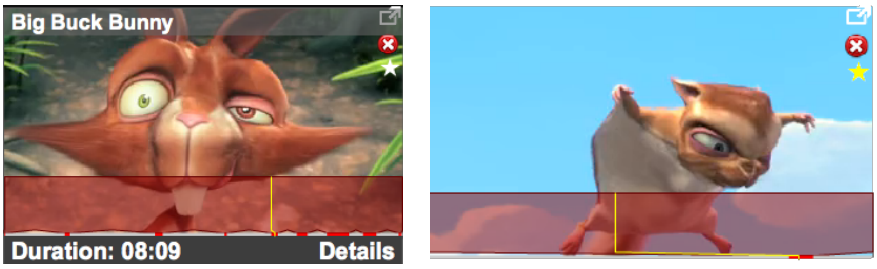


Figure 6.4: The user's history is visualized as a set of small video segments. An *intra-video segment* (left) is used to visualize an aggregated history of a single video as one thumbnail, where the union of its viewed intervals are visualized in the thumbnail timeline (red/gray bar) and the combined segment has a single seek bar. The *single video segment* (right) represents a single interval from the history, and it is the furthest possible zoom level.

depends on the visualization layout selected, which is described in Section 6.3.

The history visualizations contain an inter-video history, which shows the different videos the user viewed, with all intervals watched within a single video aggregated into an *intra-video segment* (Figure 6.4, left). The union of the viewed intervals of a single video is easily visualized in the timeline of the video segment representing the accessed video in the history. To access the detailed history of a specific video, the user can click on the 'Details' link, which opens a detailed history of the selected video only. The detailed history has a similar structure to the history mode with the exception of the removal of the Details info for the detailed history video segment as shown in Figure 6.4 (right).

The history visualization offers users the ability to manage their history where each video segment in the history can be favourited and deleted. Users had the option to edit their navigation history by hiding the intervals they did not want to keep using the button in the corresponding video segment or favouriting the intervals they liked using the button.

6.3 History Visualization Designs

Our main goal is to provide users with a history management tool via an effective visualization. Based on a review of the literature (Chapter 2) of existing interfaces and services, we made various design decisions to create a visualization of a complete history of a user's video navigation. Since the list (Figure 2.5) and grid (**??**) layouts are the most commonly used visualizations for browsing history and due to the large number of users viewing videos online, we decided to apply these layouts for our interface due to their familiarity by users and their applicability to what we are trying to visualize. We intended to investigate how users would preserve these layouts for video history and how they would help them accomplish their tasks compared to the previous design.

From our pilot study (described in Section 6.5.4) investigating the time taken to search for events, there was no significant difference between the visualizations in terms of time needed to find the events. However, users' comments indicated a strong liking (6.3 on a 7-point Likert scale) for the interface and commented that they enjoyed it. User feedback inspired us to extend the list and grid interfaces by utilizing the size of thumbnails for heuristics (in both) and expanding the list to occupy a full screen (by using two lists). Our refinement ended with two novel visualizations: Video Timeline and Video Tiles, as described below. These visualizations were tested to evaluate their performance when performing a search task, while also measuring user satisfaction.

6.3.1 Video Timeline

This visualization of video history was an extension to the familiar list visualization (and our tested single-video visualization described in Section 5.1), with the exception of having two columns of variable-sized thumbnails instead of just one. It is designed to display a large number of thumbnails within a small area, while also maintaining an explicit order, by dividing them into two columns along the user's vertical time-line.

For users to relate these thumbnails to their occurrence in time, they are attached to the user's vertical timeline, where the attachment location indicates their order with respect to the other thumbnails in the history, as illustrated in Figure

Figure 6.5: *Video Timeline* is a multiple-video history visualization design that attaches history segments to a user's vertical navigation timeline using two columns based on their view count and chosen order.

Figure 6.5. By default, the thumbnails are in reverse chronological order (based on user time); however this order can be changed to interval start time (*Time*, only available in detailed history), recently viewed (*Recent*), and viewing weight (*Most*) by using the corresponding radio buttons shown at the top right of the Figure. The interface also supports filtering the user's history based on favourite segments and/or intervals that have been watched twice or more (these features are available in both visualizations).

6.3.2 Video Tiles

This history visualization was designed to take advantage of the entire screen space, using an implicit ordering. This design is essentially a grid layout of thumbnails, which is commonly used for web browsing history, with the exception of having varied sizes for the thumbnails in our design. Moreover, in our design the grid represents a detailed video history where each grid cell corresponds to an interval within a video. Using this design allows us to display more thumbnails at once

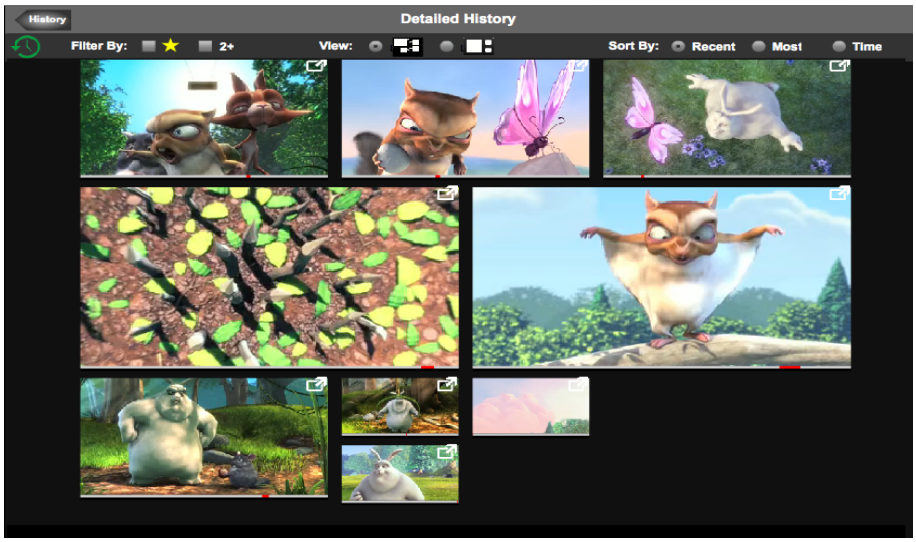


Figure 6.6: *Video Tiles* is a multiple-video history visualization design that displays history segments based on template patterns following Algorithm 4 based on their view count and chosen order.

with less scrolling needed when searching for video segments.

In our design shown in Figure 6.6, we used seven different templates (Figure 6.7), where the order of the segments and their sizes determines the template to be used and the location of segments within these templates. The template is selected based on the conditions explained in Algorithm 4. Within each template, thumbnails are displayed top-to-bottom and then left-to-right based on their order within history.

6.4 Visualization Scalability

As users view hundreds of videos and interact with them, their history will continue to grow. Clearly our visualizations must scale with the data being recorded. In our visualization, the length of a video does not have an effect on the visualization when it is viewed passively (i.e. no user interaction) where it is represented by one video segment. However, the number of interactions or seek actions a user performs with any length of a video and the number of videos viewed determines

Algorithm 4 Pattern selection: every possible pattern is compared against the current set of thumbnails to find a match.

```
P ← {Pattern 0, Pattern 1_0, Pattern 1_1,..., Pattern 6}
Retrieve ordered set of thumbnail sizes S = {S_1, S_2,..., S_n}
index ← 1; pattern ← nil
repeat
  for each p ∈ P do
    S_p = {S_index, ..., S_(index+size(p)−1)} ⊂ S
    if S_p = p then
      pattern ← p; break
    end if
  end for
  if pattern then
    Apply pattern
    index ← index+ size(pattern)
  else
    Re-order the thumbnails to fit one of the patterns
  end if
until index > n
```

the size of the history. One way to address scalability issues is to keep the size of the visualized history down, by choosing some policy to limit the number of segments visualized on screen at any one time. This leads to the need for features and interaction techniques to be able to bring the other history segments into view.

Web browsers visualize a user's browsing history by date. For instance, Google Chrome uses a fixed number of data entries to be visualized per page starting from the last visited URL arranged by date with the option of viewing previous history using the 'Older' link. When 'Older' is activated, three other options are provided: 'Newest', 'Newer', and 'Older'. Some of these web browsers provide users with filters that can be applied when viewing their browsing history, for example, by date, name, most frequently visited, and most recently visited. In our visualization for the inter-video history, we follow this approach by limiting the number of visualized items at any time and provide access to the older history items, as shown in Figure 6.3. For the detailed history visualization, we group close segments in history together as a stack of thumbnails when the number of elements exceeds the limit, as shown in Figure 6.8. Using the mouse wheel on these stacks, we can zoom

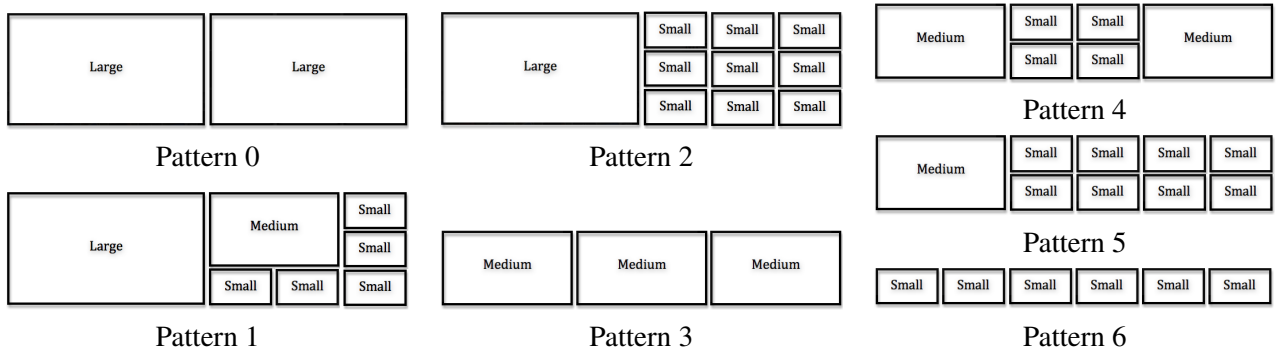


Figure 6.7: Template patterns are used for the Video Tiles visualization, to provide a clean set of thumbnails with an implicit order (top-to-bottom, left-to-right, based on a single pattern). Patterns 1, 2, 4, and 5 have alternatives where either the entire pattern is reflected or just the portion containing medium and small tiles: Pattern 1 has 8; Pattern 2 has 4; Pattern 4 has 6; Pattern 5 has 5. These are applied using Algorithm 4.

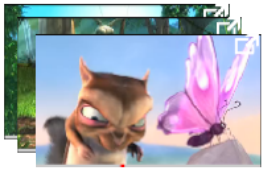


Figure 6.8: A thumbnail stack is used when the number of segments to be visualized exceeds the limit - this addresses scalability of the history visualization. The user may zoom into or out of the stack via a mouse wheel gesture to obtain control over the presented level-of-detail.

in and out, which reveals a new visualization of only that zoomed stack. The front thumbnail of the stack is represented by a seek-able and playable video segment while the other thumbnails in the stack are representative images. Our visualization also supports filtering by most frequently viewed, most recently viewed, starting timestamp, favourite intervals, and intervals viewed more than once.

6.5 Investigating Multiple-Video History Visualizations

A user study (approval certificate #: H08-03006) was performed to evaluate the different visualization layouts for navigation of video history. Our aim was to investigate the visualization layout that would be most efficient when finding previously watched segments within videos, the visualization that would make history navigation easier, and the layout that users prefer. Thus, we designed comparative studies where the participants were asked to freely watch a set of videos and then find and share their liked segments from their personal history with two different visualization layouts. We also tested whether these history visualizations performed better than without detailed history where users simply find a video from a grid of previously seen videos (Figure 6.1) and then search for intervals using the video's filmstrip (similar to what they would have in an online video website such as YouTube).

6.5.1 Apparatus

The experiment interface was developed in Flash CS4. The experiment ran on an Intel dual-processor dual-core 3 GHz Mac Pro desktop with 8GB RAM and

equipped with a 24" Dell LCD monitor with a resolution of 1920×1200 pixels at a refresh rate of 60Hz. A Microsoft optical wheel mouse was used as the input pointing device with default settings and the Adobe Air environment was set to 860×700 pixels while running the Flash program.

6.5.2 Design

Three different visualization layouts were tested. In the pilot test: List, Grid and Filmstrip were investigated, while in the actual study: Video Tiles, Video Timeline, and Filmstrip were examined. The layout order was alternated between subjects to eliminate the mode order effect. Each participant used the three visualizations to find seven different clips using each layout. For each clip, participants had to find the corresponding video first using either inter-video history (Figure 6.3) when one of the proposed visualizations was experimented, or the videos library (Figure 6.1) when Filmstrip was tested. After finding the video, participants searched for the questioned clip using the provided layout.

Each participant freely watched a set of five different short videos of length between 3 and 5 minutes (the same videos described in Section 5.2.3). They had the option to edit their navigation history by hiding the intervals they did not want to keep using the button in the corresponding video widget or favouriting the intervals they liked using the button. They were instructed to keep at least seven segments in the history of each video to be able to start the tasks. Each participant performed 21 search tasks (3 layouts $\times$ 7 segments per mode). The participants were asked to find the segments as quickly as possible. For each task, the completion time, the number of previews, and the number of scrolling events were recorded. The completion time was measured once the participant clicked on the Find button until the moment the researcher advanced the task for the participants based on the submitted segments. The navigation heuristics were also recorded during the viewing phase. The participants were asked to rank the different visualization modes based on their speed, ease and preference (1=best, 3=worst). The experiment lasted approximately one hour per participant.

6.5.3 Procedure

The experiment proceeded as follows:

1. The researchers gave the participants a walkthrough of the interface, explaining the functionality of each feature and tool within the interface and their effects when applied to a video. This stage took approximately five minutes. The participants were allowed to try the interface and ask any questions during this stage.

2. Participants were then asked to watch five different videos and re-watch any affective parts or the parts they preferred. A single video was played at a time. Participants were allowed to re-watch bits within videos. However, they could not start the searching tasks until they had watched the five videos and had at least seven segments in their navigation history of each video to make sure that they have created enough navigation history for each video. This navigation history was stored to be used for the sharing tasks.

3. Once each video was viewed, the participants were asked to name five different segments they preferred within the video, which they would share. The experimenter kept a record of these events. Once the participants named the events, they were advanced to the next video.

4. When all the videos were completely viewed, the experimenter provided the participants with the events they need to share, one event at a time. Participants started the sharing task by clicking on a 'FIND' button. Once the 'FIND' button was clicked, the history visualization was modified to display the visualization layout that is related to the task mode. Participants had to search for the segment by finding the video first from the history and then go to its details to find the required segment to be shared. In the filmstrip mode, participants had to find the video from the videos library and then used filmstrip to find the questioned event. When they found the event, they needed to play the segment on the main player. When the researcher approved the submitted segment, the participant proceeded to find the next segment. Upon the completion of finding the seven segments, the participants were advanced to

the next visualization mode to share another seven segments from their history.

5. The experiment ended once the participant had experimented with all three visualization layouts with seven segments for each mode. Finally, the participants were asked to fill out a questionnaire (Section B.5) where they ranked the modes and gave their feedback and comments about the interface, its features and their experience.

6.5.4 Pilot Test

The pilot study was conducted with five participants where each subject tried three modes to find their previously seen affective segments. It was carried out to determine user feedback and also, to test the history visualization and the experimental design. In this pilot, two history visualizations, which were chosen due to their familiarity by users viewing videos online and their applicability to what we are trying to visualize, were tested against the Filmstrip. These visualizations are:

List: The List visualization simply displays the history segments on one vertical strip as shown in Figure 6.9(top), where these segments are organized in a reversed chronological order having the last viewed segment at the top of the list. This layout is similar to our previous visualization described in Section 5.1.1, with the exception of having varied sizes for the thumbnails.

Grid: History segments in this visualization are organized into $n \times m$ grid of thumbnails as demonstrated in Figure 6.9(bottom), where they are ordered left to right and then top to bottom. The recently viewed segment appears at the most top left corner while the first viewed segment appears at the most bottom right thumbnail. The order of the thumbnails is changeable based on the user preference using the sort radio buttons. In this visualization, all thumbnails are displayed using a fixed size (medium size), which does not take the viewing weight into account. The fixed size of the thumbnails allowed this visualization to display more thumbnails at a time compared to the other visualization. Thus, less scrolling and searching time are needed for a certain segment.

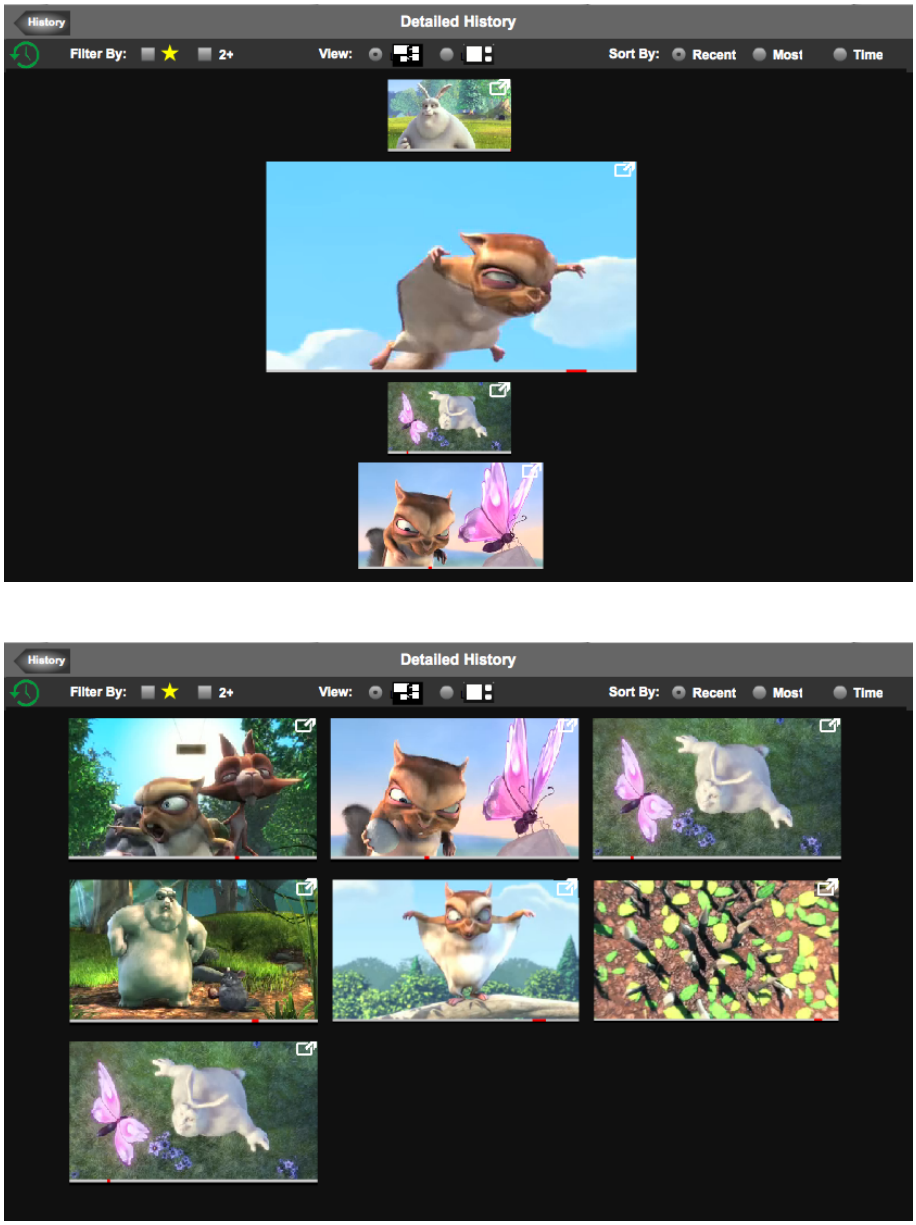


Figure 6.9: The piloted visualization designs for multiple-video history: *List* (top) and *Grid* (bottom). In *List* design, history segments are displayed in one vertical scroll-able column, while in *Grid* design, they are visualized on an $n \times m$ matrix going from left to right top to bottom.

6.5.5 Results and Lessons Learned from Pilot Test

The pilot study showed positive results on the features of the interface which encouraged us to continue running the experiment. For the finding task, each participant was able to complete nine search tasks in less than 40 seconds per segment. The mean completion time for the search task using the different modes is illustrated in Table 6.1. Search using the Filmstrip (no detailed history available) took more time than when having detailed history using the different visualizations. This indicates that having access to the user's detailed personal navigation history leverage finding previously seen interested parts within videos.

Table 6.1: Pilot study means of the completion time (in seconds), the number of previews, and the number of scrolling events for each mode. The results demonstrate that history-based search for personal affective intervals is better than search using Filmstrip and the video library.

	List	**Grid**	**Filmstrip**
Completion Time	18.78	16.47	22.52
No. of Previews	17.53	27.13	44.93
No. of Scrolls	17.27	1.13	-

Even though the completion time was not significant, from Table 6.1 we can see that the more history segments (thumbnails) displayed by the visualization in one view (i.e. less interaction to view more thumbnails), the less scroll events occurred and the less time was needed to find the segment. This was anticipated, since all the sizes makes the thumbnails content recognizable and easily identified among other thumbnails. The results for the Grid visualization support this conclusion where few scrolling events occurred and it was the fastest on average.

Participants ranked Grid as the most liked visualization, then List, and Filmstrip ranked as the least liked mode. However, for the fastest mode, they ranked Grid as the fastest, then Filmstrip, while List came last. The quantitative results for the List mode contradicts with the participants' ranking. Some participants mentioned that the List layout needed more scrolling which made them think they were slower than the Filmstrip. However, they liked having different sizes for the thumbnails to indicate how an interval was consumed. This made some participants

favour the List visualization over the Grid, which is worth investigation.

Based on these pilot results and users' feedback, we extended the two visualization layouts as explained in Section 6.3 to utilize different sizes for the Grid layout and including more thumbnails in the List design. To evaluate the new proposed visualization, we ran the same experiment but with the new visualizations: Video Timeline and Video Tiles. Moreover, to illustrate the differences between methods, more subjects were recruited for the experiment and more tasks were performed per method. In this experiment, we hypothesized that users would perform better when detailed history is available in terms of time needed to find their previously seen preferred segments, and that less navigation would be required (**H1**). In terms of the visualization layouts, we anticipated that the Video Tiles layout would outperform Video Timeline layout for time and scrolling needed, since more thumbnails can be viewed at once (**H2**). However, in terms of seeking or navigation events used, we expected no difference, since they both have the same thumbnail representations (**H3**). For modes preference, we predicted that it would highly depend on participants' familiarity with List and Grid layouts (**H4**).

6.5.6 Participants

Ten paid volunteers, 6 female and 4 male, participated in the experiment (different from those in the pilot). Participants ranged in age from 19 to 40. Each participant worked on the task individually. All participants were experienced computer users and have normal or corrected to normal vision. Seven participants watch online videos on a daily basis, two watch videos 3-5 times a week, and one watches once a week. Five of the participants watch 1-3 videos on average per day, while three watch 4-6 videos per day and two watch more than 10 videos per day on average. Table 6.2 presents participants' details.

6.5.7 Results and Discussions

The study showed strongly positive results for the interface. Each participant was able to complete all 21 search tasks in less than one minute per segment.

Table 6.2: Demographics summary for participants in the investigation of the multiple-video history visualization study. (Note: WMP = Windows Media Player, QT = QuickTime, VLC = VLC Media Player, RP = Real Player, KMP = KMP Player, Gom = Gom player, M = Mplayer, YT= YouTube)

P	Gender	Age Group	Vision	Watching Frequency	Videos per Day	Familiar Players	Frequent Player
1	Male	31-40	Normal	Daily	1-3 videos	iTunes, QT, RP, VLC, WMP	QT
2	Male	26-30	Corrected	Daily	more than 10 videos	iTunes, M, QT, RP, VLC, WMP	YT, KMP
3	Male	19-25	Normal	Daily	4-6 videos	iTunes, M, QT, RP, VLC, WMP	iTunes
4	Female	19-25	Corrected	Daily	4-6 videos	QT, VLC, WMP	WMP
5	Female	19-25	Corrected	3-5 times a week	1-3 videos	RP, VLC, WMP	VLC, WMP
6	Female	19-25	Normal	3-5 times a week	1-3 videos	iTunes, VLC, WMP	VLC
7	Female	26-30	Normal	Once a week	1-3 videos	RP, VLC	RP
8	Female	19-25	Normal	Daily	1-3 videos	iTunes, RP, VLC, WMP	VLC
9	Male	31-40	Normal	Daily	more than 10 videos	RP	YT, RP
10	Female	19-25	Normal	Daily	4-6 videos	VLC, WMP, Gom	WMP

Table 6.3: Performance comparisons for the three multiple-video history methods, using the F-test for equality of means. The results demonstrate that history-based search for personal affective intervals is more efficient than search using Filmstrip and the video library. Notes: SD is Standard Deviation; Completion time is measured in seconds; $^{*}p < 0.01$.

	Video Timeline		**Video Tiles**		**Filmstrip**		
	Mean	SD	Mean	SD	Mean	SD	F-test
Completion Time	17.76	5.28	17.88	4.91	24.31	10.42	23.01*
No. of Previews	25.11	28.24	23.41	21.03	35.53	27.56	5.03*
No. of Scrolls	6.64	7.74	1.54	3.96	-	-	28.72*

Search Task: A within-subjects Analysis of Variance (ANOVA) test was carried out to examine the significance of the results in terms of the average completion time per segment, average number of previews and average number of scrolling events. The results as illustrated in Table 6.3 showed a significant difference between methods in terms of time needed to find segments, average number of previews and average number of scrolls performed while searching for events. A post-hoc analysis showed that the filmstrip had significantly more previews and took more time than the other modes, as we hypothesized. This can be explained by the fact that when searching using personal history, participants had a defined context and mental map of their segments [83]. Some participants mentioned that for some videos, it was harder to find events due the similarity of the content of the Filmstrip segments, which made history visualization better in this case. Moreover, when analyzing the participants' navigation history, we found that participants created eleven history segments on average per video. This means that when using history, they had to search for their segment among eleven different thumbnails, whereas when using the video's filmstrip they had only six different segments. Despite this, they performed even better when using history as illustrated by the results confirming that history helped them completing the task.

Difference Between Proposed Designs: For the history visualization layouts, the Video Timeline method was not significantly different from the Video Tiles method in both the average number of previews and the completion time. The non-

significant results of the completion time might be due to the confusion caused by the flow of the Video Tiles mode as discussed below. However, a significant effect was observed on the number of scrolling events, where Video Tiles had significantly less scrolling events than Video Timeline. This proves the second part of our hypothesis and the third hypothesis. The more history segments (thumbnails) displayed by the visualization at one view (i.e. less interaction to view more thumbnails), the less scroll events occurred and less time was needed for searching. This was anticipated, since all the sizes make the thumbnails' content recognizable and easily identified among other thumbnails. The results for the Video Tiles visualization support this conclusion where less scrolling events occurred in comparison to the Video Timeline visualization. Nevertheless, as the history grows, it becomes almost impossible to view all the segments at once without the need for scrolling when different thumbnail sizes are used. If the scrolling event is disabled following [34], then thumbnails size will shrink as the history builds up making the thumbnails difficult to identify. As is known with visualization scalability of any system, there is a trade-off between size and scrolling or zooming needs.

Participants' Mode Ranking: When participants were asked to rank the different modes for preference, ease and speed, they ranked Video Timeline as the most preferred, easiest and fastest visualization, and then Filmstrip. Video Tiles ranked as the least liked mode. However, the quantitative results for the Filmstrip and Video Tiles modes contradict with the participants ranking for speed, where Filmstrip came last quantitatively and second qualitatively. Some participants mentioned that the flow of the thumbnails in the Video Tiles mode created some confusion which they thought made their performance worse, which we can see is not true. Nevertheless, as we mentioned earlier, a proposed visualization needs to ease the user task and meet their satisfaction level. Thus, this visualization needs to make the sequence of the thumbnails more obvious. Participants pointed out that the vertical timeline indication used in the Video Timeline mode helped them easily visualize the relationship between the thumbnails. Moreover, some participants indicated that since they are familiar with Video Timeline layout, it was easier for them to recognize the sequence of the thumbnails and understand the flow. This aligned with what we predicted in our last hypothesis. Thus, more exposure time

to the Video Tiles visualization might also help in understanding its representation and flow.

Features' Ranking & Usage: In terms of ease and usefulness of the interface components and features, the average ranking across all components and features was 5.82 out of 7. All features were ranked above 5 except for two items: the ease of finding previously seen segments using the Video Tiles visualization ($M = 4.7$), and using the zoom in/out functionality ($M = 4.3$). The low ranking for the ease of finding previously seen segments using the Video Tiles visualization was due to the confusion of understanding the layout and the thumbnails sequence as participants commented. For the zoom in/out functionality, some participants pointed out that it was a bit confusing where a small mouse-wheel gesture triggered multiple zoom-in events, causing some frustration. This could also explain the low usage of this feature while performing the tasks; only two participants used it when searching for events. The problem with this feature can be resolved by reducing the sensitivity of the mouse wheel. Participants appreciated having this functionality, which helped them to get a more detailed view of the Filmstrip segments.

For the frequency of the interface features usage, there were some features that were rarely or never used during the task. For instance, sorting history elements based on time was only used 27 times in the 140 different tasks (10 participants $\times$ 2 modes $\times$ 7 tasks), while sorting by recently viewed, or mostly viewed were never applied. This might be due to the cues provided by the sizes for the mostly viewed, and the interval representation within each thumbnail for the time indication. For the filters, the favourite filter was applied by one participant 12 times in the 140 tasks, while the '2+' filter was also used by a single participant for 14 out of the 140 different tasks. The low usage of the favourite filter might be due to the need of favouriting segments first and then users could apply this filter. Even though most participants favourited at least 2 segments per video, they did not apply it when performing the tasks. Most participants, when asked at the end of the experiment why they have not used some features, stated that there were so many features, which overwhelmed them and made it hard to remember when performing the task. For the '2+' filter, the participant who applied it found filtering out unwanted segments made it easier and less overwhelming when searching for what

they liked. This can be due to the behaviour participants applied when viewing the provided videos. Participants tended to watch the entire video first and then used the Filmstrip to re-watch the segments they liked the most. Thus, their history had only one extra segment for the entire video that had no impact, since it was obvious from their other history segments.

Participants' comments: Participants were also asked for their suggestions and comments on each interface component. Some participants asked for a possibility to show/hide the Filmstrip while viewing because sometimes they found it a bit distracting. Others proposed to have a button to allow a full screen view of the video. Similar to other online video websites, we can allow a full screen view with the option of showing the Filmstrip component once the cursor is moving over the video and the ability of locking it to be visible all the time. Participants found having different thumbnail sizes were very useful for finding segments from a detailed video history; however, for the inter-history visualization, the different sizes used for the videos thumbnails was criticized as being confusing. Participants indicated that sizes did not make any sense for them when used to indicate the importance of the different videos. We can represent videos with the same size while the sorting feature based on mostly viewed will allow participants to indicate the importance of their videos. Participants wanted to try more visualization layouts for the history to get some insights on the importance of the features provided.

6.6 Directions

Video Timeline and *Video Tiles* showed highly positive results and we received valuable comments from participants, which motivates us to present our interface to a larger community. Thus, we plan to design a field study where these visualizations can be investigated on a large scale where different tasks and procedures can be used to evaluate the video navigation history. Through the field study, we will further explore the different visualization approaches (e.g. Graph and 3D visualizations) and interactions required when using multiple-video navigation history in a less controlled setting to further refine the utility of our approach. Based on our observations from the study and the results in this work, we are inspired to

perform future empirical studies within social website and educational contexts to investigate how this could change the way users consume and navigate video.

6.7 Summary

To summarize, we have presented two new methods, Video Timeline and Video Tiles, to visualize, manage, and navigate a video space using a personal video history. These methods could be integrated easily in a video viewer together or individually as demonstrated in our Video History System (VHS). These visualizations are based on observations of web browsing visualizations and the increasingly temporal nature of video navigation. We performed a comparative study based on a use case of fast searching and sharing, and found significant results in favour of the Video Timeline method. The visualizations were positively perceived and showed significantly faster times for finding previously seen parts when detailed history is used. We conclude that visualizing history is a valuable addition to any video navigation interface and the visualizations we have designed are effective and useful.

In the next chapter, we look at another issue with current video navigation, which is object selection in interactive videos. We introduce and validate a new interaction technique to ease the selection problem of moving objects in video.

Chapter 7

Object Selection in Videos

Another way to offer affordances within a video interface is to use the video content or simply the objects within a video. Once these objects are annotated or tagged, they can serve as hyperlinks, which then can be used to navigate the video content, retrieve more information about these objects or simply to direct users to other related videos. This kind of affordances would enrich the video experience. However, one of the fundamental issues with this form of affordance is the selection of the interactive links or objects within these videos to traverse to the next piece of information. Due to the time-based nature of these videos, the embedded clickable anchors or annotations are visible or active for only a certain duration of the video in comparison to web pages in which hyperlinks are present at all times. Therefore, the activation and selection of these hotspots becomes difficult, as they are affected by the shape, size and the location of the object at a given time of the user's selection. Some of these factors have been studied and analyzed extensively for stationary targets while moving targets lack similar consideration. Target acquisition becomes a critical task when trying to select moving targets at different speeds and sizes, such as that which occurs in interactive video where active objects in the video are moving. This type of interaction is emerging as video becomes a ubiquitous media on the web for example. Previous research has investigated methods to ease the selection task by introducing various effective selection techniques [52, 62, 89].

In this chapter, we focus on the selection of objects within interactive videos.

To propose new techniques that alleviate the problem of selecting moving objects, we need to understand the effects of target factors in selection and how those factors interact with human cognitive, perceptual and motor control systems to optimize user performance and interaction design. Section 7.1 describes a derivation of a new model for target acquisition, which extends Fitts' Law [39] to accommodate moving targets in both one and two dimensions. The model we present calculates the index of difficulty and movement time for moving targets by incorporating target speed, relative direction and movement angle. Section 7.2 introduces our novel selection technique that we call *Hold* (also referred to as *Click-to-Pause*) to alleviate a moving target acquisition problem by temporarily pausing the content while selection is in progress to provide a static target, which removes the speed factor presented in the model. Section 7.3 outlines our experiment (approval certificate #: H10-01897), which validates the model and evaluates *Hold* against a traditional approach, and we discuss the results of the experiment and how the selection techniques fit within our model. Section 7.4 illustrates an application of our selection technique, *Hold*. Finally, Section 7.5 reflects on the impact of the technique and provides direction for future research.

7.1 Modelling Target Acquisition

Rapid aimed movements can be characterized using two different motor control models: the iterative corrections model (Fitts' model [39]) or the impulse variability model (Schmidt et al. model [103]), which depend entirely on the task demands. Wright and Meyer [119] have indicated that the iterative corrections model applies to tasks with spatially constrained movements while the impulse variability model applies to tasks with temporally constrained movements. Spatially constrained movement tasks are those where movements end within a target region while trying to minimize the average movement time. Temporally constrained movement tasks are those where movements are initiated having a specified duration in mind while the movements end nearby a target point, not a region. Since our approach involves spatially constrained movements, we characterized and derived our model by extending Fitts' law.

Fitts' Law [39] is the most commonly used approach to study new acquisition

techniques, since it was the first successful model to predict the time required to complete an acquisition task. The index of difficulty (ID) is modeled on a logarithmic scale depending on target width (W) and distance from cursor (D); movement time (MT) is modeled as a linear relation of ID:

$$ID = log_2\left(\frac{D}{W} + 1\right) \tag{7.1}$$

$$MT = a + b \times ID \tag{7.2}$$

where a, b are empirically determined constants. This equation was initially proposed for stationary targets in 1D, which also assumed that the direction of movement is collinear with W.

For moving targets, Jagacinski et al. [64] and Hoffmann [58] showed that Fitts' model, shown in Equation 7.1, failed to accurately predict the acquisition time for moving targets at a constant speed. They found that including target speed in the index of difficulty of a task showed an excellent fit with the mean acquisition time. Jagacinski et al. [64] used empirical data from their study to derive an estimate of the index of difficulty for pursuing a moving target in one-dimension (1D). Hoffmann [58] gave three different extensions to Fitts' Law: using a first order continuous control system, a second order continuous control system, and a discrete response model.

Therefore, the model used to predict the acquisition time of moving targets must include target speed. To accommodate target speed in the model, we applied a model proposed by Card [19] and extended it to moving objects. Each acquisition task involves a ballistic phase and a corrective phase, which could be considered as smaller acquisition subtasks. Card used these subtasks to derive a model similar to the approach taken in the steering law for a straight path [1]. The acquisition time is determined by human cognitive, perceptual and motor control systems. Each single movement needs perceptual processing time (τ_P), cognitive processing time (τ_C) and motor processing time (τ_M) to move the hand towards the target. Therefore, n correction movements will need $n(\tau_P + \tau_C + \tau_M)$ of time to capture the target [19].

By applying Card's model, we took the remaining distance after each move

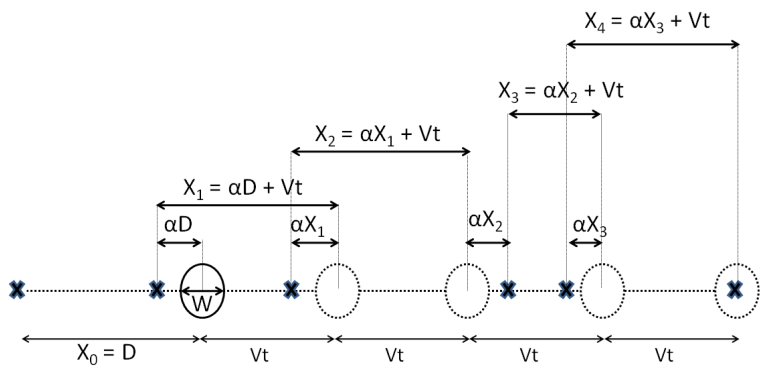


Figure 7.1: Analysis of a moving target in 1D

where the initial remaining distance (X_0) is the initial distance between the cursor and the target D. Let the relative accuracy to reach a target in each move be α, then the remaining distance after the first move will be $X_1 = \alpha X_0 = \alpha D$. We should take into consideration that our target is moving at constant speed V, which by the end of the first move after some time t will have already moved Vt. Then X_1 will be $\alpha D + Vt$ as the target moves away from the cursor, which is illustrated in Figure 7.1. Continuing with this argument, we get the following for n moves:

$$\begin{aligned} X_n &= \alpha X_{n-1} + Vt \\ &= \alpha^n D + Vt(1 + \alpha + \alpha^2 + \cdots + \alpha^{n-2} + \alpha^{n-1}) \\ &= \alpha^n D + Vt(1 - \alpha^n)/(1 - \alpha) \end{aligned}$$

using a geometric series. If the target is captured at this point then X_n should be $X_n \leq W/2$ and if we solve for n:

$$n = -log_2 \left(\frac{D \pm \frac{V}{K}}{\frac{W}{2} - \frac{V}{K}} \right) / log_2 \alpha \tag{7.3}$$

where $K = (1 - \alpha)/t$, which is empirically determined and the $\pm$ indicates the direction of movement i.e. towards or away from the cursor. Then the acquisition

time and index of difficulty is:

$$MT = n(\tau_P + \tau_C + \tau_M) \tag{7.4}$$

$$= -\frac{(\tau_P + \tau_C + \tau_M)}{log_2 \alpha} log_2 \left(\frac{D \pm \frac{V}{K}}{\frac{W}{2} - \frac{V}{K}} \right) \tag{7.5}$$

$$ID_{C1} = log_2 \left(\frac{D \pm \frac{V}{K}}{\frac{W}{2} - \frac{V}{K}} \right) \tag{7.6}$$

This model coincides with Hoffmann's first order continuous-control model [58] for the moving targets.

To modify this model to account for moving target selection in 2D, we look first at how the acquisition time of stationary objects in 2D is modeled and then extend it to moving targets. For 2D targets, several factors should be taken into consideration beyond the target width and distance constraints of the 1D Fitts' Law. 2D pointing is constrained by the target area, the location of the target from the cursor (i.e. the 2D vector representing angle and target position in 2D). Bivariate pointing was first studied by MacKenzie and Buxton [76], where they tested five different formulae to model the index of difficulty and found two of them fit with their experimental results. Their first correlated formulation substitutes the magnitude of the target in the direction of movement (W') for W and thus:

$$ID_{W'} = log_2 \left(\frac{D}{W'} + 1 \right) \tag{7.7}$$

Their second formula, which is highly correlated to their experimental data, substitutes the smaller value of a target's width (W) and height (H). The index of difficulty is then:

$$ID_{min} = log_2 \left(\frac{D}{min\,(W,H)} + 1 \right) \tag{7.8}$$

Accot and Zhai [2] later identified problems with these formulations: Equation 7.7 ignores height (shown to have an effect by Sheikh and Hoffmann [109]), while Equation 7.8 considers only one dimension and ignores the angle of approach. Accot and Zhai proposed a weighted Euclidean model (Equation 7.9),

which addressed the dimension issue.

$$ID_{WtEuc} = log_2\left(\sqrt{\left(\frac{D}{W}\right)^2 + \eta\left(\frac{D}{H}\right)^2} + 1\right) \tag{7.9}$$

Accot and Zhai's model is similar to the Euclidean norm, with the addition of the parameter η, which weights the effect of height differently from the effect of width. However, Accot and Zhai's formulation does not account for the angle of the target from the cursor and is constrained to rectangular targets. Therefore, Grossman and Balakrishnan [48] proposed a probabilistic model that is generalized to any target shape, size, orientation, location and dimension.

We adopted Accot and Zhai's weighted Euclidean model (shown in Equation 7.9) [1]. However, as we mentioned earlier this model does not account for possible differences in performance due to varying movement angles. Therefore, we applied an approach proposed by Grossman and Balakrishnan [48] for pointing targets in 3D. They accommodated angles by adding an additional empirically determined weight parameter $f_{W,H,D}(\Theta)$ (W :width, H :height and D :depth) for each component in the weighted Euclidean model. We applied their model by removing the third dimension constraint (depth). Hence our modified weighted Euclidean model becomes:

$$ID_{P2} = log_2\left(\sqrt{f_W(\Theta)\left(\frac{D}{W}\right)^2 + f_H(\Theta)\left(\frac{D}{H}\right)^2} + 1\right) \tag{7.10}$$

For the sake of completeness, this model is compared with $ID_{W'}$ (Equation 7.7) proposed by MacKenzie and Buxton [76] after extending it to accommodate the possible effects of different target dimensions and various movement angles. From this perspective, the model becomes:

$$ID_{WtW'\Theta} = log_2\left(f_{W'}(\Theta)\frac{D}{W'} + 1\right) \tag{7.11}$$

Next, we need to include the target speed to the ID_{P2} model in Equation 7.10. Therefore, we combined the model we discussed for moving targets in 1D ID_{C1} and the model for stationary targets in 2D ID_{P2}. We revised our ID_{C1} model and

broke the target velocity V into x and y components as follows:

$$V_x = V cos\left(\Theta\right), V_y = V sin\left(\Theta\right)$$

The resulting index of difficulty ID_{C2} incorporating target speed is shown in Equation 7.12. Incorporating target velocity V in a similar manner into the $ID_{WtW'\Theta}$ model results in Equation 7.13.

$$ID_{C2} = log_2\left(\sqrt{f_W\left(\Theta\right)\left(\frac{D \pm \frac{V_x}{K}}{\frac{W}{2} - \frac{V_x}{K}}\right)^2 + f_H\left(\Theta\right)\left(\frac{D \pm \frac{V_y}{K}}{\frac{H}{2} - \frac{V_y}{K}}\right)^2} + 1\right) \tag{7.12}$$

$$ID_{VWtW'\Theta} = log_2\left(f_{W'}\left(\Theta\right)\frac{D \pm \frac{V}{K}}{\frac{W'}{2} - \frac{V}{K}}\right) \tag{7.13}$$

The ID_{C2} model looks complicated as it is presented in the current formulas; however, by using vector notation, we could rewrite it in a much simpler way as shown in Equation 7.14 and it could even be updated to 3D with an update to the assumptions.

$$\vec{V} = \begin{bmatrix} \frac{V_x}{k} \\ \\ \frac{V_y}{k} \end{bmatrix}, \vec{R} = \begin{bmatrix} \frac{W}{2} \\ \\ \frac{H}{2} \end{bmatrix}, \vec{D} = \begin{bmatrix} D_x \\ \\ D_y \end{bmatrix}, and \vec{F} = \begin{bmatrix} \sqrt{f_w\left(\theta\right)} \\ \\ \sqrt{f_H\left(\theta\right)} \end{bmatrix}$$

By taking $\vec{V}$ as the velocity vector, $\vec{R}$ as the object vector, $\vec{D}$ as the distance vector between the cursor and the target and $\vec{F}$ as the weighted vector, the extended model will simply be

$$ID_{C2} = log_2\left(\left|\vec{F} \cdot \frac{\vec{D} + \vec{V}}{\vec{R} - \vec{V}}\right| + 1\right) \tag{7.14}$$

From the $ID_{VWtW'\Theta}$ model (Equation 7.13) we can see that the index of difficulty increases as the speed increases or the size decreases while keeping other factors constant. The model also predicts that targets moving towards the cursor (i.e., chasing behaviour) have a larger index of difficulty than those moving away (i.e., pursuit behaviour). To validate these models, we conducted a user study as described later in Section 7.3.

7.2 Moving Target Selection Technique

We created a selection technique called *Hold* to overcome some of the drawbacks of previous methods and to reduce the difficulty of moving target selection. Our approach removes speed as a contributing factor to the index of difficulty of selecting moving targets thus reducing the task to a simple 2D static selection task.

Hold works as follows: when a user clicks the mouse button down, the moving targets temporarily pause while the user interacts with targets. When they release the button, the target starts moving again. The active engagement of the motor system allows users to be aware of the temporary nature of the paused state, which should reduce confusion [18, 76]. This approach removes the factor of target speed from the task of selecting a moving target on the basis of its distance from a pointer and its relative size. This is in contrast to the traditional chasing technique that involves moving a cursor over a moving target and accurately selecting it before it moves out of the way (we refer to this technique as *Chase*). Pausing the interface theoretically removes the target speed, reducing the movement time of the pointer because the pointer speed no longer needs to be coupled to the target speed. For the same reason, we expect the error rate to be reduced. The main adjustment for users is that selection is done with a mouse-up event after mouse movement, which is different from the usual mental model for selection. Our experiment, described next in Section 7.3, tests this hypothesis and compares our *Hold* technique with the usual *Chase* technique as well as a hybrid of the two to see if users can seamlessly and effectively use a combination of both.

7.3 Empirical Validation of Moving Target Models

For our model validation, we ran a controlled experiment following the standard protocol used in Fitts' experiment (discrete point-select) for both 1D and 2D where participants are required to go to a designated start location and then acquire the target with a controlled set of independent variables (selection type, distance, size, velocity, angle and direction). The test environment structured as a game developed in Flash CS4 called "Catch the Wisp" (based on a previous game created by Ilich[62]) to test the three approaches: *Chase*, *Hold*, and *Hybrid*. In this game the three interaction techniques were abstracted as game objects that we developed

as *potions*. The game's mental model was designed through an iterative design strategy to avoid confusion for participants and establishes a simple analogy that minimizes the training and confounding explanation that you might see in traditional Fitts experiments directly applied to moving targets. The three techniques are:

Chase: The user pursues and selects a moving target by predicting its movement and clicking the left mouse button when the cursor is over it. In state space shown in Figure 7.2, State 0 represents a Button-Up state with no initial target selection, while State 1 represents a Button-Down state where an object has been selected.

Hold: The user is able to freeze all targets' motion by holding the left mouse button down. With the mouse button depressed, the user can move the cursor over the paused object and release the button to select it. Releasing the button while the cursor is not over a valid target will resume the motion of all targets in view. In state space as shown in Figure 7.2, State 0 represents a Button-Up state with no initial target selection and the target is in motion; State 1 represents a Button-Down state where the target has been frozen and a subsequent Button-Up event with the cursor over the target will result in target selection.

Hybrid: This model is a hybrid of the previous two models in which the user is free to chase the target and reduce the initial amplitude of movement, until they decide to click, hold the mouse button down and freeze the target for a final precision or corrective phase of movement. In this model, the target can be acquired by either a Button-Up or Button-Down event, provided the cursor is directly over the target. In state space as illustrated in Figure 7.2, State 0 represents a Button-Up state with no initial target selection and the target is in motion; State 1 represents a Button-Down state where the target has been frozen and selected if the cursor was over the target. A subsequent Button-Up event with the cursor over the target will also result in target selection.

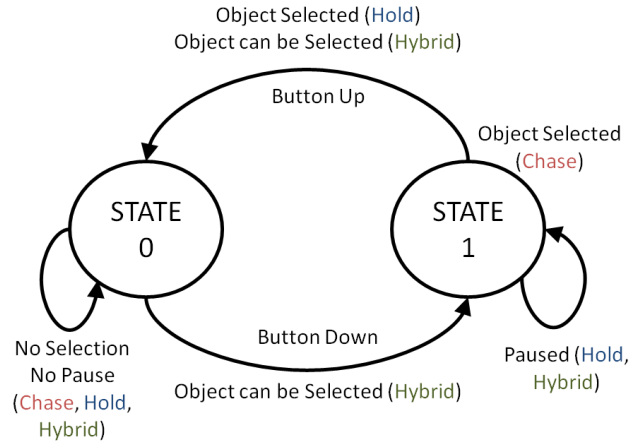


Figure 7.2: State Transition Diagram for methods of interaction.

The user study consisted of two phases in which we first compare the performance of *Chase* and *Hold*, and then observe user behavior in the *Hybrid* model (*Chase-or-Hold*) to see if users can effectively use both techniques together seamlessly.

The challenge in creating a test environment for moving targets is that we needed to ensure that we could reliably have the moving targets in a predetermined position across conditions between participants. Thus, we used a game called "Catch the Wisp" where the target object was abstracted as a *wisp*, or a ball of light from folklore, that would start at a fixed distance from the cursor. In order to make sure that the cursor is at a constant distance from the target, users would have to roll the mouse over a target start location, called a *potion*, in the game metaphor, which is located at a predefined location from the target. Users were presented with three different colours for the start location: red, blue and green representing the three different techniques *Chase*, *Hold* and *Hybrid*, respectively. The behaviour of each *potion* was framed as part of the game such as paralysis and removing a wisp's shield to provide a mental model consistent with the technique being tested. Using this game, we could observe the acquisition of a variety of targets while the users remained engaged and following instructions without making mode errors with respect to technique.

7.3.1 Apparatus

The user study was conducted on a Toshiba laptop with a 2.10GHz Core2 Duo CPU with 2GB of RAM running Windows XP Professional. For the purposes of this experiment, "Enhance pointer precision" was disabled and the pointer speed was set to 6 (out of 10). The laptop LCD display was used at a resolution of 1280×800 pixels at a refresh rate of 60Hz. A Microsoft optical wheel mouse was used as the input pointing device and the Adobe Air environment was set to 1024×640 pixels while running the Flash program.

7.3.2 Participants

Fifteen paid volunteers, eleven female and four male, participated in the experiment. Participants ranged in ages from 19 to 40, were all experienced computer users and have either normal or corrected to normal vision. None of the participants had colour blindness. All participants controlled the mouse in the experiment with their right hand. Participants report playing computer games rarely or never. Participants' details are shown in Table 7.1.

Table 7.1: Demographics summary for participants in the investigation of the target selection technique study. (Note: D: dimension)

P	Gender	Age Group	Dominant Hand	Game Play Frequency	Vision	Experiment Type
1	Female	18-21	Right	Rarely	Corrected	1D
2	Female	22-25	Right	Rarely	Normal	1D
3	Female	22-25	Right	Rarely	Corrected	1D
4	Female	18-21	Right	Rarely	Corrected	1D
5	Male	18-21	Right	Rarely	Normal	1D
6	Female	31-40	Right	Rarely	Corrected	1D
7	Female	18-21	Right	Never	Normal	2D
8	Male	31-40	Right	Rarely	Corrected	2D
9	Female	18-21	Right	Never	Corrected	2D
10	Male	26-30	Right	Rarely	Normal	2D
11	Female	22-25	Left	Never	Corrected	2D
12	Female	26-30	Left	Rarely	Corrected	2D
13	Female	18-21	Right	Never	Corrected	2D
14	Female	22-25	Right	Never	Corrected	2D
15	Male	26-30	Right	Rarely	Normal	2D

7.3.3 Procedure

Participants played the "Catch the Wisp" game and they were asked to capture a moving wisp using the mouse functions according to the test conditions based on the start location (i.e. potion) colour. The destination target (i.e. wisp) is presented as a white circle protected by a shield, which is disabled differently by each technique. For the red potion, simply rolling the mouse over the potion will disable the shield allowing users to capture the wisp by clicking on it: the *Chase* mode. For *Hold*, we use a blue potion. When rolling over the blue potion, a blue web (24×20 pixels) appears at the potion location and by depressing the left button of the mouse over the blue web the wisp's shield will be removed and the motion will be frozen as if paralyzed. Keeping the mouse button down, users can then drag a thread from the web to the wisp and release the button over the wisp to catch it. With the green potion, the shield is removed when they roll over it. At that point, users could decide to pursue the wisp and click on it to catch it or if they prefer, they can hold the left button down, freezing the target and displaying a diagonal cross hair where the cursor is. Then, they can drag the diagonal cross hair over the wisp and release the button to catch it. These three methods are shown in Figure 7.3 depicting the start mode (left panel) and the selection mode (right panel).

The game was structured as a series of trials. In each trial participants were asked to select one moving target, which starts moving from a predefined location, which is at a constant distance from a start location. Rolling over the start location activates the corresponding technique causing the start target (i.e. potion) to disappear and the target (i.e. wisp) to start moving. Participants started by completing a tutorial of 24 trials that included on-screen cues showing the next action they have to perform, such as "Move the mouse over the potion" or "Click on the web to pause the wisp". After completing the tutorial, participants started the game, which consists of sixteen sets (8 red and 8 blue). Each set contains 36 trials, so in total the game has 576 trials. The sixteen sets were organized in alternation of *Chase* (Red) and *Hold* (Blue) sets where one type of condition is presented in each set. Upon completion of the *Chase* and *Hold* sets, participants would test the *Hybrid* (Green) condition. They did twelve practice trials with the *Hybrid* condition. For the experiment, participants completed 72 trials using *Hybrid* condition

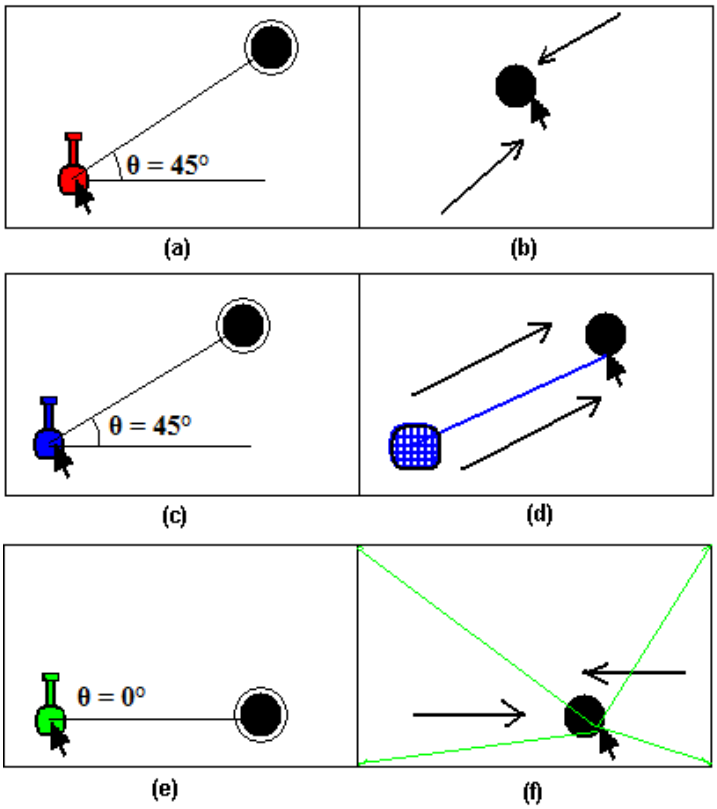


Figure 7.3: Experiment Acquisition Types, Red Potion/*Chase* (a,b), Blue Potion/*Hold* (c,d) and Green Potion/*Hybrid* (e,f)

that was specified as a bonus set to keep in line with the game metaphor for the experimental set up. For each trial, participants had a maximum of five attempts to select the target after which the trial was ended and they advanced to the next trial. At the end of each set, a message was displayed informing the participants of the number of targets they caught in that set, the total credit they achieved up to that point and their best acquisition time. Participants were allowed to take breaks after the completion of each set. They were instructed to be as accurate and as fast as possible.

Upon the completion of the second phase, participants were asked to fill out a questionnaire, shown in Section B.6, to give their comments and suggestions on each method. The experiment lasted approximately two hours per participant.

7.3.4 Design

The user study was run as two distinct experiments, divided between targets moving in 1D and in 2D, since having one subject do both conditions would have been too long. To quantitatively facilitate a comparison of performance between *Chase* trials and *Hold* trials a repeated measures within-participant factorial design was used. Selection method determined by the potion colour was the independent variable. In order to determine the relation between the task difficulty and acquisition time, we used four other independent factors. These factors are:

Size (W): The size of the wisp in each trial, as measured by the radius of its circle in pixels. The three levels used were: 10 pixels, 18 pixels or 30 pixels.

Speed (V): The constant speed at which each wisp traveled. Three levels were used: 100 pixels/sec, 175 pixels/sec or 250 pixels/sec.

Direction: The direction of the wisp from the potion. We used two levels: toward or away from the potion.

Movement Angle (θ): The angle between the wisp and the cursor. The angles are: 0 and 180 degrees in 1D and 45, 135, 225 and 315 degrees in 2D relative to a horizontal axis through the center of the potion

These factors were fully counter-balanced between trials. Within each selection method, participants in the 2D experiment tried 72 ($3W \times 3V \times 2Direction \times 4\theta$) conditions in which they carried out 4 targeting trials for each condition. Participants in the 1D experiment tried 36 ($3W \times 3V \times 2Direction \times 2\theta$) conditions in which each condition was carried out 8 times. Therefore, for each dimension, participants were presented with 576 trials distributed evenly between two selection techniques, including 288 *Chase* and 288 *Hold*. In order to compare the results with the traditional Fitts tasks, different sizes were used while keeping the distance to the target constant. Varied speeds of movement were used to determine their effects on the index of difficulty of moving targets. Finally, we considered that objects moving away from the cursor may be easier (or harder) to select than objects moving towards the cursor so had trials for each to test this.

For the second phase of the experiment where participants used the *Hybrid* method (green potion), a qualitative measure was used to observe their acquisition behavior. In this phase, participants had the option to pursue the target, freeze the target or a combination of both in order to select the moving target. In this phase, we were looking to see whether participants optimize the two since allowing a target that is moving toward you while you move towards it may achieve faster target acquisition times versus pausing it. However, it may come at the price of accuracy as well as poorer time performance if you miss out on the first rendezvous with the target.

7.3.5 Performance Measures

For this study we used the acquisition time and the number of errors as our dependent variables. The acquisition time MT was measured from the time the start location (i.e. potion) was activated, when the mouse rolled over it, until the time the moving target was captured. The number of clicks or mouse up events that did not result in a moving target's capture were counted as errors. Moreover, the position of the cursor and the moving target were recorded in every frame. Every time a participant froze the target was also recorded together with the total acquisition time in order to study participant behavior.

For the *Hybrid* set, a qualitative analysis consisted of a statistical measure of the distribution of participant behaviors among four categories:

Chase: A participant chose to pursue the target without freezing it.

Hold: A participant chose to freeze the target immediately after the activation of the start location.

Hybrid: A participant chose to start pursuing the target and then freeze the target closer to it.

Error Correction: A participant missed the target and tried to correct their miss.

We categorized each trial into one of the above four categories according to:

- If the trial had an error then it is categorized as an **Error correction**.

- If a participant did not *Hold* during the trial then a ***Chase*** behaviour was selected.
- If a participant did *Hold* then we checked the cursor position when the last *Hold* event occurred with the initial cursor and wisp position.
  - If the distance moved was less than the remaining distance to the wisp then it was considered to be a ***Hold*** behavior.
  - Else it was a ***Hybrid*** behavior.

7.3.6 Results

Both a repeated measures ANOVA analysis and a Generalized Linear Mixed Models (GLMMs) Test were carried out to test the significance of the results. We illustrate below the GLMMs Test results since our data does not have a normal distribution where it is positively skewed. We get almost similar significant effects in both. Selection technique, size, speed, angle and direction were taken as fixed factors in the GLMMs Test while subject and trials were taken as random factors. The outliers were removed based on the acquisition time and number of errors, such that any data point with extreme acquisition time where its frequency dropped to zero, or with 5 errors (since participants were allowed to have only 5 attempts in each trial) were not included. In total, 0.45% of the 1D data and 0.90% of the 2D data were removed as outliers. Six people participated in the 1D selection task experiment, and nine people participated in the 2D selection task experiment as shown in Table 7.1.

1D selection task

Phase 1: Our data was positively skewed so we used a Generalized Linear Mixed Models (GLMM) test to analyze the acquisition time. The analysis showed that the independent variables selection technique ($F(1,3099.89) = 260.26$, $p < 0.001$), size ($F(2,3099.38) = 301.26$, $p < 0.001$), speed ($F(2,3099.38) = 5.02$, $p = 0.007$), direction ($F(1,3099.27) = 6.08$, $p = 0.014$), and angle ($F(1,3099.37) = 101.36$, $p < 0.001$) had a significant effect on the acquisition time. Moving to the left ($\theta =$

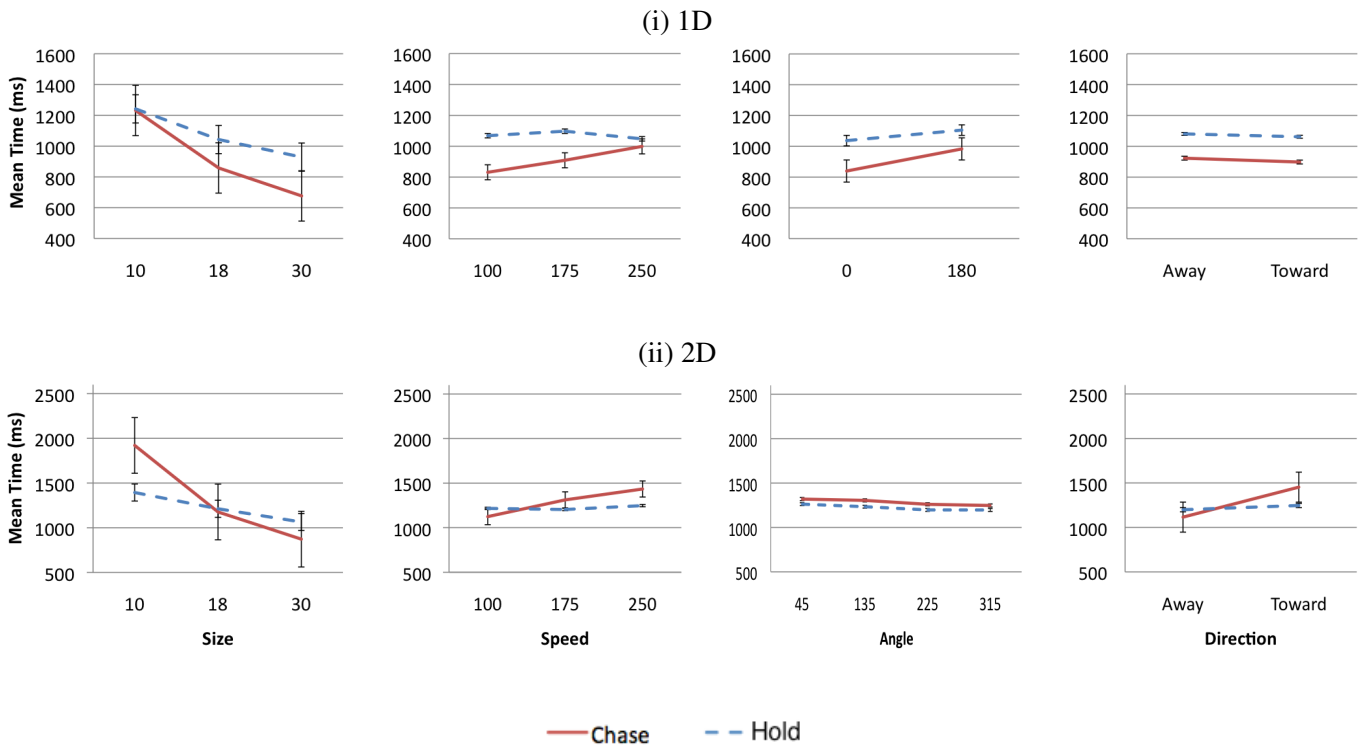


Figure 7.4: The effect of size, speed, angle and direction on mean acquisition time for *Hold* and *Chase* in (i) 1D and (ii) 2D.

180) was found faster than moving to the right (θ = 0) for both techniques. Post-hoc comparison showed that the mean acquisition time for *Hold* (M = 981.838, SD = 69.015) was longer than *Chase* (M = 835.525, SD = 63.931). This coincides with participants' feedback that the blue potion gives them the impression of having enough time to select the target and they do not have to rush, as in red potion (*Chase*) trials. Hence, it slowed their reaction and took longer to get started after rolling over the potion. Also technique $\times$ size $\times$ speed $\times$ angle interaction showed significant effect ($F(4, 3099.31)$ = 3.94, p = 0.003) on time. The overall mean acquisition times for techniques by size, speed, angle and direction are illustrated in Figure 7.4(i).

Errors were also analyzed using the same GLMMs test as acquisition time. A significant effect was observed on the number of errors for the selection technique ($F(1, 3105.66)$ = 380.17, $p < 0.0001$), size ($F(2, 3101.72)$ = 17.35, $p < 0.0001$), speed ($F(2, 3101.69)$ = 3.33, p = 0.036) and angle ($F(1, 3101.61)$ = 23.34, $p <$ 0.001). Direction ($F(1, 3100.85)$ = 0.182, p = 0.67) had no main effect on the number of errors. *Chase* contained more errors (M = 0.073, SD = 0.76) compared with *Hold* (M = 0.015, SD = 0.36), which illustrated the speed accuracy trade off in the two selection techniques consistent with participants feeling they had to rush in the *Chase* mode.

Phase 2: The 72 trials of the green potion were compared with the last 72 trials from each technique of the first phase of the experiment using the GLMMs Test. A significant effect was observed on acquisition time ($F(2, 272.09)$ = 10.49, $p <$ 0.001). *Hybrid* (M = 842.68, SD = 420.49) was faster than *Hold* (M = 1056.67, SD = 475.18) and *Chase* (M = 960.67, SD = 859.28) (time is measured in milliseconds). Moreover, a significant effect was observed on the number of errors ($F(2, 267.21)$ = 35.15, $p < 0.001$). *Hold* (M = 0.017, SD = 0.397) had fewer errors than either *Chase* (M = 0.089, SD = 0.827) or *Hybrid* (M = 0.082, SD = 0.995).

2D selection task

Phase 1: For the 2D selection task experiment, a GLMMs was conducted to compare the total acquisition time for each technique. There was a significant effect

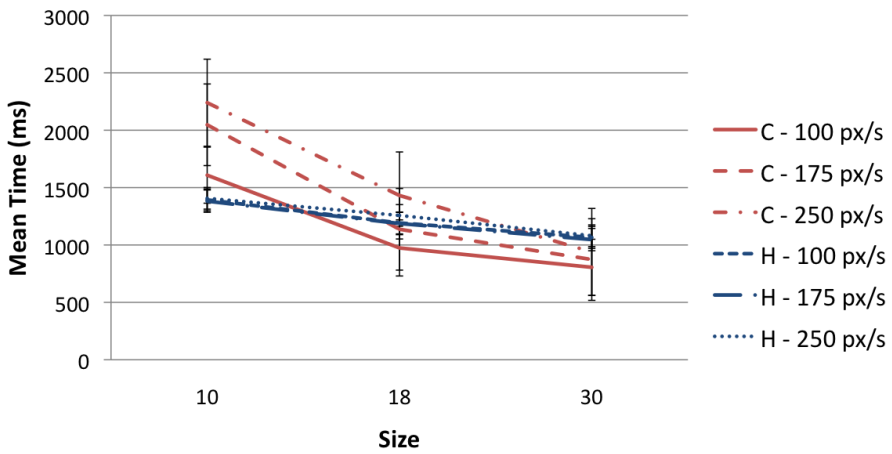


Figure 7.5: The 2D mean time by size and speed for C: *Chase* and H: *Hold*.

($F(1,4807.93) = 64.38$, $p < 0.001$) when an overall comparison was made and the mean acquisition time for *Hold* ($M = 1239.874$, $SD = 67.079$) was shorter than *Chase* ($M = 1450.053$, $SD = 127.114$). The other independent variables size ($F(2,4807.58) = 668.66$, $p < 0.001$), speed ($F(2,4807.30) = 65.50$, $p < 0.001$), direction ($F(1,4807.14) = 156.26$, $p < 0.001$), and angle ($F(3,4807.16) = 4.66$, $p = 0.003$) showed significant effect on acquisition time. Pairwise contrast between angles showed a significant difference between angle pair 45 and 315 degrees and this coincides with results found in [116]. This effect is most likely due to the fact that targets on either end of a vector are equally difficult to select and movement direction had no effect on acquisition time as illustrated in [48]. The effect of the angle on acquisition time was dependent on both size and direction indicated by the interaction size $\times$ direction $\times$ angle ($F(6,4807.15) = 2.24$, $p = 0.037$), and also mode, size, speed and direction indicated by mode $\times$ size $\times$ speed $\times$ direction $\times$ angle ($F(12,4807.14) = 2.29$, $p = 0.007$). The mean acquisition times for technique by size, speed, angle and direction are illustrated in Figure 7.4(ii). In addition, significant effects were observed for size and speed combinations; however, *Hold* exhibited a lower mean acquisition time for the small target as well as the fast moving targets as shown in Figure 7.5.

A significant effect of technique was also observed on the number of errors

($F(1,4812.41)$ = 754.92, $p < 0.001$), *Hold* contained fewer errors (M = 0.009, SD = 0.014), compared with *Chase* (M = 0.083, SD = 0.098). The main effect of size ($F(2,4810.50)$ = 277.73, $p < 0.001$), speed ($F(2,4808.88)$ = 56.68, $p < 0.001$) and direction ($F(1,4808.01)$ = 5.56, p = 0.018) were also found to be significant. However, angle showed no significant main effect on errors.

Phase 2: A GLMMs test was also conducted to analyze the 72 trials of the 2D *Hybrid* method. A significant effect was observed on acquisition time ($F(2,291.23)$ = 8.98, $p < 0.001$). *Hybrid* (M = 1162.07, SD = 635.63) was faster than *Hold* (M = 1198.16, SD = 499.15) and *Chase* (M = 1333.09, SD = 1343.40). Moreover, a significant effect was observed on the number of errors ($F(2,97.46)$ = 55.82, $p <$ 0.001). *Hold* (M = 0.01, SD = 0.306) had fewer errors than either *Chase* (M = 0.096, SD = 0.96) or *Hybrid* (M = 0.054, SD = 0.76).

Model Fitting

By a least-squares fit method, we estimated the coefficients of the models described earlier for *Chase* and *Hold*. We adopted the original Fitts' Law [39] in Equation 7.1 for trials involving *Hold* in 1D while we tested Equation 7.10 and Equation 7.11 for *Hold* trials in 2D. For trials involving *Chase* in 1D, we adopted Equation 7.6 while Equation 7.12 and Equation 7.13 were adopted for trials in 2D. For 1D targets, the Fitts' model (R^2 = 0.9869) and ID_{C1} (moving away R^2 = 0.9662, moving toward R^2 = 0.9755) have shown very good fits with the experimental data. The ID_{P2} (R^2 = 0.9717) and $ID_{W_tW'\Theta}$ (R^2 = 0.9505) models for stationary targets in 2D have also shown an excellent fit with the experimental data of *Hold*.

For 2D moving targets models, the $ID_{VW_tW'\Theta}$ (R^2 = 0.9099) model showed a very good correlation with the data while the ID_{C2} model exhibited poor correlation with some angles. The poor correlation can be explained as discussed earlier that angle pairs (45, 225) and (135, 315) lay in the same diagonal vector. As in previous research, each pair was considered as one movement angle and this was shown in [116] where a significant difference existed only between angles 45 and 315. When each pair was considered as one movement angle, the correlation improved (R^2 = 0.9027) as shown in Table 7.2 and it exhibited similar performance as

Table 7.2: Summary of coefficients estimates, corresponding standard errors, and R^2 values for the regression of the 2D moving targets models, where a and b correspond to coefficients of Fitts' Law in (Equation 7.2)

ID model	Parameter Estimates (with std error indicated below)							
	Direction	Θ	$a(ms)$	$b(ms/bit)$	$k(/ms)$	$f_w(\theta)$	$f_h(\theta)$	R^2
$ID_{VWtW'\Theta}$	Away	45	87.354 49.60	370.06 15.61	28.23 0.35	0.24 0.073		0.9381
		135	282.96 112.63	243.4 27.39	31.62 0.51	0.23 0.032		0.9013
		225	139.56 64.47	361.94 33.93	29.76 1.40	0.21 0.059		0.9107
		315	127.62 46.72	351.85 24.88	38.76 5.62	0.23 0.073		0.9362
	Toward	45	-236.29 57.39	419.41 13.81	32.43 3.12	0.47 0.032		0.8784
		135	-150.85 47.97	414.87 22.36	30.03 0.39	0.49 0.012		0.9094
		225	-285.31 49.69	448.01 69.19	27.97 0.28	0.42 0.0656		0.8941
		315	16.175 65.37	380.01 32.70	27.01 0.086	0.37 0.024		0.9246
ID_{C2}	Away	45	-1277.5 225.73	487.46 63.62	25.01 4.90	0.73 0.65	0.73 0.20	0.9279
		135	-1129.3 267.46	474.05 49.12	25.03 2.95	0.67 0.80	0.67 0.60	0.9182
		225	-1220.2 480.87	480.65 96.72	29.32 3.06	0.52 0.61	0.78 0.15	0.894
		315	-929.04 310.55	414.91 66.15	32.54 1.60	0.53 0.49	0.95 0.38	0.9458
	Toward	45	-1071.4 254.11	468.54 57.25	25.01 3.72	1.14 0.45	1.21 0.15	0.8891
		135	-1478.7 254.13	544.83 41.98	25.01 2.08	1.85 0.16	1.78 0.22	0.9116
		225	-1869.5 469.17	612 89.91	25.01 3.31	1.21 0.79	1.02 0.68	0.9124
		315	-1694.8 546.92	570.69 106.23	25.001 1.33	0.67 0.70	2.28 0.47	0.8911

the $ID_{VWtW'\Theta}$. Another factor could be the target dimension, the ID_{C2} model gives different weights for width and height of the target; however, we used a circular target in our experiment where both height and width are equal. Figure 7.6 illustrates the average acquisition time versus the index of difficulty for both models.

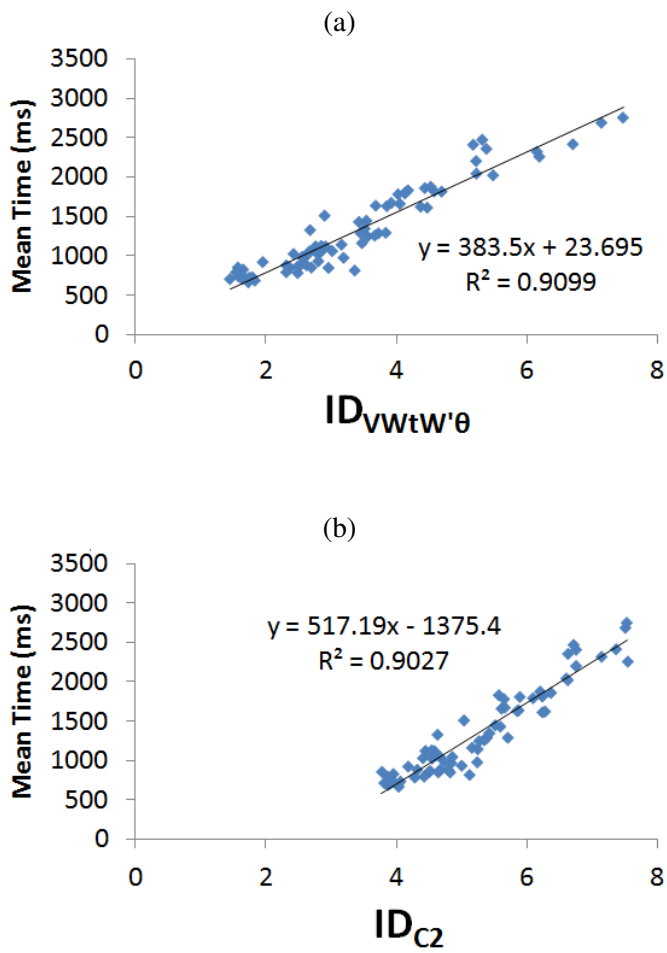


Figure 7.6: Mean Acquisition Time Vs. Index of difficulty for *Chase* in 2D using (a) $ID_{VWtW'\Theta}$, and (b) ID_{C2} models.

7.3.7 Discussion

For 1D we observed from the first experiment that *Chase* exhibited lower acquisition time than *Hold* in all conditions. This trend can be interpreted in different ways. The overhead of clicking on the web and sustaining the dragging motion outweighed the benefit of freezing the target. In addition, as the target is frozen it gives an illusion of the time also being paused. In this state, participants did not rush in selecting the target resulting in taking longer. Adding a score in the game

based on time taken, target size and speed does not help some users in mitigating this effect. We think that adding a time limit to capture a target, as some participants suggested, would help. Questionnaire results agreed with these findings. Participants preferred *Chase*, as they felt it was faster and they did not account for the accuracy. Also participants tended to take longer to precisely release the button over the target in *Hold*, which resulted in fewer errors but longer acquisition time.

However, for 2D, *Hold* showed faster acquisition time for conditions involving a target size of 20px (small) as well as conditions involving a target speed of 175px/s (moderately fast) and 250px/s (fast). This contradicts the results for the selection task in 1D. We believe that this can be explained by the distance the target had to travel was restricted to a horizontal path in 1D. Because of this restriction in 1D, the target was more likely to rebound off the end and approach the cursor while in 2D, the target moves at an angle and thus would take longer time to hit a wall and rebound towards the cursor again for users to take advantage. We confirmed that speed has little impact on *Hold* as observed in Figure 7.4. In both 1D and 2D, participants sacrificed accuracy for speed in *Chase* by attempting to click on the target in rapid succession, while they sacrificed speed for accuracy in *Hold* by carefully positioning the cursor over the target before releasing the mouse button.

The results of the second phase of the experiment in 1D showed that *Hybrid* resulted in reduction in acquisition time of 12% over *Chase* and 20% over *Hold*. While in 2D, it showed a reduction of 13% over the *Chase* and 3% over the *Hold* suggesting participants are performing optimization seamlessly. The cursor and target position logs for each participant were analyzed to categorize the technique used for each trial. The results are summarized by the percentage of trials for which each technique was chosen by target size, speed, angle, and direction in Figure 7.7(i) and Figure 7.7(ii). In 2D, participants tended to use a *Hybrid* approach more often while in 1D, participants tended to use *Chase* as they thought it was faster. Participants in the 1D experiment commented that they had used *Chase* more often in the second phase due to the unfamiliarity with *Hybrid* and they claimed that *Chase* is faster, which would optimize their acquisition time.

The distribution of the ratios (*Chase* / (*Chase* + *Hold*)) and (*Hold* / (*Chase* + *Hold*)) for second phase of the experiment was also analyzed and it is summarized by target size, speed, angle, and direction in Figure 7.8. Participants tended to

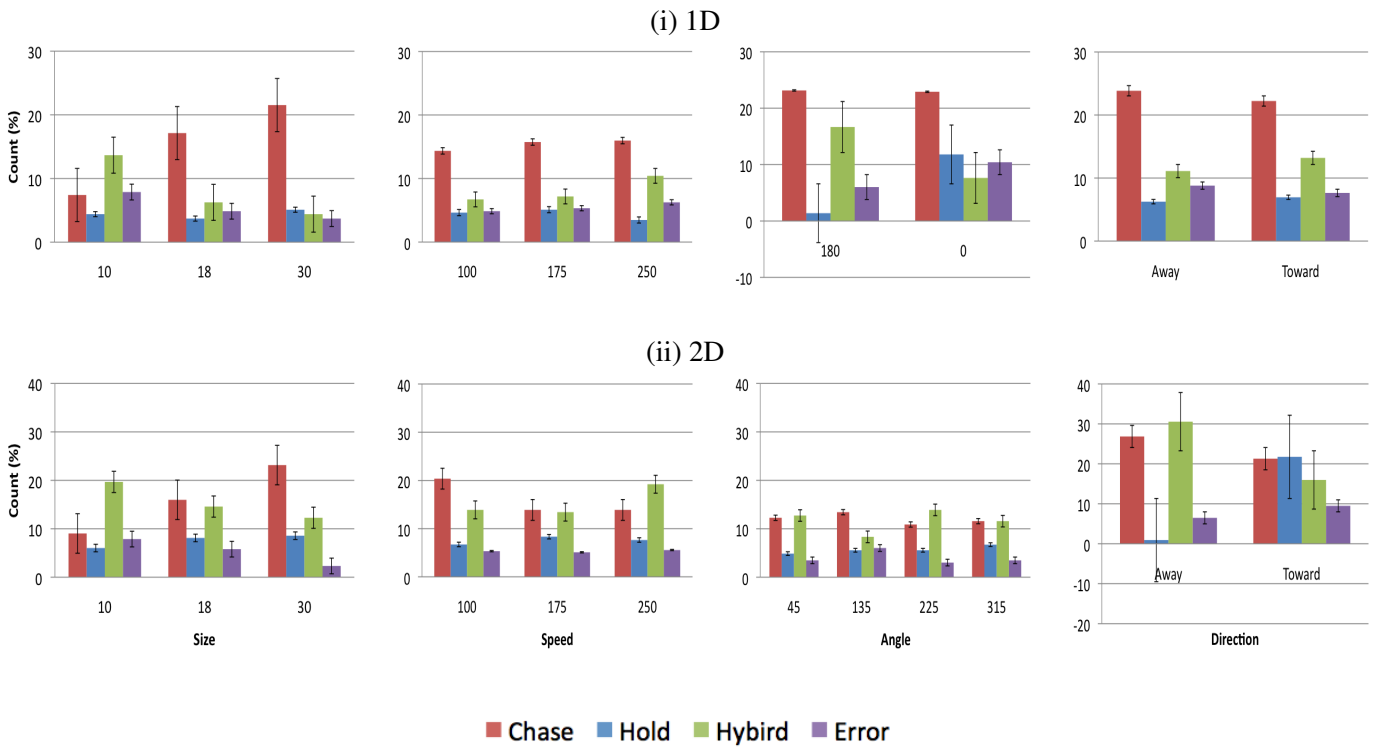


Figure 7.7: Technique chosen by size, speed, angle, and direction in (i) 1D and (ii) 2D.

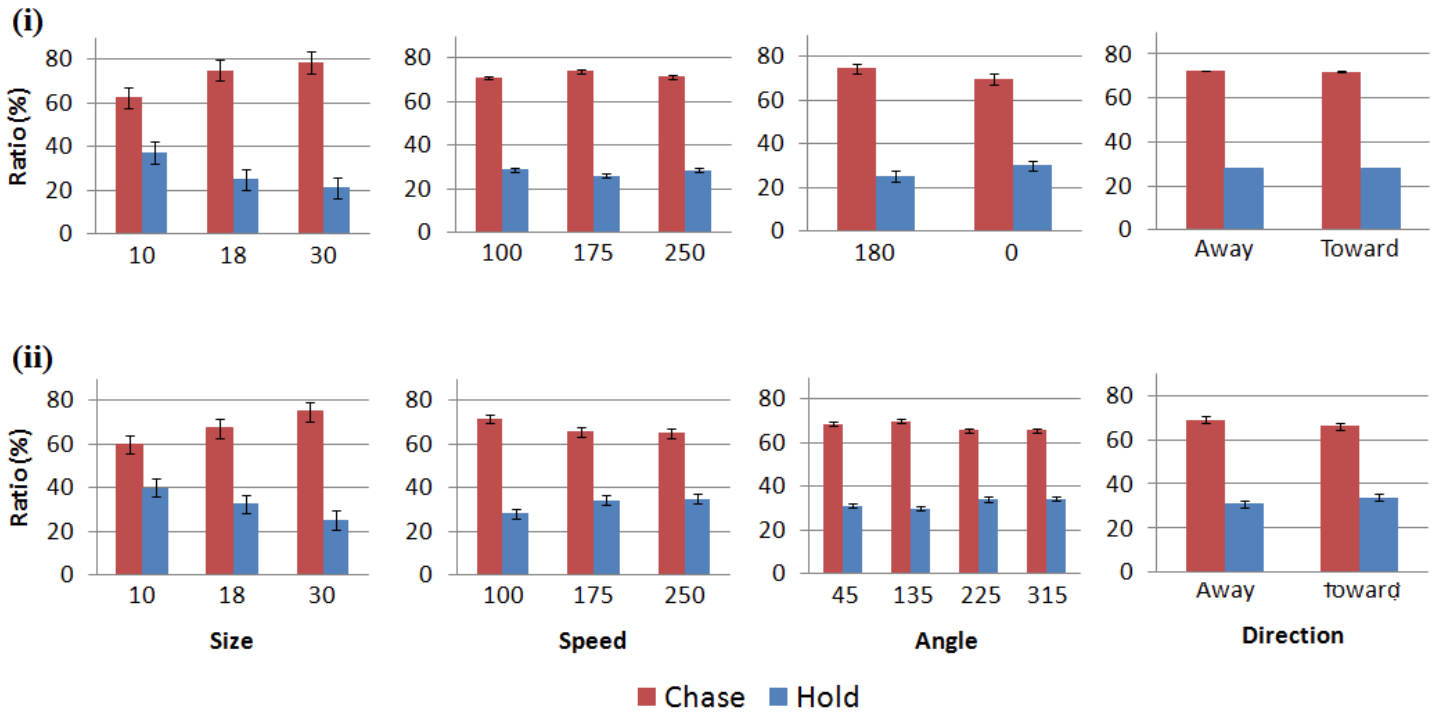


Figure 7.8: The ratio distribution of *Hold* and *Chase* using *Hybrid* in (i) 1D and (ii) 2D space

chase the target most of the time across the different conditions. However they froze the target more often as the target got smaller and faster, which indicates a seamless and effective use of techniques.

7.4 Hold in a Rich Media Interface

Based on the promising results, a rich media interface called *MediaDiver* was developed, which employs the *Hold* technique for viewing and annotating multiple-views video. *MediaDiver* allows users to experience novel moving target selection methods (*Hold* and *Chase*), new multi-view selection techniques, automated quality of view analysis to switch viewpoints to follow targets, integrated annotation methods for viewing or authoring meta-content and advanced context sensitive transport and timeline functions. As users have become increasingly sophisticated when managing navigation and viewing of hyper-documents, they transfer their expectations to new media. *MediaDiver* is a demonstration of the technology required to meet these expectations for video. Thus, users can directly click on objects in the video to link to more information or other video, easily change camera views and mark-up the video with their own content. The applications of this technology stretch from home video management to broadcast quality media production, which may be consumed on both desktop and mobile platforms.

7.4.1 MediaDiver Interface

MediaDiver is an interactive interface for experiencing, viewing and annotating complex video domains and its associated metadata content suitable for computers and mobile devices. Rather than playing videos linearly from start to finish without any viewer's interaction, *MediaDiver* extends the role of viewing to interaction with multi-view video and context-connected video-space, which allows customized viewing based on a user's preferences. It enables the viewer to interact, browse, explore and annotate (i.e. tag) the multi-stream context videos.

The central component in the main window of the *MediaDiver*, shown in Figure 7.9, is the player, which shows the main video content and allows users to interact with the video's objects. A timeline at the bottom of the player shows the current time offset from the time base of the video source. Users can control the

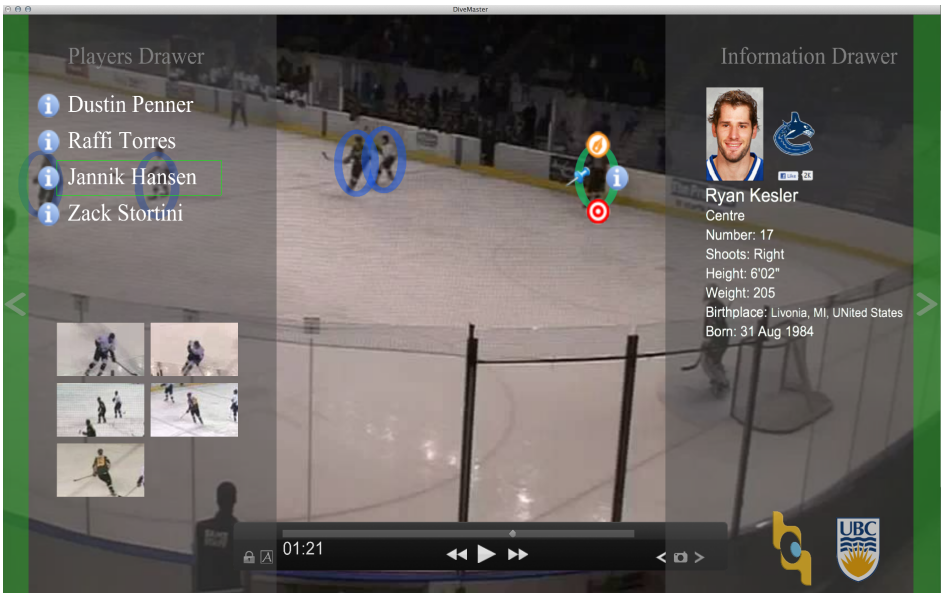


Figure 7.9: Overview of *MediaDiver* in view mode. Select-able players are highlighted when rolling over the player showing the ability to select them. Selection enables highlighting, information retrieval, pinning and tagging. Pinned players are shown to the left in the players' drawer while player's information can be retrieved and shown on the right.

playback of the video using the Play button to start playing the video or pausing it when it is playing. One of the major challenges of interaction with rich media is the selection of a moving target. Using our novel method *Hold*, a target may be selected by holding down the mouse button to pause the video, moving to the target and releasing to select. Selection serves as a context switch, which enables functions such as highlighting, information retrieval, pinning and tagging. Select-able objects are highlighted when rolling over the player, showing the ability to select them.

Our interface provides several additional options after selection. With the mouse button held down over a target or rolled over the target, a pie menu appears with options for following the target, information retrieval, pinning and annotation. The follow button (◎) function is to keep the selected player within the current view so that when a player leaves the current field of view the interface

Figure 7.10: Overview of *MediaDiver* in annotation mode. Each rectangular border represents an annotated object. Users can edit each annotation by moving or deleting it. They can also add new annotation by clicking on the object they want to annotate.

switches to a view in which the player is visible. The information retrieval button ( ) allows a user to request a player specific details, which are displayed in the information drawer on the right of the interface (Figure 7.9). Users can pin players to the players' drawer, shown on the left of Figure 7.9, for more efficient future re-selection by clicking on . Selecting a player from the players' drawers shows all the views in which the player appears as screenshots of the player at that moment within the other views as shown in Figure 7.9. Clicking on any of the screenshots changes the playing video to the corresponding camera view. Thus it allows users to follow the selected player and change the camera view based on that player. The pinned player's information can also be retrieved by clicking on button next to the player's name from the list, which will open the information drawer with the player's information. The information and players' drawers can be open or closed any time using the arrows in the corresponding drawer. This allows users to watch their videos without any distractions.

Annotation is integrated into *MediaDiver* and enables enthusiasts to create their

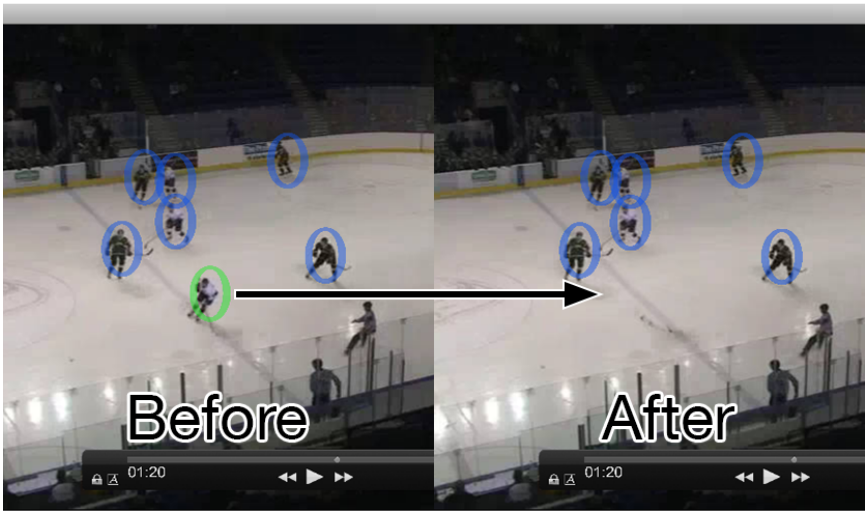


Figure 7.11: *MediaDiver* hiding feature allows users to focus on specific player while watching. The selected player is shown in the right image and after pressing the 'D' key, the player is removed from the video as shown on the left.

own rich media experiences. Our delayed annotation method allows users to add bookmarks (using) while viewing so that subsequently they may return to add their own annotations. Alternatively users may annotate at any time by enabling the editor interface (using) and add new or edit existing content. In the annotation mode, shown in Figure 7.10, users can manipulate the annotations or tags using , , and to create, edit, move and delete annotations respectively.

With annotations, users can perform more advanced tasks within the interface. For example, if a user wants to study or analyze only one specific player without getting distracted by the other players, *MediaDiver* offers a feature to hide one object or multiple objects from videos. Users can select the object they want to hide; once it is selected they can press the 'D' key from the keyboard and the object will be removed (i.e. hidden) as shown in Figure 7.11. The accuracy of this feature; however, depends on the correctness of the annotations provided to the system. Users also have the option to bring all the hidden objects back by pressing on the 'S' key from the keyboard.

Our multiple view rich media interface allows users to navigate video content

by switching views using gestures or manual view selection buttons (). Absolute view selection is supported with our grid interface (Figure 7.12(a)), and relative view selection can be accomplished with our video flow visualization (Figure 7.12(b)).

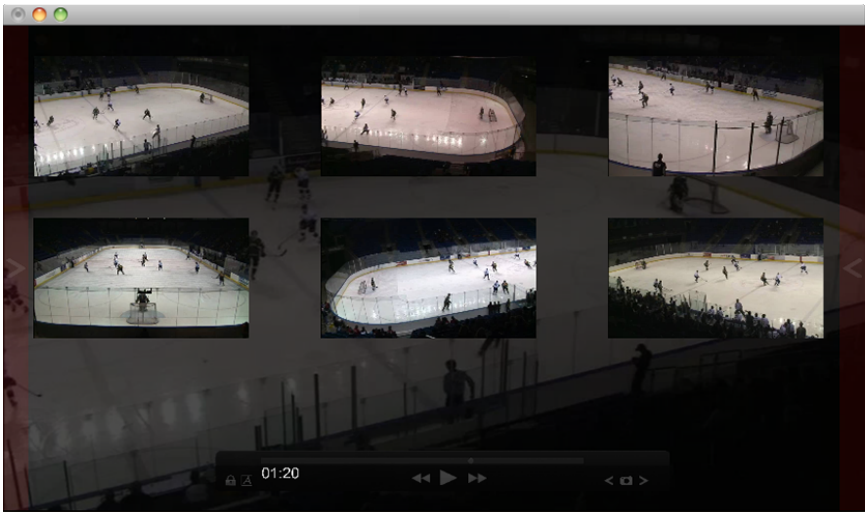
7.4.2 MediaDiver Applications

We imagine that our viewing and annotation environment, *MediaDiver*, will be critical to support interactive video being a primary media of web applications in the future, such as video sharing and social networking sites. Interactive annotated video supports home video organization and production for personal use and sharing over the network. Moreover, the *MediaDiver* is helpful for broadcasters looking to create meta content. For example, in sports broadcasts, the *MediaDiver* supports the following example functions:

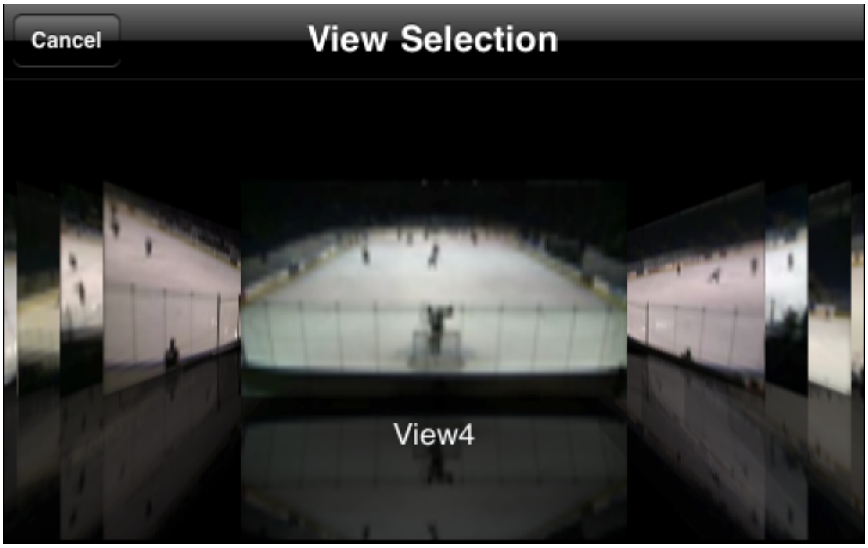
- Users can specify that a player should remain in view so that cameras are automatically adjusted to keep him in view;
- Links to hypertext data such as player information or statistics associated with the player, or other video (e.g. interviews);
- Users can pin player objects for future easy access;
- Users can easily adjust the camera view;
- Users can edit the current video annotation and easily view these changes.

7.5 Directions

We have shown that performance time for moving targets in 2D can be predicted for most situations using our models. For some variants of angles, the correlation is not as accurate as we would like, motivating further studies to test various angles and adjust the model. The work we have presented has been validated for linear movement and the mouse as an input device, which provides a foundation for future research and validation of pointing models for chaotic targets and other input devices (e.g. finger). We anticipate *Hold* will work well for complex movement types since a less predictable motion is more difficult to select with *Chase*.



(a) Absolute View



(b) Relative View

Figure 7.12: Switching between views in *MediaDiver* using (a) absolute view and (b) relative view.

Moreover, we think the chase technique will be disadvantaged for touch due to the repetition of tapping, lifting required when a selection fails; such as when trying to select small, fast moving targets. However, we believe *Hybrid* would work well because one touch anywhere pauses the video and users just need to slide their finger to the target. Thus, if the first touch misses, keeping the finger down and sliding it to the target will allow easy correction. We suspect that users will optimize the selection time by approaching the target for a rendezvous before touching the display. Further, we also foresee that rapid aimed movements, such as moving target selection may be a hybrid of an iterative correction model and an impulse variability model [99], suggesting a new area for future research. The first movement towards the target could be considered as an impulse task, where user tries to hit the target within a specific time in their mind, while the next moves are corrective movements. Thus, a hybrid model of Fitts' law [39] and Schmidt's law [103] may be an effective way to characterize rapid aimed movements.

We have also shown how *Hold* can be used in a rich media interface, the *MediaDiver*. Deploying this interface online would allow us to explore how users interact with the interface and how often *Hold* is applied for both annotation and information retrieval.

7.6 Summary

To summarize, in this chapter we introduced *Hold* and *Hybrid* novel moving target selection techniques. We performed studies evaluating these techniques and investigated the effect of target size, speed, movement angle, movement direction and their interactions on acquisition performance in both 1D and 2D. In 2D, our *Hold* method provides faster and more accurate selection performance for small or fast moving targets. Building upon prior work on 1D and 2D selection tasks, we introduced and validated variants of Fitts' Law that model selection of moving targets in both 1D and 2D. We have shown that our models can predict the performance time for moving targets.

Moving targets are likely to be common in future interfaces and in these environments accuracy is generally more important than speed. We have shown that our *Hold* and *Hybrid* approaches provide an effective interaction technique and offer

a fast annotation method that can be easily integrated into interfaces with moving targets as presented in the *MediaDiver*.

In the next chapter, we look at how to integrate our concepts described in this dissertation into a mobile application. This will allow us to reach a large audience and collect more representative results.

Chapter 8

Video Viewing History in a Mobile Application

Video viewing has become incredibly popular on mobile devices where they making up almost 40% of YouTube's global watch time [1]. According to BI Intelligence[2] report, globally, 15% of all time spent watching online videos is spent viewing on smartphones, tablets and other mobile devices. This offers another platform to investigate the video viewing history where it can reach a larger group of people from different backgrounds. Thus, we decided to develop a mobile application version of our history-based interface to be able to test our concepts with a larger group of audiences. In this chapter, we describe our mobile application (*Mevie*) and in Section 8.1 the challenges we faced. Section 8.2 describes the components of *Mevie* and the gestures proposed to interact with these components. A workflow of the application is presented in Section 8.3. We designed a short-term interview study (approval certificate #: H13-01589) to investigate our designs using the application, detailed in Section 8.4. Finally Section 8.5 addresses the refinements and directions for future research.

[1] https://www.youtube.com/yt/press/en-GB/statistics.html
[2] http://www.businessinsider.com/mobile-video-statistics-and-growth-2013-12

8.1 Mevie: History-based Mobile Application

Mevie is a history-based video viewing mobile application developed for the iOS platform, specifically for the Apple iPad. It encapsulates our research prototypes into a unified high-fidelity application for people to try emerging video experiences. It supports fast navigation to popular (or unseen) parts, search, quick retrieval of previously-viewed intervals, quick authoring for share-able movies, and directly previewing content, without interrupting normal playback. Many of these features are enabled by the use of our video history mechanism described in the previous chapters of this dissertation.

Integrating the designs and features into a mobile application introduced several challenges due to the differences in the platforms. One of these challenges is the change from cursor-based input to touch input, as some of our previous interactions include cursor hovering, which is impossible in the current mobile platform. Thus, a few modifications and new interactions have been introduced to replace these. The design language is another challenge that required us to convert our code classes and designs from ActionScript to Objective C, the language supported by the Apple iPad. The third challenge is the limited screen size available for the mobile application, which forced us to redesign some of the visualizations to fit the layout specified for our design described in previous chapters. Despite all these challenges, moving to a mobile application is worth doing since *Mevie* will allow us to reach a larger audience to evaluate our concepts and designs.

8.2 Mevie Components

The main goal of *Mevie* is to allow users to view different videos available on their devices and to access and manage their viewing history. Thus the interface is designed around allowing users to very quickly and easily access their video viewing history. Users can use the application to share different parts of videos from either their viewing history or the filmstrip. *Mevie* consists of three main modes: a home screen featuring videos that users can choose from, a main video viewer, and a history.

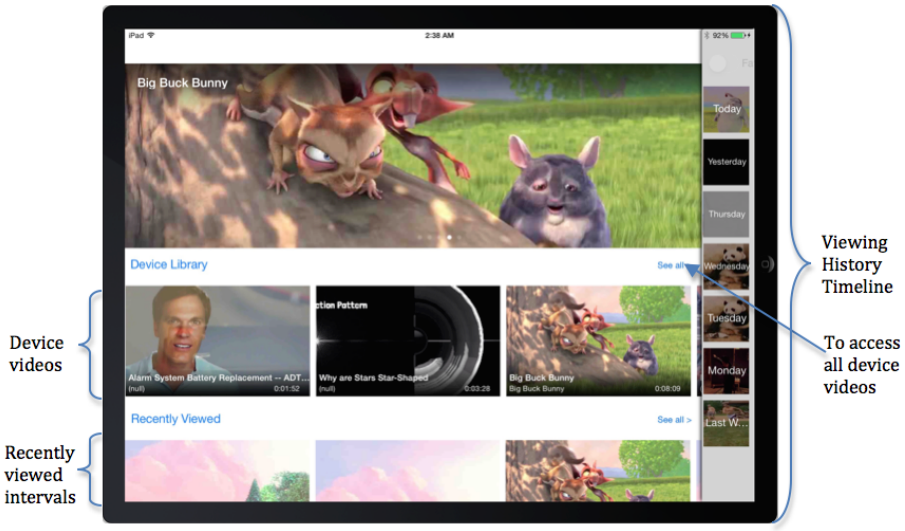


Figure 8.1: Mevie's Home screen illustrates lists of videos and clips that are accessible to a user. Each video or clip is represented by a small video preview; tapping on any preview transitions to the viewer and starts playing the video. More videos can be accessed by tapping on 'See All' for each corresponding list.

8.2.1 Home Screen

The interface first presents the home screen with lists of clips and videos as shown in Figure 8.1. The suggestions are pulled from the user's device, and the user's history in term of favourites, mostly watched, recently watched and recently shared. Users may also access a full list of videos or clips in each category by tapping on 'See All'. For example, tapping on 'See All' in the Device Library brings a grid of all videos available on the device as illustrated in Figure 8.2. Each video is represented by a thumbnail extracted from the video along with the video details such as the name and duration. Tapping any of the thumbnails transitions to the Viewer, shown in Figure 8.3, and starts playing the video.

8.2.2 Viewer

The Viewer allows users to watch any video and control its playback. As shown in Figure 8.3, the Viewer consists of a video player showing the content and a filmstrip

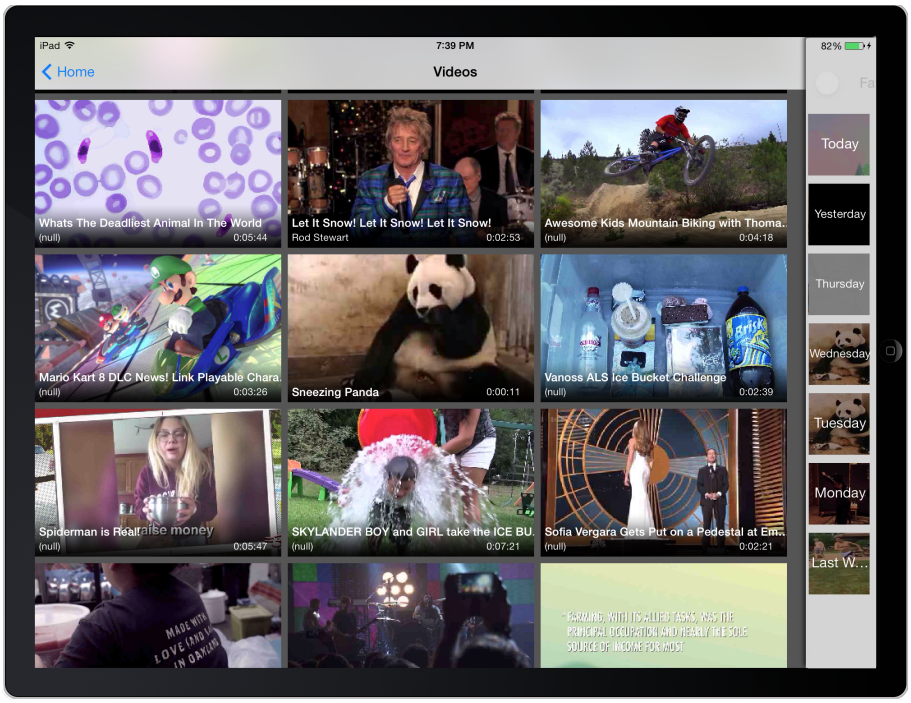


Figure 8.2: A grid of all videos available in the device library. Each video is represented by a thumbnail with a video name and duration. Tapping on any of the thumbnail transitions to the Viewer and starts playing the video.

for navigation. There are twenty thumbnails in the thin filmstrip at the bottom of the player, each representing one-twentieth of the video currently playing. Panning or dragging across the filmstrip shows a preview of the corresponding frame in the video, and causes the main video player to seek to that frame. Tapping at any location on the filmstrip also causes the video to seek to that location. As the video plays, a playhead moves across the filmstrip, indicating the current time location of the video.

Users can also navigate or seek the video by dragging or a long press on the video player itself as illustrated in Figure 8.4 where a bigger filmstrip appears consisting of five thumbnails. Dragging across this filmstrip also causes the main video player to seek to the corresponding frame. Tapping on ▭ or double tapping on the thin filmstrip also allows users to have a bigger locked filmstrip at the bottom

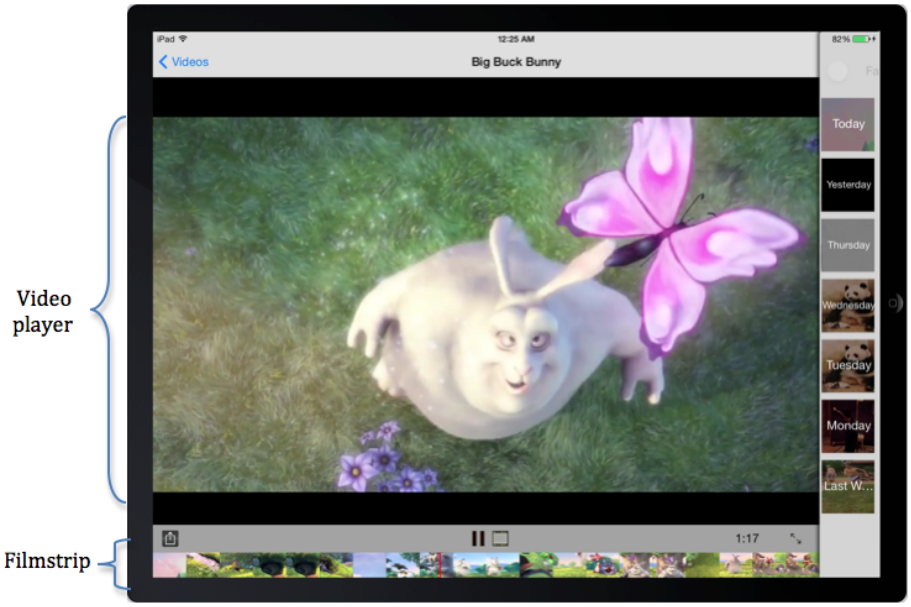


Figure 8.3: Mevie's Viewer screen consists of the main video player and the Filmstrip at the bottom. Tapping or dragging on the filmstrip allows users to seek the video being played.

of the player. Users can switch between the thin one and this bigger one using the same icon ▭.

A video can be viewed in full screen by tapping on ⤡ to remove any components that may cause any disruption while viewing. This option was provided since many participants in our previous studies (presented in earlier chapters) suggested having this option. In the Viewer screen, users also have the option to share the entire video or just portions of it. Tapping on ⬆ opens the sharing drawer from the filmstrip shown at the bottom, as illustrated in Figure 8.5 and automatically selects the last watched segment of the video. Users can share the last watched portion of the video, mostly watched portion of the video, or they can manually select any interval using gestures: one finger drag from start location to the end of the interval or placing two fingers on the filmstrip, with the left most finger indicating the start point and the right most finger indicating the end point. The start and the end of the interval can be easily adjusted by dragging the markers at the edges. This is made even easier by illustrating the start and the end previews with their corresponding

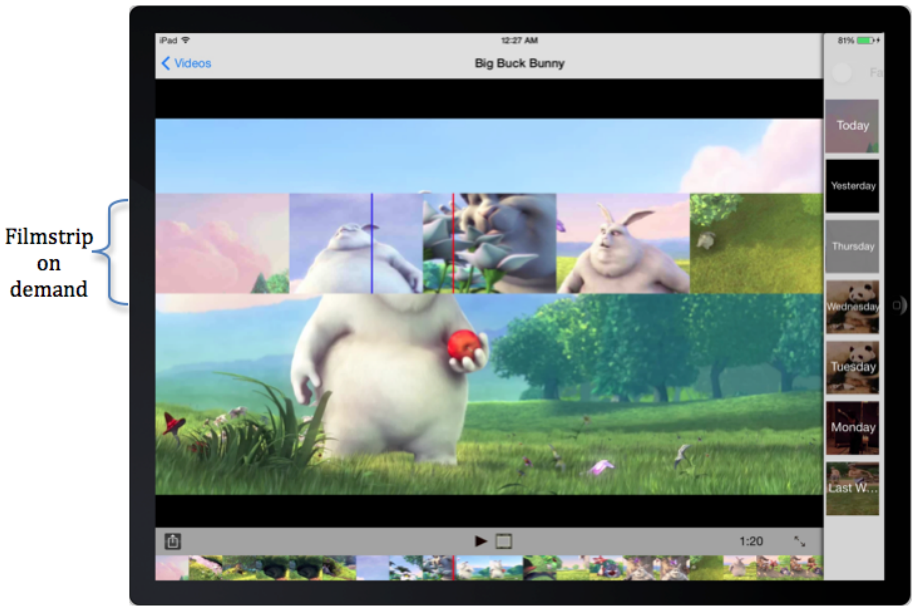


Figure 8.4: Dragging on the Viewer's main player brings up a filmstrip (as shown in the middle of the player) that viewers can use to navigate the current video.

timestamps above the filmstrip as shown in Figure 8.5. The user can also deselect the interval by tapping anywhere on the filmstrip, and reselect using the different options. Once users are satisfied with the selection, they can tap on 'Share' to bring up the traditional iOS sharing popup that allows users to share to all connected accounts of social networks capable of sharing video. Tapping on ⬆ returns users to the viewing mode where dragging on the filmstrip does not mean selection, rather, it is seeking the video in the main player.

8.2.3 Viewing History Timeline

The history represents the core component of the application. It is a collection of how a user watched a video space, and is a sequential log of all navigational actions taken by the user. Users can access their viewing history any time from the Viewing History Timeline available on the right of the application as a drawer. This component offers a detailed visualization of when and which videos a user

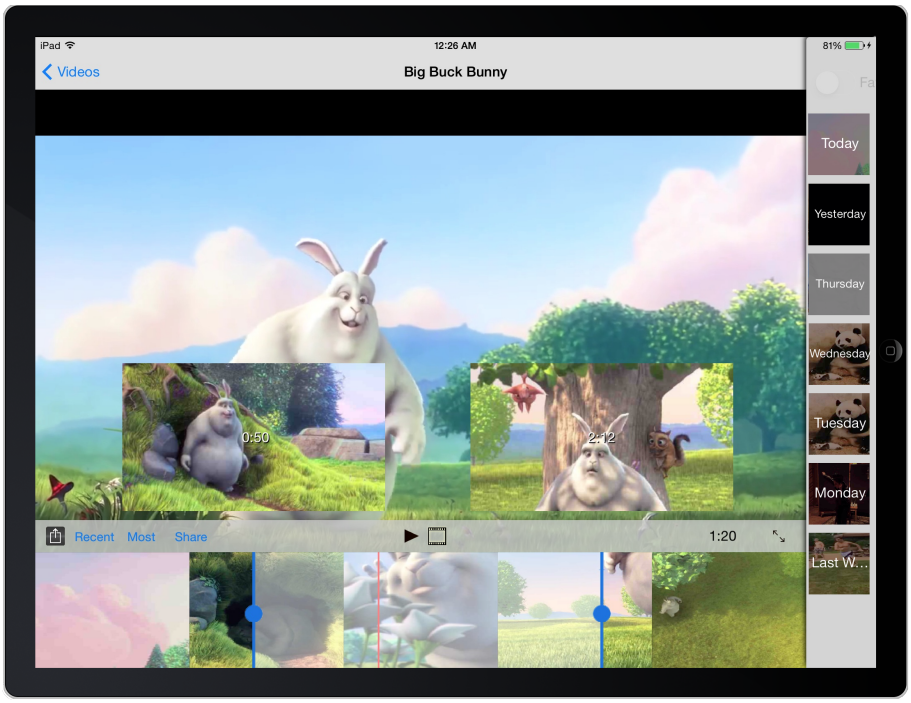


Figure 8.5: Sharing is possible from the filmstrip in the Viewer Screen using manual selection, most watched, or last watched clips. In the sharing mode, users can manually select any interval to be shared using gestures: two fingers or a panning gesture. The start and end of the selected interval can be adjusted by dragging the needed edge.

has watched. Users can even check which parts of each video they have actually watched. The Video History Timeline consists of two scroll-able columns where the first column shows the dates when there was at least one video being accessed. Each date is represented by a thumbnail illustrating the most viewed video on that date and a label indicating the date. These dates are sorted in a reverse chronological order (i.e. from recent). To avoid excessive scrolling, dates are grouped when they get older. For example, it starts from today, yesterday, then by day name if it is within the current week, followed by last week, this month, then by month name if they are within the current year; finally older ones are grouped by year as shown in Figure 8.6. This helps minimize the number of elements shown on the screen at the same time, which helps with the scalability of the interface.

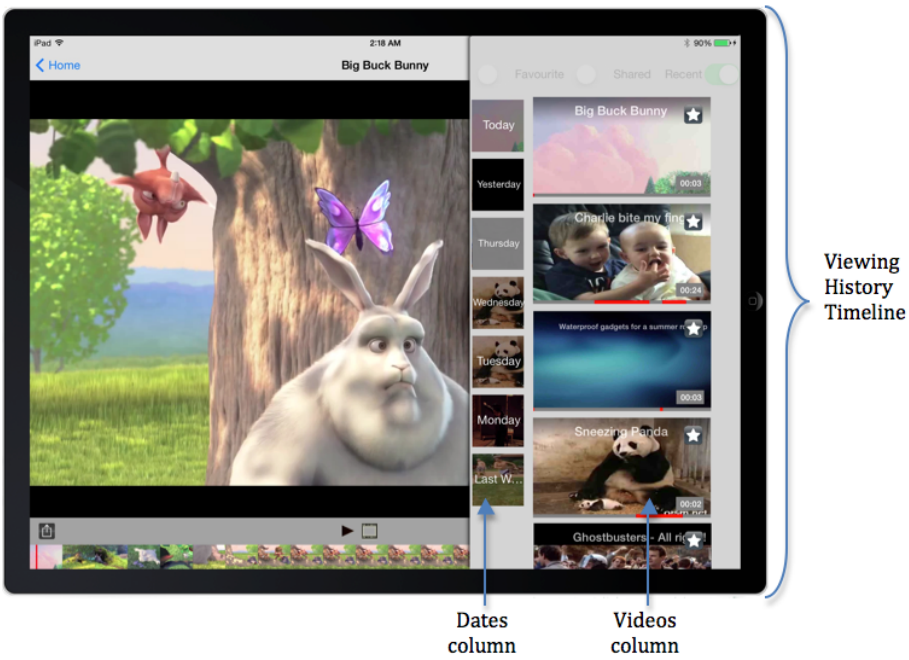


Figure 8.6: Mevie's Viewing History Timeline represents a sequential log of how a user viewed videos. It consists of two scroll-able columns where the first column shows the dates when there was at least one video being accessed and the second column shows all accessed videos. Each viewed video is represented by a seek-able and favourite-able thumbnail.

Pulling the drawer to the left shows the second column of this component, which illustrates the individual videos accessed on each date label. Each video is represented by a seek-able and favourite-able thumbnail with a temporal visualization following the design described in Section 4.3, where the portions watched from each video are highlighted in red as illustrated in Figure 8.6. This helps users easily check which parts have already been seen. Users can seek a video thumbnail by dragging horizontally. Favouriting any video can be achieved by tapping on ☆ on the corresponding thumbnail.

Users can check their entire history by dragging vertically. Moreover, videos within a specific date label can be accessed by tapping on the corresponding date thumbnail. Doing so transitions users to the summary history described below

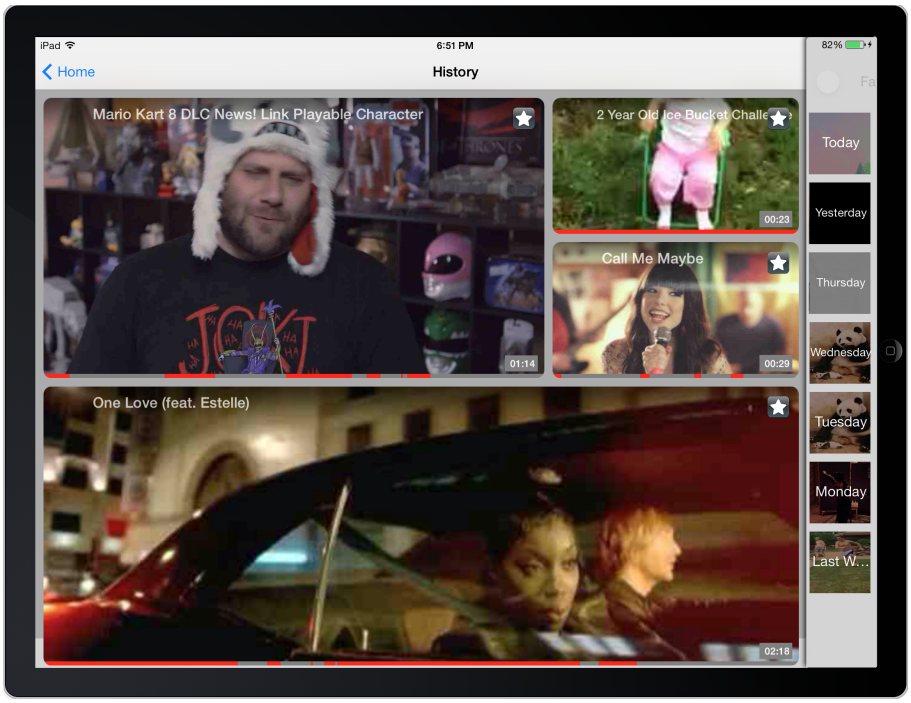


Figure 8.7: Summary Viewing History illustrates a collection of all videos accessed within a specific date. Each video is represented by a thumbnail that can seeked, favourited, played, deleted and open in the Viewer. The small timeline at the bottom of a thumbnail represents the entire video, with the viewed intervals highlighted in red.

(Figure 8.7). The detailed history of each video can also be accessed by tapping the corresponding video thumbnail from the second column of the Video History Timeline component.

8.2.4 Summary and Detailed Viewing History

A summary of the videos watched within a specific date can be accessed by tapping on the corresponding date thumbnail from the Viewing History Timeline drawer. Each video in the Summary Viewing History component, shown in Figure 8.6, is represented by seek-able, play-able, delete-able, share-able, favourite-able, and size-able thumbnail with a temporal visualization as described in Section 4.3. Each thumbnail shows a summary of which parts have been seen in the corresponding

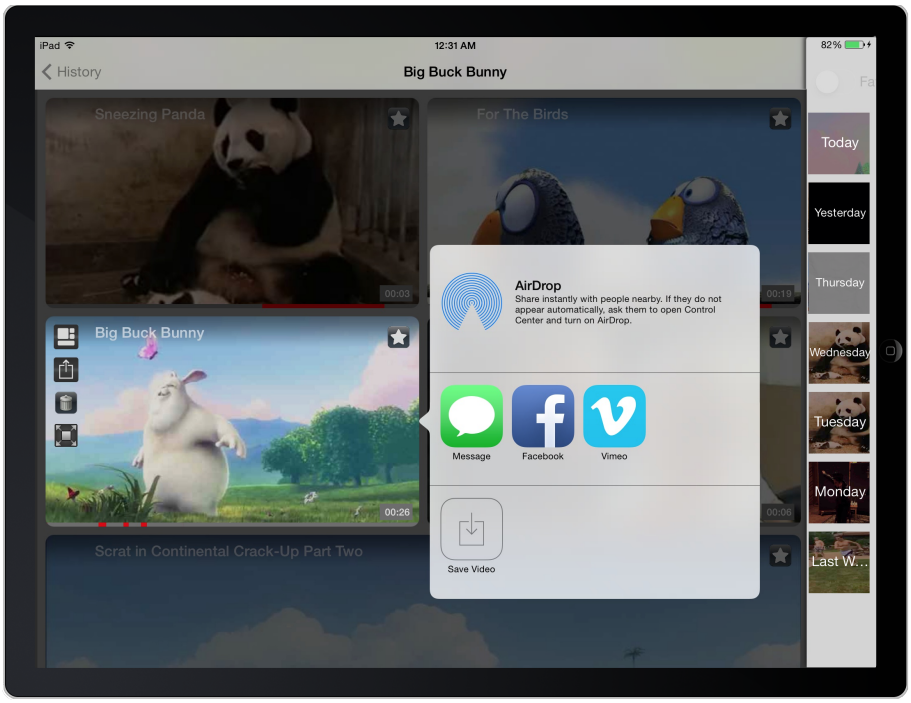


Figure 8.8: Any viewed interval or video in the user's history can be easily shared by tapping on ⬆, which brings a list of all available social networks.

video. The bottom of a thumbnail has a small timeline (i.e. temporal visualization) that represents the entire video, with the viewed intervals highlighted in red as shown in Figure 8.7. Dragging or panning horizontally on the thumbnail allows users to seek its contents. Users can also play the content of the video summary in the thumbnail using a double tap. If users wish to play it in the main player, they can do so by tapping on ⛶, which will transition to the Viewer screen. Favouriting and deleting a video from the history can be achieved by tapping on the corresponding ★ and 🗑 icons respectively. Similar to Filmstrip in the Viewer, users have the option to share some content from the history by tapping on ⬆, which opens a list of the available social networks for sharing video content (Figure 8.8). Since a video thumbnail is a collection of different portions from the video that might not be adjacent, tapping on ⬆ will combine these portions into a single mashup video. Hence, this allows users to easily generate short summaries or abstractions of their

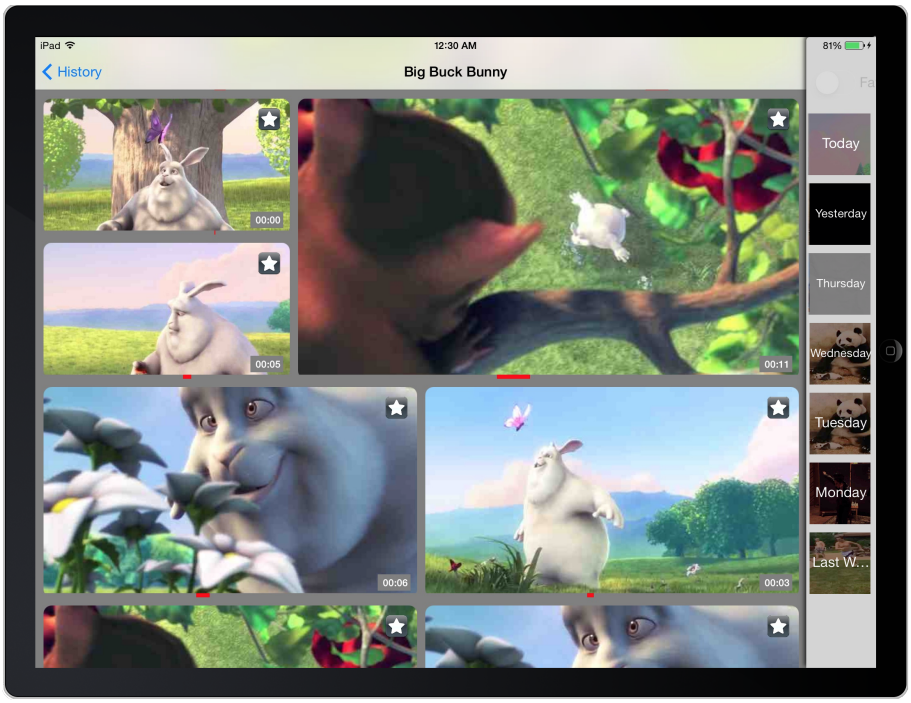


Figure 8.9: Detailed Viewing History illustrates how a single video has been viewed showing all viewed intervals from this video and in the same order how they were viewed.

video based on what they have seen from it. To access how each video has been viewed and to get more details about individual viewed portions from the video, users can tap on , which transitions them to the Detailed Viewing History.

The Detailed Viewing History component shown in Figure 8.9 visualizes each interval a user has watched from the selected video. It applies exactly the same features and designs as the Summary Viewing history with two exceptions. First, a thumbnail in Detailed Viewing History represents a single interval from a video rather than a union of multiple intervals as in the Summary Viewing History. Second, thumbnails in the Detailed Viewing History do not have the details icon ( ) since these are the highest level of the detail. All other features work exactly as explained earlier in the Summary Viewing History thumbnails.

Both the Detailed and the Summary viewing history apply the Video Timeline and Video Tiles design layouts discussed in Section 6.3 to visualize and arrange

their thumbnails. Thumbnails use different sizes to communicate how often the content was viewed: size is determined based on a weight factor that is derived from the view count and duration of the segment as explained in Section 6.3. In the evaluation section (Section 8.4) of this chapter, we are going to investigate the Video Tiles design for both components since we already tested both in Section 6.5. The application can provide both so that users can select the one they prefer.

8.3 Mevie Workflow

Mevie starts by opening the home screen from which users can select a video or a clip they would like to watch. The suggestions are pulled from the user's device, history, and favourites lists. Users can watch a video from their device by selecting the corresponding video from the device library list. If they want to check all videos available on the device, they can tap 'See All' as shown in Figure 8.1. This brings a grid of all videos (Figure 8.2) from which users can select. Users can also watch their favourited clips by tapping the one they like from the favourites list. They have the option to watch a clip they have already seen by choosing one from the recently viewed clips or by finding one from the history timeline (Figure 8.6).

Tapping any of the thumbnails transitions to the Viewer (Figure 8.3) and the video or clip starts playing. Users can seek the video using the provided filmstrip at the bottom of the player or by directly dragging in the player itself (Figure 8.4). Every time the user seeks the video, their history gets updated and a new interval is added. The Viewer offers users the ability to easily share clips from the video by tapping . If users were just watching a portion that they enjoyed and would like to share, they can tap 'recent' which selects the last portion they watched and then with one tap ('share') they can share it with others. However, if the selected portion is not the exact bit they would like to share, then they can refine the selection by adjusting its start and end simply by dragging the edges (shown in Figure 8.5) and then share. Users can also share the most watched part in the video or they can manually select any part they would like to share.

Accessing and managing viewing history can be easily accomplished any time using the viewing history timeline (Figure 8.6), which is always visible to the right of *Mevie*. Users can look for videos by date or by video. Tapping on any of the

dates brings a summary history (Figure 8.7) for that specific date where all videos accessed within that date are shown. Users have a summary visualization of what content has been seen from each video. They can play the summary of each video (i.e. union of the video's intervals) by double tapping the targeted video, and they can share this summary by tapping the share icon. Any history item can be favourited or deleted using the corresponding icons (and respectively). Users can access the details of how each video was viewed by tapping on its corresponding icon, which transitions to the detailed video history (Figure 8.9. All intervals viewed from the selected video are presented in the same order they were viewed. Users can play, share, favourite, and delete any of these intervals.

Users can always return to the previous screen by using the navigation controller shown at the top of the application. This allows them to always have the ability to go back to the home screen to select other videos to be watched and enjoyed.

8.4 Mevie in a User Study

We designed an interview study to explore how users behave when they have access to their personal navigation history and what they think about having access to their history. Our main goal was to investigate the design and usability of our history-based mobile application (*Mevie*). We anticipate that people will like having their history available for them and will appreciate the ease of sharing in *Mevie*.

8.4.1 Apparatus

Our mobile application was implemented using objective C and was tested on an Apple iPad Air and Apple iPad 4 running iOS 7.1.2.

8.4.2 Participants

26 paid volunteers, ten female and sixteen male, participated in the experiment. Participants ranged in ages from 19 to older than 60 years; thirteen participants were from 19-24, ten were 25-29, one was 30-39, and two were 60 years or older. Each participant worked on the task individually. Since the experiment was run at the Kaiser building, our participants were affiliated with UBC and all were from

the Engineering field except one who was from Science. One of our participants was a faculty member, four were undergraduate students, seventeen were graduate students, 3 were visiting students and one was a staff member. Fifteen participants watch online videos on a daily basis while the rest watch videos a few times a week. About participants' video sharing frequency, we had six who share videos a few times a week, three share once a week, twelve of the participants rarely share videos, while five reported that they never shared videos. Participants' details are shown in Table 8.1.

Table 8.1: Demographics summary for participants in the *Mevie* user study. (Note: Sci: Science, Eng: Engineering, FT/week: A few times a week)

P	Gender	Age	Occupation	Area	Watching Frequency	Sharing Frequency	Tested Method
1	Male	60+	Faculty	Sci	FT/week	Rarely	History
2	Male	19-24	Undergrad	Eng	Daily	Once a week	Filmstrip
3	Female	19-24	Grad	Eng	FT/week	Never	History
4	Male	19-24	Grad	Eng	Daily	Never	Filmstrip
5	Male	25-29	Grad	Eng	FT/week	Rarely	History
6	Female	19-24	Undergrad	Eng	FT/week	FT/week	Filmstrip
7	Female	60+	Staff	Eng	FT/week	Rarely	Filmstrip
8	Female	19-24	Visiting Grad	Eng	Daily	Never	History
9	Female	19-24	Visiting Stud	Eng	FT/week	Rarely	Filmstrip
10	Male	19-24	Undergrad	Eng	Daily	Rarely	History
11	Male	25-29	Grad	Eng	FT/week	Rarely	Filmstrip
12	Female	30-39	Grad	Eng	Daily	Once a week	Filmstrip
13	Male	19-24	Grad	Eng	Daily	Rarely	History
14	Female	19-24	Summer Stud	Eng	Daily	Rarely	History
15	Male	25-29	Grad	Eng	Daily	FT/week	Filmstrip
16	Male	25-29	Grad	Eng	Daily	FT/week	History
17	Male	25-29	Grad	Eng	Daily	Rarely	Filmstrip
18	Male	25-29	Grad	Eng	FT/week	Never	History
19	Male	25-29	Undergrad	Eng	Daily	FT/week	History
20	Male	25-29	Grad	Eng	Daily	Once a week	History
21	Female	19-24	Grad	Eng	Daily	Rarely	History
22	Male	19-24	Grad	Eng	Daily	FT/week	Filmstrip
23	Male	19-24	Grad	Eng	FT/week	Rarely	Filmstrip
24	Female	19-24	Grad	Eng	FT/week	Never	Filmstrip
25	Female	25-29	Grad	Eng	FT/week	Rarely	Filmstrip
26	Male	25-29	Grad	Eng	Daily	FT/week	History

8.4.3 Design and Procedure

We designed an interview study where participants had the opportunity to try *Mevie* and provide feedback on their experience and suggestions on the application. To compare our concept of having history to what normally users have in other video viewers, we defined sharing as the task for the experiment so that participants can imagine how this application can be used. Participants were divided into two groups where the first group ($n = 13$) shared video clips using the filmstrip in the Viewer screen (Figure 8.5) while the second group ($n = 13$) used history (Figure 8.8) to share clips.

Participants started by providing their consent to participate in the study. Then, a researcher gave them a walkthrough of the application where the researcher explained each screen, how to play and navigate a video, how history is created, how to share clips using filmstrip, how to play, favourite, delete, and share clips from history, and how to access summary and detailed history. Participants were then given the device to try the application by themselves for three minutes and were asked to talk through their experience as the audio was recorded. The researcher made sure that participants tried every aspect of the interface including viewing, navigating, and sharing from both filmstrip and history. During the free play, the researcher made note of two clips the participant watched from two different videos to be used for the sharing task. After the free play, the researcher asked the participant to share the two clips using the history or filmstrip depending on the participant's group. The sharing task was timed and then used for the comparison between the two methods: filmstrip and history.

Once participants found both events they were asked few questions in an interview setting while being recorded. They were asked about what they think about the application, whether they would use it, whether they would actively use the history, which method they preferred for sharing, and if they have any suggestions or recommendations of improvement for the application. After answering these questions, participants were asked to fill a short questionnaire (Section B.7) to provide feedback on the history and the interface in general.

8.4.4 Tested Videos

For users to try the application, some videos were needed to be available in the device. Based on the five most viewed categories found in Chapter 3, we chose 10 most viewed videos in each category from YouTube for users to test in *Mevie*. Thus, participants had 50 videos to use while exploring and testing the application.

8.4.5 Results and Discussions

The main goal of our study was to evaluate how simple and intuitive having history within a video player interface is for users. When we asked participants if having history as a part of the video player was useful, they highly agreed ($Md = 6$ on a 7-likert scale) that it is useful and helpful ($Md = 5$ out of 7) to have it always available. Some participants indicated that being able to hide the history and bring it up when it is needed would be better for them especially when they are watching a video in the Viewer. In order to increase the utility of the history, we developed the application around the history, and made it easily accessible in all modes inside the interface. We wanted to know more specifically if it is easy and natural to have the viewing history in a drawer. The viewing history timeline in a drawer was found easy to browse ($Md = 6$) and was rated highly as a natural way to present the history ($Md = 6$). These results indicate that participants positively perceived having history in a video player and it did not bother them having the history always available in the application.

About the ease of finding and sharing clips from both filmstrip and history, participants highly agreed that finding clips was easy using both ($Md = 6$) and they strongly agreed that sharing was easy in both components ($Md = 7$). This was also seen on the time measured for the sharing task using the two methods, where an independent-sample t-test revealed insignificant difference between the two methods, which we expected due to the short time participants had to use the application. The mean time needed to find and share clips using history was 25.2 seconds while with filmstrip they took 27.9 seconds on average. However, when participants were interviewed and asked for their preferred sharing method, filmstrip or history, thirteen participants preferred the filmstrip while four preferred history and the rest ($n = 9$) liked them both. Participants indicated that if they have

seen the clip they want to share then they might use the history for that but if the clip has not been seen before or needs refining at its start and end then they will use filmstrip for sharing. Some pointed out that for immediate sharing while viewing, they would use filmstrip but to look for previously seen clips or for later share, they would use history. These results indicate that our designs and interactions used for sharing clips were found very easy to use for both filmstrip and history.

Participants were also interviewed to get their thoughts about the application and any comments/suggestions to improve it. In regard to what participants thought about *Mevie*, most participants reported that it is good ($n = 15$), cool ($n = 3$), awesome ($n = 2$), neat ($n = 3$), useful ($n = 7$) and interesting ($n = 5$). The ease of sharing within the application was really appreciated by participants and they found it better than what is currently available on other video players. Appendix C, Table C.1 shows all participants' response and few of them are highlighted below.

> *"I think it's good, because when I watch another video, I want to skip forward, I guess I need to remember what part I have already seen. In here I can see, what part. This is a very advanced application."*

> *"I like this sharing thing because sometimes you just want to share part of the video, sometimes videos are really long and you only want to share a few seconds. I think this is a very good, application for your app. The history is also useful. Especially because you can see the most viewed part of each video, instead of only the beginning or end, so it is more representative for you to remember."*

> *"I think best part of it is the feature that we can share part of the selected video because usually when we watch video online, we share the whole video. But in this one, we can select the best part we want to watch. Because usually, when I watch the videos shared by my friends, it's very long, and I don't want to go through all of it, and if using this app, we can just watch the best parts, or share the best parts we want to share. So this is the best thing about this app."*

> *"It's very useful because sometimes I need to replay part of the video and it's really only this part of the video I'm interested in. I always have some*

difficulties to find which part of the video I should go down and find where is it. It's a really good application."

To explore whether users were willing to use our application as their main video player, we asked them if they would use *Mevie* if they had it. All participants reported that they would definitely use it and they would like to have it integrated to online video providers such as YouTube and Vimeo because they think most people use these websites to watch and share videos. Participants' replies are shown in Appendix C, Table C.2 with some highlighted here.

"I don't know, because, when I watch things like this, I usually go on YouTube. If it was part of YouTube, then yeah I would definitely do it. Especially if I wanted to share it. I think it's really clever ... It's really user friendly."

"I'm not a video lover. I have seen editing video parts, those take more steps. I prefer this app if I do it."

"Yeah, I think it would be all about the database. If you could access YouTube, or other websites, Vimeo etc."

Most participants, as we mentioned earlier, acknowledged how easy it was to share using *Mevie* where most preferred using the filmstrip rather than the history. So to investigate the usability of the history, we asked participants whether they would actively use the history. Most of them answered 'Yes' and for those who said 'No' they think others might use it but not themselves. Some participants think they will definitely use it to look up things they watched a long time ago. Participants' answer are presented in Appendix C, Table C.3 and a few of them are highlighted here.

"Yeah, the history would be really good because you can go back and check certain things or see, depends on what you had or that you kept. I think the history would be great. That would be a good thing. I would really like that."

"Yes, I would use it. Yesterday, I had the problem of finding a part of the video, because I need those part of videos to watch it again."

> *"The history is not really necessary for me. It's a good part of the app. It should be for some other users. Not for me. I always watch a video one or two times at most and share it."*

> *"I would use history, but I don't know how much of the sharing mechanism I would use. The history is very natural. Visibility of the drawer provides good affordance for something new."*

Participants suggested many areas for improvement in the application. They would like to be able to tag, annotate or comment on the videos and history items. They also would like to have some mechanisms or tools to search for clips. Using tags or filters will allow users to search their history. We are planning to add these features in a modified version of the application. A few participants pointed out that most times thumbnails look similar and participant 16 suggested to blank those with similar colours. We think having a search functionality will mitigate this problem since we cannot do anything with thumbnail colours, which depend on the video content. In regard to the sizes used in the application, one participant criticized the size of the buttons for the history segments (Figure 8.8) as being too small, while another did not like the font sizes in the labels for the history timeline drawer (Figure 8.6), he preferred a consistent medium font size. This participant also indicated that the theme colours were not attractive and he suggested using light colours instead, while another participant liked the colours. One of the participants suggested different interactions for history segments to bring up the context menu (shown in Figure 8.8) where he recommended a long press instead of a single tap. A single tap would then be used to play the video instead of using double tap, which is not commonly used in mobile applications. Some participants would like to play a selection from the filmstrip to be able to easily assess its content before sharing. A participant indicated that having the viewing history drawer to the right was confusing and not natural since it is used differently in the YouTube application. YouTube uses the right drawer to show related videos. Thus, he suggested having it at the bottom instead. However this will conflict with the Control Center drawer in iOS. Investigating *Mevie* with more users will reveal if this is an issue or not. Two participants proposed having a tutorial on how to use the application since it is new, which we think is necessary for the released version to help users

master *Mevie*'s functionality and power. Some participants suggested including online videos along with personal videos to provide more options for users and utilize the collective wisdom in recommending clips from videos.

In summary, *Mevie* received a lot of positive feedback and comments from several participants. They appreciated having their history as a part of a video viewing application and they noted its importance for viewing and sharing videos. Some participants gave us some insight on how they could see the interface and its history being used. One participant suggested that they could see the history used to quickly create highlight videos for online video game streams. This can be extended to any lengthy video, such as a movie, television show, or sports event, instructional video, or documentary. Another participant stated that they would use the history to find videos of a similar nature. They foresee its best usability in online videos since most people now watch videos on these providers (e.g. YouTube and Vimeo). Some features need to be added to *Mevie* to make search even faster as suggested by some participants. *Mevie* already can filter history items based on the favourite items and we are looking at offering users the ability to add tags to their history items and then use these tags to find their requested clips following Craggs et al. [35]. Participants also mentioned that it would be better if they could refine history items before sharing to have similar flexibility as sharing from filmstrip. Therefore, we will look at implementing this feature for history items and as future work we will allow them to add clips to a playlist and refine each clip before combining them in one video.

Even though participants only had around five minutes to play with *Mevie*, they were astonished by how easy it is to use, find and share videos. They welcomed and appreciated having the history within the application. This indicates that this might be the time to integrate video players into a history-based video viewing interface.

8.5 Directions

The application's validation and iteration has only begun, and should be evaluated further with more people and in a longitudinal field study. Future field studies can be done with no researchers' influence on users' experience where *Mevie* will be

released for installation and users can try it on their own devices and videos. This will allow us to gather more data and feedback that will help in improving the design and introducing new features for the application.

We intend to integrate *Mevie* with one of the online video providers like YouTube or Vimeo. This will provide users with a larger collection of videos that they can view and experience. This will also attract more users to install our application and thus more data can be collected and employed in the next iterations of the mobile application.

8.6 Summary

To summarize, we have presented an integration of our history-based video viewing interface into a mobile iOS application called *Mevie* developed for Apple iPad. It offers users the ability to view, navigate, share, and manage their videos in addition to the accessibility of their personal viewing history. Through a short-term interview study, we were able to explore how people perceived having their history available in a video player and test the usability of the application. Participants rated the application features highly, liked having the history and were very enthusiastic with the ease of sharing video portions using *Mevie*. They also recommended integrating our application with online video providers such as YouTube and Vimeo.

In the final chapter, we summarize the contributions of the dissertation and propose a few directions for future investigation.

Chapter 9

Conclusions

The key goal of this dissertation was to create validated design strategies for future video interfaces to match users' ways of interacting with contemporary video. Further, these strategies are translated into tools that can be easily integrated into existing systems. This chapter concludes the dissertation by outlining and summarizing the contributions that came out of this research. It also discusses the limitations of the research and the directions for future research. Finally, concluding comments are presented.

9.1 Dissertation Contributions

The main contributions of this dissertation include a behavioural analysis and characterization of how users interact with online videos, creation and validation of new techniques for video navigation, and a validated mathematical model for moving object acquisition. While these contributions were tested in a desktop platform, they can be easily transferred to mobile devices, as shown in Chapter 8, where we think such an interface would be most likely to be used.

9.1.1 Behavioural Analysis of Users' Video Viewing Actions

Previous research on users' viewing behaviours were either undertaken for specific types of videos (e.g. education, instructional videos), specific groups of users (e.g. learners), or analyzed specific interactions (e.g. play, pause, rewind, skip). All of

the work presented in the literature analyzed the collective data of multiple users and presented their findings and suggestions based on that. None of the previous work, to the best of our knowledge, has looked at the individual's interactions and if they exhibited similar behaviours. This is important if we are going to propose a new interface and tools for personal usage.

To achieve this, we developed a web browser extension for Google Chrome that allowed us to collect users' viewing behaviours while watching videos on a YouTube web browser on a desktop platform. Five months of collected traces were analyzed to understand how people navigate and view videos and to determine their viewing behaviour. We concluded that people are actively watching videos and demonstrated that people re-watch all video types equally and that they do it often. The data revealed that when users accessed videos the next time, it is mostly to refer back or re-watch/enjoy something that has been seen previously and not to be resumed from where it was left. We also showed that the drop-off has little to no correlation with the video length and popularity indicating the subjectivity of users being uninterested in the content. At the individual level of the analysis, we discovered that most behaviours occurred for almost every user. We found that most users were re-watchers (i.e. watch portions of a video multiple times), skippers (i.e. jump around a video to find specific information or to pass over irrelevant and uninteresting parts) or both.

Explicit users' metadata can complement the implicit data gathered automatically from users' actions. It can add explanations and justifications for the actions that users applied while watching. This will provide a deep understanding of users' viewing behaviours and better correlation of actions and purposes. This offers an important factor for consideration in future studies for a more comprehensive understanding.

9.1.2 Video Navigation Techniques

Our main goal was to create and validate design approaches to match video viewing behaviour and cognitive strategies to improve people's ability to manage and share contemporary video content. We did this by creating different prototypes of History-based Video interfaces targeting three different contexts, as well as creat-

ing a mobile version to integrate these strategies. We used the findings from our first contribution to design and implement an interface that keeps track of users' viewed portions from each video and provide access to this data. This allowed users to view, navigate and find moments from previously seen videos faster. In order for users to be able to access their history data, different visualizations were designed and evaluated using three contexts as follows:

1. **Single-video history:** We used a chronological ordered list of thumbnails as a visualization of the viewed portions within a video. Using this visualization we were able to offer a faster tool to navigate to different parts within a video and provide a quick search tool for previously seen content in a video. By evaluating this visualization against the filmstrip, we proved that users were faster in finding answers to specific questions from previously seen content using our new visualization.

2. **Summarized history in a filmstrip:** Active interactions within a single video creates a long list of viewed segments, which may hinder the search task or make it difficult to get an overview of how a video is viewed. Hence, we employed the viewing statistics of a video to construct a histogram-based filmstrip, which we call *VCR*. It makes it easier for users to visualize the popularity of different portions of the video content. Through a user study, we proved that searching for specific events in a previously seen video using *VCR* outperformed the state-of-the-art filmstrip and people appreciated it as a summarization tool.

3. **Multiple-video history:** After proving that viewing history provides a faster search tool for previously seen content within a single-video, we wanted to test how this would work in multiple-video context. Thus, we designed two different visualizations for multiple-video history: *Video Timeline* and *Video Tiles*. Testing these designs revealed that searching for previously seen events was significantly faster using both designs in comparison to the state-of-the-art method.

9.1.3 Moving Object Selection Model

To alleviate the burden of selecting hotspots in interactive videos, we modeled the time needed to select moving objects and validated this model. Based on this model a novel acquisition technique (*Hold*) was developed. Our *Hold* technique temporarily pauses the content while selection is in progress to provide a static target, which removes the speed factor presented in the model.

Formulating a mathematical model

By adapting previously proposed models in the literature for static targets, we developed a new model to estimate the time needed to select moving targets. This model estimates the time based on the target size, speed, direction and angle of movement. Our model, in vector notation, suggests its applicability for 3D moving targets, which is worth investigating in future studies. It would allow us to check whether the same factors determine the selection time needed in a 3D environment.

Validating the model

A user study was designed to validate the proposed model for the moving target. The data showed a very good correlation with the model, verifying its validity.

Applying the selection technique in a working interface

To test the feasibility of our proposed selection technique (*Hold*), we designed and implemented an interactive rich media interface, called *MediaDiver* for experiencing, viewing and annotating complex video domains and their associated metadata content. It enables viewers to interact, browse, explore and annotate (i.e. tag) multi-stream sport videos. Making this framework accessible for others would allow us to thoroughly investigate the techniques and how it is applied.

9.2 Limitations

This section describes some limitations to the work presented in this dissertation. Some of these are also addressed in Section 9.3. The first limitation is that all of our studies (aside from the behavioural study described in Chapters 4, 5 and 6) were

short-term controlled laboratory experiments, which may not reflect the normal interactions people may exhibit when they are in their own private place and level of comfort. We tried to address this issue by asking participants to choose the events they liked instead of imposing our defined events; however, giving users the interface installed on their own platforms would provide more valid measures for our evaluations.

The second limitation is the feedback and ratings of our evaluated designs. Due to the newness of our designs there is a possibility of bias towards either the known design or our new ones. Also, participants wanting to please researchers is another issue that can cause biased results. Having only one method tested per participant might help but users may still compare it to an existing tool they might know of. Having said that, the collected data from the quantitative measures strongly proves our results but we still need users' feedback since it is an important factor in designing user interfaces.

The third limitation is the number and type of videos used for evaluations. Due to the time constraint of the laboratory experiments, only a few videos were tested to investigate the techniques proposed, so these designs were not validated for general types of video but rather there was evidence to support our hypothesis. The ideal usage of these methods and interface would be to allow people to watch any kind of videos they prefer without influencing their decisions. This would be accomplished in a field study, which is the next step in the future evaluation.

The fourth limitation is that our samples and demographics do not reflect the whole population since most of our participants are students and range in age from 19 to 40. Thus the results presented in this dissertation cannot be generalized to every person who watches videos. However, it is really to help those people who have the behaviours we presented in this dissertation for example re-watchers, skippers and revisitors. We are showing that we can make better interfaces tailored to the behaviours we found. Thus not only the people we are studying might benefit, but also other people who use videos like this group will be able to use it. But we do not know that yet. So releasing our mobile application (described in Chapter 8) or integrating the concepts defined in this dissertation into an online video interface such as Vimeo or YouTube will allow us to cover broader groups and collect more data. This will enable us to test how our findings fit with the larger group and to see

how many people will actually find these features useful beyond our small group.

9.3 Directions

We discussed some of the directions for the research components of this dissertation at the end of each chapter. Here we review some of these directions and provide broader directions for future research in releasing our interface and designs for personal use and integrating them for educational classroom participation and reflection.

Behavioural analysis of video viewing using cross platforms

The behavioural study presented in this dissertation was developed for a desktop platform and more specifically works only in the Google Chrome browser, which limited the size of the data collected. According to W3Schools.com[1] statistics of browser usage in July 2014, Chrome was the most used browser (59.8%) and Firefox came second with 24.9% usage. This indicates that our results did not account for the 40% of users who watch videos on browsers other than Chrome. Moreover, these percentages count for the wide variety of platforms and devices, and according to YouTube[2] 40% of their viewed videos were watched on mobile devices. This indicates that more people are viewing videos on a variety of devices and platforms, which encourages us to develop a cross-platform plug-in that would provide us with a richer resource of data. Furthermore, as we mentioned in Section 9.1.1 collecting explicit data along with implicit interactions would give us a broader understanding of users' viewing behaviour by correlating these to their intention. This will provide a resource for other researchers working in this area to assist their decisions when it comes to proposing new interactions and applications for video interfaces.

Introduce new features by allowing explicit-user metadata

One of the big issues with history is the burden of searching such data. Researchers have proposed several techniques that include the use of text search. Providing

[1] http://www.w3schools.com/browsers/browsers_stats.asp
[2] https://www.youtube.com/yt/press/en-GB/statistics.html

users with a tagging tool with which they can annotate different parts from their viewing history would allow them to filter out a large portion of their history. Having this feature along with our proposed filtering tools described in Chapter 6 would speed up the search.

Evaluating the interface and methods in a field study

As mentioned in the limitations in Section 9.2, our results are based on a small group of users who experimented with the different aspects of this dissertation in a short-term controlled laboratory study. The controlled lab experiment helped us to explore the complexity of the video space to be evaluated for certain methods and especially the history. A more comprehensive experiment would not have given us the specific insights we have achieved with this type of experiment - under perfect conditions. And now since we have proven their validity in such a controlled study, we can investigate it further using a field study with a larger group of users from different backgrounds can try it. This can be achieved by releasing our application (described in Chapter 8) or deploying our interface online where it is linked to an online video website to allow users access to any video they prefer. This might also allow us to evaluate the interface on multiple platforms. Furthermore, a longitudinal study would allow us to investigate the performance of a long-term history. This will provide a good opportunity to test the scalability of the interface.

With regard to *VCR*, as discussed in Section 5.2, using the crowd-sourced data provided a potential tool for recommending clips from videos to new viewers. Integrating the *VCR* into an online video website such YouTube would provide a potential test bench for the method since YouTube already has this data recorded but they are only available to the video's owner. This would allow us to explore how people are going to use it when accessing new videos. Would they watch what other people viewed from that video to drive their decision or would they completely ignore it? We can also offer the users the choice to switch between their own *VCR* and crowd-sourced ones to give them more options for navigation.

Evaluate the scalability of the visualization

The scalibility of our designs described in Sections 5.1 and 6.3 were not fully tested since we had only short-term user studies. The segments presented in these designs are a collection of thumbnails extracted from the video once the video is being played in the main player. These thumbnails are stored on the local storage, which means having a large number of long viewed segments in the history will take a much more memory. This can be solved by placing these thumbnails on a server hosted online instead of using local storage; however, the privacy issue will arise. Also, accessing the interface when users are not connected to the internet would be impossible. Another option would be creating thumbnails when needed and removing them otherwise. For example, for these segments, users starting a seek would create their corresponding thumbnails on demand with a short delay at the beginning. This might affect the responsiveness of the interface, which could create frustration from the users' end. Therefore, as future research we would like to explore the different options for keeping these thumbnails and evaluate them in a large scale study.

Integrate the interface with MOOCs

One of the potential areas that can benefit from our interface aside from home usage is the field of education and specifically the massive open online courses (MOOCs). MOOCs provide course material such as videos, readings and problem sets along with interactive user forms. It helps with collecting both implicit and explicit users' behaviours while they watch and review lectures' video. Kim et. al [69, 70] have been studying learners' behaviour on such platforms; however, they have not provided learners with the options we described in this dissertation. Integrating our tested designs with theirs would allow learners to more productively review their lectures and reflect on them based on their personal viewing and other learners' viewing patterns.

9.4 Concluding Remarks

To conclude, this dissertation has presented a version of the future video interfaces, introducing new ways for navigating and thinking about videos. The current model

that people use for working with videos is being shaped by the interface they use. So it is a constraint on the way they think of the time based media of video. Changing the underlying mechanisms of video can change the way we think about video as a medium. It changes who uses the medium, how we use video and how we think about the video viewing experience.

By doing what we have shown in this dissertation and from working on all of these experiments with people, we have seen that as people start using the interface they start thinking of video as an expressive media in which they can express themselves, much like text. However, much remains to prove this, which requires further studies looking at how changes of the design influence people's views of time-based media. By making video easy to archive, search, retrieve, author, and share, video becomes as much an extension as text or pictures to human expression. This enables thinking through video in the way we think when we write because authoring, editing, experiencing, and sharing will become part of our natural human cognitive process. Therefore, we will have new ways of thinking that are not constrained by the linear time of video. This will open the door to new uses of the medium in various contexts such as education, training, entertainment, social documentation, communication, and marketing.

This dissertation lays the foundations for that change as we began to see behaviour changes once the features are integrated into people's video viewing experience. We believe there is a whole future in making that a reality and we take the first steps in solving parts of it in this dissertation. This really shows that people start thinking differently and there is a shift in thinking continuously occurring.

Bibliography

[1] J. Accot and S. Zhai. Beyond fitts' law: Models for trajectory-based hci tasks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 295–302, New York, NY, USA, 1997. ACM.

[2] J. Accot and S. Zhai. Refining fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 193–200, New York, NY, USA, 2003. ACM.

[3] M. Ajmal, M. H. Ashraf, M. Shakir, Y. Abbas, and F. A. Shah. Video summarization: Techniques and classification. In *Proceedings of the 2012 International Conference on Computer Vision and Graphics*, volume 7594 of *ICCVG'12*, pages 1–13, Berlin, Heidelberg, 2012. Springer-Verlag.

[4] A. Al Hajri, S. Fels, G. Miller, and M. Ilich. Moving target selection in 2d graphical user interfaces. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction*, volume 6947 of *INTERACT '11*, pages 141–161, Berlin, Heidelberg, 2011. Springer-Verlag.

[5] A. Al-Hajri, G. Miller, S. Fels, and M. Fong. Video navigation with a personal viewing history. In *Proceedings of the 14th IFIP TC 13 International Conferenceon Human-computer Interaction*, volume 8119 of *INTERACT '13*, pages 352–369, Berlin, Heidelberg, 2013. Springer-Verlag.

[6] A. Al-Hajri, M. Fong, G. Miller, and S. Fels. Fast forward with your vcr: Visualizing single-video viewing statistics for navigation and sharing. In *Proceedings of the 2014 Graphics Interface Conference*, GI '14, pages 123–128, Toronto, Ont., Canada, 2014. Canadian Information Processing Society.

[7] A. Al-Hajri, M. Fong, G. Miller, and S. S. Fels. How personal video navigation history can be visualized. In *Proceedings of the 41st*

International Conference and Exhibition on Computer Graphics and Interactive Techniques; Posters, SIGGRAPH '14, page 105, New York, NY, USA, August 2014. ACM.

[8] A. Al-Hajri, G. Miller, M. Fong, and S. S. Fels. Visualization of personal history for video navigation. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1187–1196, New York, NY, USA, 2014. ACM.

[9] J. Alexander, A. Cockburn, S. Fitchett, C. Gutwin, and S. Greenberg. Revisiting read wear: Analysis, design, and evaluation of a footprints scrollbar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1665–1674, New York, NY, USA, 2009. ACM.

[10] J. M. Almeida, J. Krueger, D. L. Eager, and M. K. Vernon. Analysis of educational media server workloads. In *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, NOSSDAV '01, pages 21–30, New York, NY, USA, 2001. ACM.

[11] T. Asano, E. Sharlin, Y. Kitamura, K. Takashima, and F. Kishino. Predictive interaction using the delphian desktop. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, pages 133–141, New York, NY, USA, 2005. ACM.

[12] M. Avlonitis and K. Chorianopoulos. Video pulses: User-based modeling of interesting video segments. *Advances in Multimedia*, 2014, 2014.

[13] P. Baudisch, E. Cutrell, D. Robbins, M. Czerwinski, P. Tandler, B. Bederson, and A. Zierlinger. Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch and pen operated systems. In *Interact 2003 - Ninth IFIP TC13 International Conference on Human-Computer Interaction*, pages 57–64, 2003.

[14] P. Baudisch, A. Zotov, E. Cutrell, and K. Hinckley. Starburst: A target expansion algorithm for non-uniform target distributions. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '08, pages 129–137, New York, NY, USA, 2008. ACM.

[15] L. Bergman, V. Castelli, T. Lau, and D. Oblinger. Docwizards: A system for authoring follow-me documentation wizards. In *Proceedings of the*

18th Annual ACM Symposium on User Interface Software and Technology, UIST '05, pages 191–200, New York, NY, USA, 2005. ACM.

[16] R. Blanch, Y. Guiard, and M. Beaudouin-Lafon. Semantic pointing: Improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 519–526, New York, NY, USA, 2004. ACM.

[17] C. Brooks, C. Thompson, and J. Greer. Visualizing lecture capture usage: A learning analytics case study. In *Proceedings of the LAK 2013 Workshop on Analytics on Video-based Learning*, volume 983 of *WAVe 2013*, pages 9–14. CEUR-WS, 2013.

[18] W. Buxton. A three-state model of graphical input. In *Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction*, INTERACT '90, pages 449–456, Amsterdam, The Netherlands, The Netherlands, 1990. North-Holland Publishing Co.

[19] S. K. Card. The model human processor - a model for making engineering calculations of human performance. *Human Factors Society, Annual Meeting, 25th, Rochester, NY; United States*, pages 301–305, October 1981.

[20] S. K. Card, G. G. Robertson, and W. York. The webbook and the web forager: Video use scenarios for a world-wide web information workspace. In *Conference Companion on Human Factors in Computing Systems*, CHI '96, pages 416–417, New York, NY, USA, 1996. ACM.

[21] S. K. Card, P. Pirolli, M. Van Der Wege, J. B. Morrison, R. W. Reeder, P. K. Schraedley, and J. Boshart. Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 498–505, New York, NY, USA, 2001. ACM.

[22] A. Carlier, V. Charvillat, W. T. Ooi, R. Grigoras, and G. Morin. Crowdsourced automatic zoom and scroll for video retargeting. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 201–210, New York, NY, USA, 2010. ACM.

[23] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN Systems*, 27(6): 1065–1073, Apr. 1995.

[24] O. Chapuis, J.-B. Labrune, and E. Pietriga. Dynaspot: Speed-dependent area cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1391–1400, New York, NY, USA, 2009. ACM.

[25] K.-Y. Cheng, S.-J. Luo, B.-Y. Chen, and H.-H. Chu. Smartplayer: User-centric video fast-forwarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 789–798, New York, NY, USA, 2009. ACM.

[26] X. Cheng, C. Dale, and J. Liu. Understanding the characteristics of internet short video sharing: Youtube as a case study. In *Technical Report arXiv:0707.3670v1 cs.NI*, New York, NY, USA, 2007. Cornell University, arXiv e-prints.

[27] J. P. Chin, V. A. Diehl, and K. L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '88, pages 213–218, New York, NY, USA, 1988. ACM.

[28] K. Chorianopoulos. Collective intelligence within web video. *Human-centric Computing and Information Sciences*, 3(1):10, 2013.

[29] K. Chorianopoulos and M. N. Giannakos. Merging learner performance with browsing behavior in video lectures. In *Proceedings of the LAK 2013 Workshop on Analytics on Video-based Learning*, volume 983 of *WAVe 2013*, pages 37–41. CEUR-WS, 2013.

[30] K. Chorianopoulos, M. N. Giannakos, and N. Chrisochoides. Open system for video learning analytics. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 153–154, New York, NY, USA, 2014. ACM.

[31] M. Christel and N. Moraveji. Finding the right shots: Assessing usability and performance of a digital video library interface. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, MULTIMEDIA '04, pages 732–739, New York, NY, USA, 2004. ACM.

[32] A. Cockburn and A. Firth. Improving the acquisition of small targets. In E. ONeill, P. Palanque, and P. Johnson, editors, *People and Computers XVII Designing for Society*, pages 181–196. Springer London, 2004.

[33] A. Cockburn and B. Mckenzie. What do web users do? an empirical analysis of web use. *International Journal of Human-Computer Studies*, 54 (6):903–922, June 2001.

[34] A. Cockburn, C. Gutwin, and J. Alexander. Faster document navigation with space-filling thumbnails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1–10, New York, NY, USA, 2006. ACM.

[35] B. Craggs, M. Kilgallon Scott, and J. Alexander. Thumbreels: Query sensitive web video previews based on temporal, crowdsourced, semantic tagging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1217–1220, New York, NY, USA, 2014. ACM.

[36] R. Crane and D. Sornette. Viral, quality, and junk videos on youtube: Separating content from noise in an information-rich environment. In *Proceedings of AAAI symposium on Social Information Processing*, pages 18–20, Menlo Park, California, USA, 2008. The AAAI Press.

[37] P. Dragicevic, G. Ramos, J. Bibliowitcz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh. Video browsing by direct manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 237–246, New York, NY, USA, 2008. ACM.

[38] S. M. Drucker, A. Glatzer, S. De Mar, and C. Wong. Smartskip: Consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 219–226, New York, NY, USA, 2002. ACM.

[39] P. M. Fitts. The information capacity of the human motor system in controllong the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[40] M. Fong, A. Al Hajri, G. Miller, and S. Fels. Casual authoring using a video navigation history. In *Proceedings of the 2014 Graphics Interface Conference*, GI '14, pages 109–114, Toronto, Ont., Canada, 2014. Canadian Information Processing Society.

[41] E. Frécon and G. Smith. Webpath - a three-dimensional web history. In *Proceedings of the 1998 IEEE Symposium on Information Visualization*, INFOVIS '98, pages 3–10, Washington, DC, USA, 1998. IEEE Computer Society.

[42] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: A view from the edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, pages 15–28, New York, NY, USA, 2007. ACM.

[43] A. Girgensohn, L. Wilcox, F. Shipman, and S. Bly. Designing affordances for the navigation of detail-on-demand hypervideo. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 290–297, New York, NY, USA, 2004. ACM.

[44] C. Gkonela and K. Chorianopoulos. Videoskip: Event detection in social web videos with an implicit user heuristic. *Multimedia Tools and Applications*, 69(2):383–396, Mar. 2014.

[45] D. B. Goldman, C. Gonterman, B. Curless, D. Salesin, and S. M. Seitz. Video object annotation, navigation, and composition. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 3–12, New York, NY, USA, 2008. ACM.

[46] V. Gopalakrishnan, R. Jana, K. K. Ramakrishnan, D. F. Swayne, and V. A. Vaishampayan. Understanding couch potatoes: Measurement and modeling of interactive usage of iptv at large scale. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, pages 225–242, New York, NY, USA, 2011. ACM.

[47] T. Grossman and R. Balakrishnan. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 281–290, New York, NY, USA, 2005. ACM.

[48] T. Grossman and R. Balakrishnan. A probabilistic approach to modeling two-dimensional pointing. *ACM Trans. Comput.-Hum. Interact.*, 12(3): 435–459, Sept. 2005.

[49] T. Grossman, J. Matejka, and G. Fitzmaurice. Chronicle: Capture, exploration, and playback of document workflow histories. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 143–152, New York, NY, USA, 2010. ACM.

[50] Y. Guiard, R. Blanch, and M. Beaudouin-Lafon. Object pointing: A complement to bitmap pointing in guis. In *Proceedings of Graphics Interface 2004*, GI '04, pages 9–16, School of Computer Science,

University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.

[51] F. Guimbretiére, M. Dixon, and K. Hinckley. Experiscope: An analysis tool for interaction data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1333–1342, New York, NY, USA, 2007. ACM.

[52] T. J. Gunn, P. Irani, and J. Anderson. An evaluation of techniques for selecting moving targets. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 3329–3334, New York, NY, USA, 2009. ACM.

[53] K. Hasan, T. Grossman, and P. Irani. Comet and target ghost: Techniques for selecting moving targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 839–848, New York, NY, USA, 2011. ACM.

[54] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, Nov 2008.

[55] W. R. Hendee and P. N. T. Wells, editors. *The Perception of Visual Information*. Springer New York, 1997.

[56] W. C. Hill, J. D. Hollan, D. Wroblewski, and T. McCandless. Edit wear and read wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 3–9, New York, NY, USA, 1992. ACM.

[57] G. Hodgkinson. *Visually Representative Web History Browser*. PhD thesis, Massey University, Wellington, may 2007.

[58] E. R. HOFFMANN. Capture of moving targets: a modification of fitts' law. *Ergonomics*, 34(2):211–220, 1991.

[59] C. Huang, J. Li, and K. W. Ross. Can internet video-on-demand be profitable? In *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '07, pages 133–144, New York, NY, USA, 2007. ACM.

[60] D. Hupp and R. C. Miller. Smart bookmarks: automatic retroactive macro recording on the web. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, UIST '07, pages 81–90, New York, NY, USA, 2007. ACM.

[61] K. W. Hwang, D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, V. Misra, K. K. Ramakrishnan, and D. F. Swayne. Leveraging video viewing patterns for optimal content placement. In *Proceedings of the 11th International IFIP TC 6 Conference on Networking - Volume Part II*, IFIP'12, pages 44–58, Berlin, Heidelberg, 2012. Springer-Verlag.

[62] M. V. Ilich. Moving target selection in interactive video. Master's thesis, University of British Columbia, Vancouver, Canada, December 2009.

[63] D. Jackson, J. Nicholson, G. Stoeckigt, R. Wrobel, A. Thieme, and P. Olivier. Panopticon: A parallel video overview system. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 123–130, New York, NY, USA, 2013. ACM.

[64] R. J. Jagacinski, D. W. Repperger, S. L. Ward, and M. S. Moran. A test of fitts' law with moving targets. *Human Factors*, 22(2):225–233, 1980.

[65] H. Joho, J. M. Jose, R. Valenti, and N. Sebe. Exploiting facial expressions for affective video summarisation. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '09, pages 31:1–31:8, New York, NY, USA, 2009. ACM.

[66] P. Kabbash and W. A. S. Buxton. The “prince” technique: Fitts' law and selection using area cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 273–279, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[67] T. Karrer, M. Weiss, E. Lee, and J. Borchers. Dragon: A direct manipulation interface for frame-accurate in-scene video navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 247–250, New York, NY, USA, 2008. ACM.

[68] G. H. Khaksari. Direct manipulation of web browsing history. *International Journal of Computer Science and Information Technology*, 3 (5):185–192, October 2011.

[69] J. Kim, P. J. Guo, C. J. Cai, S.-w. D. Li, K. Z. Gajos, and R. C. Miller. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, New York, NY, USA, 2014. ACM.

[70] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 31–40, New York, NY, USA, 2014. ACM.

[71] D. Kimber, T. Dunnigan, A. Girgensohn, F. Shipman, T. Turner, and T. Yang. Trailblazing: Video playback control by direct object manipulation. In *Proceedings of 2007 IEEE International Conference on Multimedia and Expo*, pages 1015 –1018, july 2007.

[72] S. S. Krishnan and R. K. Sitaraman. Understanding the effectiveness of video ads: A measurement study. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 149–162, New York, NY, USA, 2013. ACM.

[73] I. Leftheriotis, C. Gkonela, and K. Chorianopoulos. Efficient video indexing on the web: A system that crowdsources user interactions with a video player. In F. Alvarez and C. Costa, editors, *User Centric Media*, volume 60 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 123–131. Springer Berlin Heidelberg, 2012.

[74] F. C. Li, A. Gupta, E. Sanocki, L.-w. He, and Y. Rui. Browsing digital video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 169–176, New York, NY, USA, 2000. ACM.

[75] I. Li, J. Nichols, T. Lau, C. Drews, and A. Cypher. Here's what i did: sharing and reusing web activity with actionshot. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 723–732, New York, NY, USA, 2010. ACM.

[76] I. S. MacKenzie and W. Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 219–226, New York, NY, USA, 1992. ACM.

[77] R. Martin and H. Holtzman. Newstream: A multi-device, cross-medium, and socially aware approach to news content. In *Proceedings of the 8th International Interactive Conference on Interactive TV&Video*, EuroITV '10, pages 83–90, New York, NY, USA, 2010. ACM.

[78] T. Masui, K. Kashiwagi, and G. R. Borden, IV. Elastic graphical interfaces to precise data manipulation. In *Conference Companion on Human Factors in Computing Systems*, CHI '95, pages 143–144, New York, NY, USA, 1995. ACM.

[79] J. Matejka, T. Grossman, and G. Fitzmaurice. Swifter: Improved online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1159–1168, New York, NY, USA, 2013. ACM.

[80] M. Mayer. *Visualizing Web Sessions: Improving Web Browser History by a Better Understanding of Web Page Revisitation and a New Session- and Task-Based, Visual Web History Approach*. PhD thesis, Universitt Hamburg, 2007.

[81] M. McGuffin and R. Balakrishnan. Acquisition of expanding targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 57–64, New York, NY, USA, 2002. ACM.

[82] R. Mertens, R. Farzan, and P. Brusilovsky. Social navigation in web lectures. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, HYPERTEXT '06, pages 41–44, New York, NY, USA, 2006. ACM.

[83] N. Milic-Frayling, R. Sommerer, and K. Rodden. Webscout: support for revisitation of web pages within a navigation session. In *Proceedings of IEEE/WIC International Conference on Web Intelligence*, WI '03, pages 689 – 693, oct. 2003.

[84] G. Miller, S. Fels, M. Finke, W. Motz, W. Eagleston, and C. Eagleston. Minidiver: A novel mobile media playback interface for rich video content on an iphonetm. In *Proceedings of the 8th International Conference on Entertainment Computing*, ICEC '09, pages 98–109, Berlin, Heidelberg, 2009. Springer-Verlag.

[85] G. Miller, S. Fels, A. Al Hajri, M. Ilich, Z. Foley-Fisher, M. Fernandez, and D. Jang. Mediadiver: Viewing and annotating multi-view video. In *CHI*

'11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, pages 1141–1146, New York, NY, USA, 2011. ACM.

[86] A. G. Money and H. Agius. Video summarisation: A conceptual framework and survey of the state of the art. *Journal of Visual Communication and Image Representation*, 19(2):121–143, Feb 2008.

[87] A. G. Money and H. Agius. Analysing user physiological responses for affective video summarisation. *Displays*, 30(2):59 – 70, 2009.

[88] T.-J. K. P. Monserrat, S. Zhao, K. McGee, and A. V. Pandey. Notevideo: Facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1139–1148, New York, NY, USA, 2013. ACM.

[89] D. Mould and C. Gutwin. The effects of feedback on targeting with multiple moving targets. In *Proceedings of Graphics Interface 2004*, GI '04, pages 25–32, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.

[90] T. Nadeem and B. Killam. A study of three browser history mechanisms for web navigation. In *Proceedings of the Fifth International Conference on Information Visualisation*, pages 13–21, 2001.

[91] T. Nakamura and T. Igarashi. An application-independent system for visualizing user operation history. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 23–32, New York, NY, USA, 2008. ACM.

[92] J. Nicholson, M. Huber, D. Jackson, and P. Olivier. Panopticon as an elearning support search tool. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1221–1224, New York, NY, USA, 2014. ACM.

[93] D. A. Norman. *The Design of Everyday Things*. Basic Books, Inc., New York, NY, USA, 2002. ISBN 9780465067107.

[94] D. R. Olsen and B. Moon. Video summarization based on user interaction. In *Proceddings of the 9th International Interactive Conference on Interactive Television*, EuroITV '11, pages 115–122, New York, NY, USA, 2011. ACM.

[95] A. Pavel, C. Reed, B. Hartmann, and M. Agrawala. Video digests: A browsable, skimmable format for informational lecture videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 573–582, New York, NY, USA, 2014. ACM.

[96] W.-T. Peng, W.-T. Chu, C.-H. Chang, C.-N. Chou, W.-J. Huang, W.-Y. Chang, and Y.-P. Hung. Editing by viewing: Automatic home video summarization by viewing behavior analysis. *Multimedia, IEEE Transactions on*, 13(3):539–550, June 2011.

[97] G. Ramos and R. Balakrishnan. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pages 105–114, New York, NY, USA, 2003. ACM.

[98] E. Risko, T. Foulsham, S. Dawson, and A. Kingstone. The collaborative lecture annotation system (clas): A new tool for distributed learning. *Learning Technologies, IEEE Transactions on*, 6(1):4–13, Jan 2013.

[99] D. A. Rosenbaum. *Human Motor Control*. Academic Press, 1991. URL http://books.google.ca/books?id=w3kVQBqSlEQC.

[100] C. A. Russell and S. J. Levy. The temporal and focal dynamics of volitional reconsumption: A phenomenological investigation of repeated hedonic experiences. *Journal of Consumer Research*, 39(2):341–359, August 2012.

[101] N. Sawhney, D. Balcom, and I. Smith. Hypercafe: Narrative and aesthetic properties of hypervideo. In *Proceedings of the the Seventh ACM Conference on Hypertext*, HYPERTEXT '96, pages 1–10, New York, NY, USA, 1996. ACM.

[102] N. Sawhney, D. Balcom, and I. Smith. Authoring and navigating video in space and time. *IEEE MultiMedia*, 4(4):30–39, Oct. 1997.

[103] R. A. Schmidt, H. N. Zelaznik, B. Hawkins, J. S. Frank, and J. T. J. Quinn. Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review*, 86(5):415–451, Sept 1979.

[104] D. Shamma, J. Yew, L. Kennedy, and E. Churchill. Viral actions: Predicting video view counts using synchronous sharing behaviors. In *Proceedings of the fifth International AAAI Conference on Weblogs and Social Media*, ICWSM '11, Menlo Park, California, USA, 2011. The AAAI Press.

[105] D. A. Shamma, R. Shaw, P. L. Shafton, and Y. Liu. Watch what i watch: Using community activity to understand content. In *Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval*, MIR '07, pages 275–284, New York, NY, USA, 2007. ACM.

[106] D. A. Shamma, L. Kennedy, and E. F. Churchill. Tweet the debates: Understanding community annotation of uncollected sources. In *Proceedings of the First SIGMM Workshop on Social Media*, WSM '09, pages 3–10, New York, NY, USA, 2009. ACM.

[107] R. Shaw and M. Davis. Toward emergent representations for video. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, pages 431–434, New York, NY, USA, 2005. ACM.

[108] R. Shaw and P. Schmitz. Community annotation and remix: A research platform and pilot deployment. In *Proceedings of the 1st ACM International Workshop on Human-centered Multimedia*, HCM '06, pages 89–98, New York, NY, USA, 2006. ACM.

[109] I. Sheikh and E. Hoffmann. Effect of target shape on movement time in a fitts task. *Ergonomics*, 37(9):1533–1548, 1994.

[110] R. J. Sundstrom. Method and system for automatically linking a cursor to a hotspot in a hypervideo stream, Jan 2009.

[111] T. Syeda-Mahmood and D. Ponceleon. Learning video browsing behavior and its application in the generation of video previews. In *Proceedings of the Ninth ACM International Conference on Multimedia*, MULTIMEDIA '01, pages 119–128, New York, NY, USA, 2001. ACM.

[112] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, Aug. 2010.

[113] S. Teelink and R. F. Erbacher. Improving the computer forensic analysis process through visualization. *Communications of the ACM - Next-generation cyber forensics*, 49(2):71–75, Feb. 2006.

[114] B. T. Truong and S. Venkatesh. Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 3(1), Feb. 2007.

[115] E. Vartiainen, V. Roto, and J. Kaasalainen. Graphical history list with multi-window support on a mobile web browser. In *Proceedings of the Third International Conference on Internet and Web Applications and Services*, ICIW '08, pages 121 –129, june 2008.

[116] T. G. Whisenand and H. H. Emurian. Some effects of angle of approach on icon selection. In *Conference Companion on Human Factors in Computing Systems*, CHI '95, pages 298–299, New York, NY, USA, 1995. ACM.

[117] J. O. Wobbrock, J. Fogarty, S.-Y. S. Liu, S. Kimuro, and S. Harada. The angle mouse: Target-agnostic dynamic gain adjustment based on angular deviation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1401–1410, New York, NY, USA, 2009. ACM.

[118] A. Worden, N. Walker, K. Bharat, and S. Hudson. Making computers easier for older adults to use: Area cursors and sticky icons. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 266–271, New York, NY, USA, 1997. ACM.

[119] C. E. Wright and D. E. Meyer. Conditions for a linear speed-accuracy trade-off in aimed movements. *Quarterly Journal of Experimental Psychology*, 35(A):279–296, 1983.

[120] T. Yamaguchi, H. Hattori, T. Ito, and T. Shintani. On a web browsing support system with 3d visualization. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, WWW Alt. '04, pages 316–317, New York, NY, USA, 2004. ACM.

[121] M. M. Yeung and B.-L. Yeo. Video visualization for compact presentation and fast browsing of pictorial content. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(5):771–785, Oct. 1997.

[122] J. Yew and D. A. Shamma. Know your data: Understanding implicit usage versus explicit action in video content classification. In *Proceedings of SPIE, Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V*, volume 7881, 2011.

[123] J. Yew, D. A. Shamma, and E. F. Churchill. Knowing funny: Genre perception and categorization in social video sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 297–306, New York, NY, USA, 2011. ACM.

[124] H. Yin, X. Liu, F. Qiu, N. Xia, C. Lin, H. Zhang, V. Sekar, and G. Min. Inside the bird's nest: Measurements of large-scale live vod from the 2008 olympics. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '09, pages 442–455, New York, NY, USA, 2009. ACM.

[125] B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang. Video summarization based on user log enhanced link analysis. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, MULTIMEDIA '03, pages 382–391, New York, NY, USA, 2003. ACM.

[126] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding user behavior in large-scale video-on-demand systems. In *Proceedings of the 1st ACM SIGOPS/EuroSys European Conference on Computer Systems 2006*, EuroSys '06, pages 333–344, New York, NY, USA, 2006. ACM.

[127] S. Zhai, S. Conversy, M. Beaudouin-Lafon, and Y. Guiard. Human on-line response to target expansion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 177–184, New York, NY, USA, 2003. ACM.

Appendix A

List of Publications

Many of the contributions and ideas described in this dissertation have been previously presented in publications and oral presentations, which are listed here.

A.1 Conference Publications

Abir Al-Hajri, Matthew Fong, Gregor Miller, and Sidney S. Fels. How personal video navigation history can be visualized. In *Proceedings of the 41st International Conference and Exhibition on Computer Graphics and Interactive Techniques; Posters*, SIGGRAPH 14, pages 105:1, New York, NY, USA, August 2014. ACM

Abir Al-Hajri, Matthew Fong, Gregor Miller, and Sidney Fels. Fast forward with your VCR: Visualizing single-video viewing statistics for navigation and sharing. In *Proceedings of the 2014 Graphics Interface Conference*, GI 14, pages 123-128, Toronto, Ont., Canada, 2014. Canadian Information Processing Society.

Abir Al-Hajri, Gregor Miller, Matthew Fong, and Sidney S. Fels. Visualization of personal history for video navigation. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1187-1196, New York, NY, USA, 2014. ACM

Abir Al-Hajri, Gregor Miller, Sidney Fels, and Matthew Fong. Video navi-

gation with a personal viewing history. In *Proceedings of the 14th IFIP TC 13 International Conference on Human-Computer Interaction - Part III*, volume 8119 of INTERACT '13, pages 352-369, Berlin, Heidelberg, September 2-6 2013. Springer-Verlag

Abir Al Hajri, Sidney Fels, Gregor Miller, and Michael Ilich. Moving target selection in 2D graphical user interfaces. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Part II*, volume 6947 of INTERACT '11, pages 141-161, Berlin, Heidelberg, September 5-9 2011. Springer-Verlag

Gregor Miller, Sidney Fels, **Abir Al Hajri**, Michael Ilich, Zoltan Foley-Fisher, Manuel Fernandez, and Daesik Jang. Mediadiver: Viewing and annotating multi-view video. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1141-1146, New York, NY, USA, 2011. ACM

A.2 Interactive Demonstrations

Demonstrating *Mevie* (Chapter 8) and let attendees play with it

- August 2014 - at posters session in SIGGRAPH 2014, Vancouver, BC, Canada
- May 2014 - at GRAND Experiences in GRAND 2014, Ottawa, ON, Canada
- April 2014 - at Interactivity session in CHI 2014, Toronto, ON, Canada

Demonstrating our interface (Chapters 5 and 6) and let attendees play with it

- May 2014 - at GRAND Experiences in GRAND 2014, Ottawa, ON, Canada
- April 2014 - at Interactivity session in CHI 2014, Toronto, ON, Canada

Demonstrating *MediaDiver* (Section 7.4) and let attendees play with it

- MAY 2011 - at Interactivity session in CHI 2011, Vancouver, BC, Canada

A.3 Research Talks

May 2014 - A Comparative Evaluation of Methods for Moving Target Selection

- GRAND 2014, Ottawa, ON, Canada

May 2014 - Casual authoring using a video navigation history

- GRAND 2014, Ottawa, ON, Canada

A.4 Additional Publications

Matthew Fong, **Abir Al Hajri**, Gregor Miller, and Sidney Fels. *Casual authoring using a video navigation history*. In Proceedings of the 2014 Graphics Interface Conference, GI '14, pages 109-114, Toronto, Ont., Canada, 2014. Canadian Information Processing Society

Noreen Kamal, **Abir Al Hajri**, and Sidney Fels. DreamThrower: An audio/visual display for influencing dreams. *Entertainment Computing*, 3(4):121-128, 2012.

Noreen Kamal, Ling Tsou, **Abir Al Hajri**, and Sidney Fels. DreamThrower: Creating, throwing and catching dreams for collaborative dream sharing. In *Proceedings of the 9th International Conference on Entertainment Computing*, ICEC '10, pages 20-31, Berlin, Heidelberg, 2010. Springer-Verlag

Appendix B

User Studies Questionnaire

B.1 User Study on Video Viewing Behaviour

VIDEO VIEWING BEHAVIOUR

Thank you for participating in this user study.

Please fill the form (these details are required for research purposes only and will not be disclosed)

1. Gender:
☐ Male ☐ Female

2. Age:
☐ 19 - 25 ☐ 26 - 30 ☐ 31 - 40 ☐ 41 - 50 ☐ 51 - 60 ☐ 61 or over

3. How often do you watch videos online?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

4. How many videos do you watch on average per session?
☐ None ☐ 1 - 3 videos ☐ 4 - 6 videos ☐ 7 - 10 videos ☐ more than 10 videos

B.2 User Study on the Feasibility of Video Viewing History

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

VIDEO NAVIGATION **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

PERSONAL DETAILS

(these details are required for research purposes only and will not be disclosed)

1. Gender:
☐ Male ☐ Female

2. Age:
☐ 19 - 25 ☐ 26 - 30 ☐ 31 - 40 ☐ 41 - 50 ☐ 51 - 60 ☐ 61 or over

3. Are you an experienced computer user?
☐ Yes ☐ No

4. How often do you watch videos?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

5. Have you used any video editing software (e.g. iMovie, MovieMaker. etc.)?
☐ Yes ☐ No

6. If yes, which software have you used? _____

7. How often do you use it?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

INTERFACE FEATURES

A. For each feature, please rate how **EASY** did you find it to use the feature? "I found it easy to use" (i.e. 1: strongly disagree, . . . 7: strongly agree).

Component	**Feature**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
• Video Player		☐	☐	☐	☐	☐	☐	☐
	• Play/Pause button	☐	☐	☐	☐	☐	☐	☐
	•Seek	☐	☐	☐	☐	☐	☐	☐
	• Frame Preview at the seek bar	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• Grid of Videos		☐	☐	☐	☐	☐	☐	☐
	• Select video to play	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• History Timeline		☐	☐	☐	☐	☐	☐	☐
	• Select video to play using history	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment using thumb-nails	☐	☐	☐	☐	☐	☐	☐
	• Different colors for videos	☐	☐	☐	☐	☐	☐	☐
	• Different rectangle for each interval	☐	☐	☐	☐	☐	☐	☐
	• Popularity bar for each thumbnail	☐	☐	☐	☐	☐	☐	☐
	• Go back to previous history	☐	☐	☐	☐	☐	☐	☐
	• Understand your history	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• Video Mashup		☐	☐	☐	☐	☐	☐	☐
	• Add clip	☐	☐	☐	☐	☐	☐	☐
	• Remove video	☐	☐	☐	☐	☐	☐	☐
	• Modify clip timing	☐	☐	☐	☐	☐	☐	☐
	• Reorder videos	☐	☐	☐	☐	☐	☐	☐
	• Play clip	☐	☐	☐	☐	☐	☐	☐
	• Remove clip	☐	☐	☐	☐	☐	☐	☐
• Trailer Preview		☐	☐	☐	☐	☐	☐	☐
	• Preview trailer before exporting	☐	☐	☐	☐	☐	☐	☐
	• Replay trailer	☐	☐	☐	☐	☐	☐	☐
	• Modify trailer	☐	☐	☐	☐	☐	☐	☐

B. For each feature, please rate how **USEFUL** did you find it to use the feature? "I think it would be useful" (i.e. 1: strongly disagree, . . . 7: strongly agree).

Component	**Feature**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
• Video Player		☐	☐	☐	☐	☐	☐	☐
	• Play/Pause button	☐	☐	☐	☐	☐	☐	☐
	•Seek	☐	☐	☐	☐	☐	☐	☐
	• Frame Preview at the seek bar	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• Grid of Videos		☐	☐	☐	☐	☐	☐	☐
	• Select video to play	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• History Timeline		☐	☐	☐	☐	☐	☐	☐
	• Select video to play using history	☐	☐	☐	☐	☐	☐	☐
	• Play specific segment using thumb-nails	☐	☐	☐	☐	☐	☐	☐
	• Different colors for videos	☐	☐	☐	☐	☐	☐	☐
	• Different rectangle for each interval	☐	☐	☐	☐	☐	☐	☐
	• Popularity bar for each thumbnail	☐	☐	☐	☐	☐	☐	☐
	• Go back to previous history	☐	☐	☐	☐	☐	☐	☐
	• Understand your history	☐	☐	☐	☐	☐	☐	☐
	• Create Trailer	☐	☐	☐	☐	☐	☐	☐
• Video Mashup		☐	☐	☐	☐	☐	☐	☐
	• Add clip	☐	☐	☐	☐	☐	☐	☐
	• Remove video	☐	☐	☐	☐	☐	☐	☐
	• Modify clip timing	☐	☐	☐	☐	☐	☐	☐
	• Reorder videos	☐	☐	☐	☐	☐	☐	☐
	• Play clip	☐	☐	☐	☐	☐	☐	☐
	• Remove clip	☐	☐	☐	☐	☐	☐	☐
• Trailer Preview		☐	☐	☐	☐	☐	☐	☐
	• Preview trailer before exporting	☐	☐	☐	☐	☐	☐	☐
	• Replay trailer	☐	☐	☐	☐	☐	☐	☐
	• Modify trailer	☐	☐	☐	☐	☐	☐	☐

C. Please rank the following components for trailer creation on each of the conditions? (i.e. 1:best, 2: Second, ..)

	Grid of Videos	History Timeline	Hybrid
Easy			
Fast			

D. Please provide us with your comments and suggestions on each of the following components:

	Comments	Suggestions
Video Player		
Grid of Videos		
History Timeline		
Video Mashup		
Trailer Preview		

E. Any other general comments or suggestions about the study.

B.3 List of Thumbnails User Study

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

SINGLE-VIDEO HISTORY VISUALIZATION **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

Part 1: Demographic Description

1. Gender: ☐ Male ☐ Female

2. Age:
☐ 19 - 25 ☐ 26 - 30 ☐ 31 - 40 ☐ 41 - 50 ☐ 51 - 60 ☐ 61 or over

3. Your major:
☐ Mathematics / quantitative ☐ Arts / humanities ☐ Computer science
☐ Natural sciences / medicine ☐ Architecture / design ☐ Engineering
☐ Social / behavioral sciences ☐ Business / management ☐ Education
☐ Other (specify) ——————————

4. Are you an experienced computer user?
☐ Yes ☐ No

5. How often do you watch videos?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

Part 2: Type of System to be rated

6. Name of Navigation Method: ☐ The Filmstrip ☐ The User Timeline
7. Length of time you have worked on the interface using this method ☐ Less than 30 min ☐ 30 min to one hour

Part 3: User Evaluation of the Interface				
Please check the numbers, which most appropriately reflect your impressions about using this method. Not Applicable = **NA**. Please add your written comments below the corresponding item.				
OVERALL REACTION TO THE COMPONENT		1 2 3 4 5 6 7		NA
8.	Unimpressive	☐☐☐☐☐☐☐	Impressive	☐
9.	Frustrating	☐☐☐☐☐☐☐	Satisfying	☐
10.	Dull	☐☐☐☐☐☐☐	Stimulating	☐
11.	Difficult	☐☐☐☐☐☐☐	Easy	☐
12.	Ineffective	☐☐☐☐☐☐☐	Powerful	☐
13.	Rigid	☐☐☐☐☐☐☐	Flexible	☐
14.	Hindrance	☐☐☐☐☐☐☐	Helpful	☐
15.	Useless	☐☐☐☐☐☐☐	Useful	☐
LEARNING		1 2 3 4 5 6 7		NA
16. Learning to use the interface	Difficult	☐☐☐☐☐☐☐	Easy	☐
16.1. Getting started	Difficult	☐☐☐☐☐☐☐	Easy	☐
16.2. Learning advanced features	Difficult	☐☐☐☐☐☐☐	Easy	☐
16.3. Time to learn to use the interface	Too long	☐☐☐☐☐☐☐	Just right	☐
16.4. Remembering how to use it.	Difficult	☐☐☐☐☐☐☐	Easy	☐
17. Exploring new features by trial and error	Discouraged	☐☐☐☐☐☐☐	Encouraged	☐
17.1. Exploration of features	Uncomfortable	☐☐☐☐☐☐☐	Enjoyable	☐
17.2. Discovering new features	Difficult	☐☐☐☐☐☐☐	Easy	☐
18. Tasks can be performed in a straight-forward manner	Never	☐☐☐☐☐☐☐	Always	☐
18.1. Number of steps per task	Too many	☐☐☐☐☐☐☐	Just right	☐
18.2. Steps to complete a task follow a logical sequence	Never	☐☐☐☐☐☐☐	Always	☐
18.3. Completion of task	Unclear	☐☐☐☐☐☐☐	Clear	☐
19. I quickly became skillful with it.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
20. I can use it without written instructions.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐

GENERAL IMPRESSIONS		1 2 3 4 5 6 7		NA
21. Screens are aesthetically pleasing	Not at all	☐☐☐☐☐☐☐	Very much	☐
21.1. Screen designs and layout are attractive	Not at all	☐☐☐☐☐☐☐	Very much	☐
21.2. Use of colors	Unattractive	☐☐☐☐☐☐☐	Attractive	☐
21.3. Use of color combinations	Unattractive	☐☐☐☐☐☐☐	Attractive	☐
22. I don't notice any inconsistencies as I use it.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
23. Interface is impressive	Not at all	☐☐☐☐☐☐☐	Very much	☐
23.1. Interface can do a great deal	Not at all	☐☐☐☐☐☐☐	Very much	☐
23.2. Such an interface in a home would be	Useless	☐☐☐☐☐☐☐	Useful	☐
24. Both occasional and regular users would like it.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
25. Interface is fun to use	Not at all	☐☐☐☐☐☐☐	Very much	☐
25.1. Interface maintains ones interest	Never	☐☐☐☐☐☐☐	Always	☐
25.2. Interface would remain interesting	Unlikely	☐☐☐☐☐☐☐	Likely	☐
26. It is user friendly.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
27. Using it is effortless.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
28. I can use it successfully every time.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
SATISFACTION		1 2 3 4 5 6 7		NA
29. I am satisfied with it.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
30. I would recommend it to a friend.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
31. It is fun to use.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐

32. It works the way I want it to work.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
33. It is wonderful.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
34. I feel I need to have it.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
35. It is pleasant to use.	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
MAIN VIDEO PLAYER		1 2 3 4 5 6 7		NA
36. Location of the video player	Inadequate	☐☐☐☐☐☐☐	Adequate	☐
37. Size of the video player	Too small	☐☐☐☐☐☐☐	Large enough	☐
38. Click on the video player to play/pause video	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
39. System response to the click for play/pause	Too slow	☐☐☐☐☐☐☐	Fast enough	☐
40. Size of play/pause button	Too small	☐☐☐☐☐☐☐	Large enough	☐
41. Size of playhead	Too small	☐☐☐☐☐☐☐	Large enough	☐
42. System response to seek on video timeline	Too slow	☐☐☐☐☐☐☐	Fast enough	☐
43. Seeking a video	Difficult	☐☐☐☐☐☐☐	Easy	☐
44. Size of timeline	Too small	☐☐☐☐☐☐☐	Large enough	☐
45. Current video timestamp label	Useless	☐☐☐☐☐☐☐	Useful	☐
46. Size of video timestamp label	Too small	☐☐☐☐☐☐☐	Large enough	☐
VIDEO SEGMENT		1 2 3 4 5 6 7		NA
47. Highlight video segment	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
48. Size of the video segment	Too small	☐☐☐☐☐☐☐	Large enough	☐
49. Playing each video segment	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
50. Size of the video segment navigation area	Too small	☐☐☐☐☐☐☐	Large enough	☐

51. Color of the video segment navigation area	Irritating	☐☐☐☐☐☐☐	Pleasing	☐
52. Size of the navigation seek thumb	Too small	☐☐☐☐☐☐☐	Large enough	☐
53. Navigating video segments	Difficult	☐☐☐☐☐☐☐	Easy	☐
54. System response to the seek	Too slow	☐☐☐☐☐☐☐	Fast enough	☐
55. Drag to play from specific time	Useless	☐☐☐☐☐☐☐	Useful	☐
56. Drag the whole segment to play the video segment	Useless	☐☐☐☐☐☐☐	Useful	☐
HISTORY TIMELINE		1 2 3 4 5 6 7		NA
57. Location of user timeline	Inadequate	☐☐☐☐☐☐☐	Adequate	☐
58. Size of user timeline	Too small	☐☐☐☐☐☐☐	Large enough	☐
59. Having your own history	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
60. History makes tasks Completion	Difficult	☐☐☐☐☐☐☐	Easy	☐
61. Having control over the creation of the segments	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
62. Having variable length intervals	Useless	☐☐☐☐☐☐☐	Useful	☐
63. Having watched intervals only	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
64. Vertical scrolling	Difficult	☐☐☐☐☐☐☐	Easy	☐
65. Playing segment from history	Difficult	☐☐☐☐☐☐☐	Easy	☐
FILMSTRIP		1 2 3 4 5 6 7		NA
66. Location of filmstrip	Inadequate	☐☐☐☐☐☐☐	Adequate	☐
67. Size of filmstrip	Too small	☐☐☐☐☐☐☐	Large enough	☐
68. Systematically created intervals	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐
69. Having fixed length intervals	Useless	☐☐☐☐☐☐☐	Useful	☐
70. Having access to watched and unwatched intervals	Useless	☐☐☐☐☐☐☐	Useful	☐
71. Horizontal scrolling	Difficult	☐☐☐☐☐☐☐	Easy	☐
72. Playing video segment from filmstrip	Difficult	☐☐☐☐☐☐☐	Easy	☐
73. Zoom in the filmstrip	Unhelpful	☐☐☐☐☐☐☐	Helpful	☐

Part 4: OTHER REACTIONS, IMPRESSIONS, AND COMMENTS
74. Comments:

List the most negative aspect(s):
1.
2.
3.

List the most positive aspect(s):
1.
2.
3.

B.4 VCR User Study

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

SINGLE-VIDEO HISTORY VISUALIZATION **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

PERSONAL DETAILS

(these details are required for research purposes only and will not be disclosed)

1. Gender:
☐ Male ☐ Female

2. Age:
☐ 19 - 25 ☐ 26 - 30 ☐ 31 - 40 ☐ 41 - 50 ☐ 51 - 60 ☐ 61 or over

3. Are you an experienced computer user?
☐ Yes ☐ No

4. Your vision is:
☐ Normal ☐ Corrected to normal

5. Do you have colour blindness?
☐ No ☐ Yes. Type: ————————

6. How often do you watch TV/Movies?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

7. How often do you watch videos online (e.g. YouTube, Vimeo, . . . etc)?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

8. How many videos do you watch on average per day?
☐ None ☐ 1 - 3 videos ☐ 4 - 6 videos ☐ 7 - 10 videos ☐ more than 10 videos

9. What video players are you familiar with?
☐ iTunes ☐ MPlayer ☐ QuickTime ☐ Real player ☐ VLC media player
☐ Windows media player ☐ Other: _____

10. Which video players do you often use? _____

INTERFACE FEATURES

A. Please rate each Component/Feature as specified below.

Component/Feature		1 2 3 4 5 6 7	
11. **Overall interface**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
11.1. Learning to use the interface	Difficult	☐☐☐☐☐☐☐	Easy
11.2. Getting started	Difficult	☐☐☐☐☐☐☐	Easy
11.3. Remembering how to use it	Difficult	☐☐☐☐☐☐☐	Easy
12. **Video Player**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
12.1. Play/Pause button	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
12.2. Seek	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13. **Thumbnail**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.1. Understand interval representation	Difficult	☐☐☐☐☐☐☐	Easy
13.2. Preview each thumbnail	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.3. Seek/Navigate each thumbnail	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.4. Play a segment into the main player	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.5. Play from a specific event or video frame into the main player	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful

13.6. Play/Pause overlay	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.7. Find a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.8. Use of size for segment importance	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14. **Filmstrip**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.1. Understand what it represents	Difficult	☐☐☐☐☐☐☐	Easy
14.2. Entire video representation	Useless	☐☐☐☐☐☐☐	Useful
14.3. Play specific segment	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.4. Seek to a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.5. Preview each segment	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.6. Navigate the entire video	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.7. Zoom in/out	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.8. Find a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15. **View Count Record (VCR)**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15.1. Understand a single video history	Difficult	☐☐☐☐☐☐☐	Easy
15.2. Understand your footprints on a video	Difficult	☐☐☐☐☐☐☐	Easy
15.3. footprints Representation on a video	Useless	☐☐☐☐☐☐☐	Useful
15.4. Find mostly viewed segments	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15.5. Select a specific event to start playing	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15.6. Navigate to the mostly viewed segments	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful

B. Please order the two components: Filmstrip, View Count Record (VCR), for finding previously seen events in terms of ease, less time needed, and preference. For example, start from the easiest one and end with the least easy component in finding an event.
Easy: _____
Fast: _____
Preference: _____

C. Please provide us with your comments and suggestions on each of the following components:

	Comments	Suggestions
Video Player		
Thumbnail		
Filmstrip		
View Count Record (VCR)		

D. Any other general comments or suggestions about the study and/or interface.

B.5 Multiple Video Viewing History Visualization User Study

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

MULTIPLE-VIDEO HISTORY VISUALIZATION **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this user study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

PERSONAL DETAILS

(these details are required for research purposes only and will not be disclosed)

1. Gender: ☐ Male ☐ Female

2. Age:
☐ 19 - 25 ☐ 26 - 30 ☐ 31 - 40 ☐ 41 - 50 ☐ 51 - 60 ☐ 61 or over

3. Are you an experienced computer user?
☐ Yes ☐ No

4. Your vision is:
☐ Normal ☐ Corrected to normal

5. Do you have colour blindness?
☐ No ☐ Yes. Type: ————

6. How often do you watch TV/Movies?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

7. How often do you watch videos online (e.g. YouTube, Vimeo, etc)?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

8. How many videos do you watch on average per day?
☐ None ☐ 1 - 3 videos ☐ 4 - 6 videos ☐ 7 - 10 videos ☐ more than 10 videos

9. What video players are you familiar with?
☐ iTunes ☐ MPlayer ☐ QuickTime ☐ Real player ☐ VLC media player
☐ Windows media player ☐ Other: _____

10. Which video players do you often use? _____

INTERFACE FEATURES

A. Please rate each Component/Feature as specified below.

Component/Feature		1 2 3 4 5 6 7	
11. **Overall interface**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
11.1. Learning to use the interface	Difficult	☐☐☐☐☐☐☐	Easy
11.2. Getting started	Difficult	☐☐☐☐☐☐☐	Easy
11.3. Remembering how to use it	Difficult	☐☐☐☐☐☐☐	Easy
12. **Video Player**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
12.1. Play/Pause button	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
12.2. Seek	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13. **Thumbnail**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.1. Understand interval representation	Difficult	☐☐☐☐☐☐☐	Easy
13.2. Preview each thumbnail	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.3. Seek/Navigate each thumbnail	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.4. Play a segment into the main player	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.5. Play from a specific event or video frame into the main player	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful

13.6. Play/Pause overlay	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.7. Find a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.8. Use of size for segment importance	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.9. Hide a history element	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.10. Show all hidden history elements	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.11. Favourite a history element	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
13.12. Unfavourite a history element	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14. **Filmstrip**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.1. Understand what it represents	Difficult	☐☐☐☐☐☐☐	Easy
14.2. Entire video representation	Useless	☐☐☐☐☐☐☐	Useful
14.3. Play specific segment	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.4. Seek to a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.5. Preview each segment	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.6. Navigate the entire video	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.7. Zoom in/out	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
14.8. Find a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15. **Inter History**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
15.1. Understand your multi-videos history	Difficult	☐☐☐☐☐☐☐	Easy
15.2. Understand a single video history representation	Difficult	☐☐☐☐☐☐☐	Easy

15.3. Change Visualization layout	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
15.4. Sort history elements	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
15.5. Access the detailed history	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
15.6. Scroll the history	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
15.7. Find a specific video history	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
15.8. Select a specific event to start playing	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16. **Intra History**	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.1. Understand a detailed single video history	Difficult	□□□□□□□	Easy
16.2. Understand a single history element representation	Difficult	□□□□□□□	Easy
16.3. Find a specific history element	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.4. Find a previously seen event	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.5. Select a specific event to start playing	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.6. Sort history elements	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.7. Change visualization layout	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.8. Filter history based on view count	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
16.9. Filter history based on favourited items	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
17. **Video Tiles Visualization**	Difficult	□□□□□□□	Easy
	Useless	□□□□□□□	Useful
17.1. Understand the visualization layout	Difficult	□□□□□□□	Easy

17.2. Find specific history element	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
17.3. Different sizes for history elements	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
17.4. Navigate the history	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
18. **Video Timeline Visualization**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
18.1. Understand the visualization layout	Difficult	☐☐☐☐☐☐☐	Easy
18.2. Find specific history element	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
18.3. Different sizes for history elements	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
18.4. Navigate the history	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
19. **Videos Library**	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
19.1. Select a video to play	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful
19.2. Play from a specific event	Difficult	☐☐☐☐☐☐☐	Easy
	Useless	☐☐☐☐☐☐☐	Useful

B. Please order the following components: Filmstrip, Video Timeline, and Video Tiles, for finding previously seen events in terms of ease, less time needed, and preference. For example, start from the easiest one and end with the least easy component in finding an event.

Easy: _____

Fast: _____

Preference: _____

C. Please provide us with your comments and suggestions on each of the following components:

	Comments	Suggestions
Videos Library		
Video Player		
Thumbnail		
Filmstrip		
Inter History		
Intra History		
Video Timeline		
Video Tiles		

D. Any other general comments or suggestions about the study and/or interface.

B.6 Object Selection User Study

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

VIDEO NAVIGATION & HISTORY VISUALIZATIONS **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this user study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

PERSONAL DETAILS

(these details are required for research purposes only and will not be disclosed)

1. Gender: ☐ Male ☐ Female

2. Age:
☐ 18-21 ☐ 22-25 ☐ 26-30 ☐ 31-40 ☐ 41-50 ☐ 51-60 ☐ 61 or over

3. You are: ☐ Right handed user ☐ Left handed user

4. Are you an experienced computer user? ☐ Yes ☐ No

5. How often do you use computers?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

6. How often do you play computer games?
☐ Daily ☐ 3 - 5 times a week ☐ Once a week ☐ Rarely ☐ Never

7. Your vision is:
☐ Normal ☐ Corrected to normal ☐ Other. Specify:————

8. Do you have colour blindness? ☐ No ☐ Yes. Type: ————

USER'S FEEDBACK

A. Please rate how closely you agree with each of the following statements

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
9. I caught wisps faster with the blue potion than with the red.	☐	☐	☐	☐	☐
10. I made fewer mistakes with the blue potion than with the red.	☐	☐	☐	☐	☐
11. I preferred using the blue potion for smaller wisps.	☐	☐	☐	☐	☐
12. I preferred using the blue potion for faster wisps.	☐	☐	☐	☐	☐
13. I preferred using the red potion for larger wisps.	☐	☐	☐	☐	☐
14. The faster wisps were too difficult to capture using the red potion.	☐	☐	☐	☐	☐
15. The red potion made catching wisps moving towards the potion easier than the blue.	☐	☐	☐	☐	☐
16. The green potion made catching wisps easier than blue.	☐	☐	☐	☐	☐
17. With the green potion, I chose to chase wisps before freezing.	☐	☐	☐	☐	☐
18. With the green potion, I chose to pause before catching the wisp.	☐	☐	☐	☐	☐

B. Please rank the potions according to which makes it faster to catch the wisp in each condition. If you have any comments please write them down in the comments column.

	Red	Blue	Green	Comments
Smaller wisps				
Larger wisps				
Faster wisps				
Slower wisps				
Wisps moving towards potion				
Wisps moving away from potion				
Overall performance				

C. Please rank the potions according to which has fewer errors when used to catch the wisp in each condition. If you have any comments please write them down in the comments column.

	Red	Blue	Green	Comments
Smaller wisps				
Larger wisps				
Faster wisps				
Slower wisps				
Wisps moving towards potion				
Wisps moving away from potion				
Overall performance				

D. What are the advantages and disadvantages of the blue potion?

E. What are the advantages and disadvantages of the green potion?

F. How to improve the effects of the blue potion?

G. How to improve the effects of the green potion?

H. Do you have any comments/thoughts about the whole user study?

B.7 Mevie User Study

THE UNIVERSITY OF BRITISH COLUMBIA

Human Communication Technologies Lab
Department of Electrical and Computer Engineering
University of British Columbia
2366 Main Mall
Vancouver, BC, Canada V6T 1Z4

VIDEO HISTORY-BASED MOBILE APPLICATION **SUBJECT #:**

Thank you for participating in this user study. We would like you to give us your feedback on this user study. If you have any questions regarding the completion of this questionnaire please do not hesitate to ask at any time.

Part 1: Demographic Description

1. Gender:
☐ Male ☐ Female

2. Age:
☐ 19 - 24 ☐ 25 - 29 ☐ 30 - 39 ☐ 40 - 49 ☐ 50 - 59 ☐ 60 or over

3. Your occupation:
☐ UBC Undergraduate student ☐ UBC Graduate student ☐ UBC Faculty
☐ UBC Staff ☐ Other (specify) _____

4. 4. Area of occupation:
☐ Science ☐ Health ☐ Engineering ☐ Arts ☐ Business / Law
☐ Other (specify) _____

5. How often do you watch videos online?
☐ Daily ☐ A few times a week ☐ Once a week ☐ Rarely ☐ Never

6. How often do you share videos you've watched online?
☐ Daily ☐ A few times a week ☐ Once a week ☐ Rarely ☐ Never

Part 2: Interface				
Please check the numbers, which most appropriately reflect your impressions about using this application. Not Applicable = **NA**. Please add your written comments below the corresponding item.				
		1 2 3 4 5 6 7		NA
7. Having this history as part of the video player is useful	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
8. Having the history always available (in the drawer) is helpful	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
9. Presenting the history as a pull-out drawer feels natural	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
10. Browsing the history drawer is easy	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
11. Finding a video using the history drawer is easy	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
12. Finding a video in the fullscreen history is easy	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
13. Sharing a clip from the fullscreen history is easy	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
14. Sharing a clip from the video player's filmstrip is easy	Strongly disagree	☐☐☐☐☐☐☐	Strongly agree	☐
Part 3: Interview Questions				
1. What do you think about the application?				
2. Would you use this application?				
3. Would you actively use a video history?				
4. Which method did you prefer for sharing: filmstrip or history?				
5. Do you have any suggestions/recommendations of improvement for the application?				

Appendix C

Additional Experiments' Data

C.1 Behavioural User Study

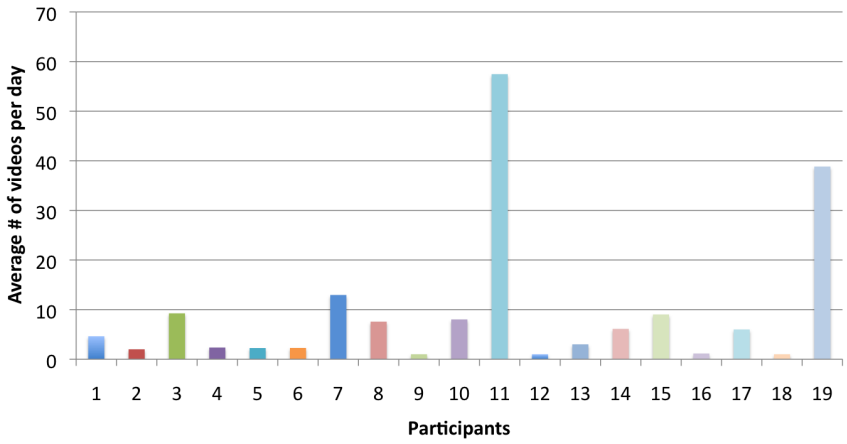


Figure C.1: Average number of visited videos per day for each participant. Participant 11 watched the most number of videos on average per day, while participants 9, 12, 16 and 18 watched only one video on average per day.

The total number of unique videos cannot be used to judge how often participants watched videos since this data is collected over a variable duration based on participants as can be seen from Table 3.1. Thus, we tried to look at how many videos were viewed on average per participant's active day, shown in Figure C.1. Participant 11 watched the most

number of videos per day which can be an indication of re-visitation or replaying previously seen videos. Looking at Figure C.1 and the number of videos per session participants reported in Table 3.1, we found that eleven participants misconstrued their average number of videos per session and this is clear from participant 18 who reported that he watches more than 10 videos per session but the data revealed that he actually watched only one video on average per day. Moreover, when data was analyzed for how often videos were watched, the results also showed that nine participants misconstrued how often they watch videos. For example, participants 12 and 18 reported that they watch videos on daily basis, however, the data showed that participant 12 rarely watches videos while participant 18 watches videos once a week. This might not be true as some participants may watch videos in their reported frequency but on different devices which do not have our extension installed on it and this means missing data from these participants in our dataset. Being able to interview participants after the experiment would allow us to justify this difference. However, due to some privacy issues participants did not want to be recognized in person and linked to their viewed videos.

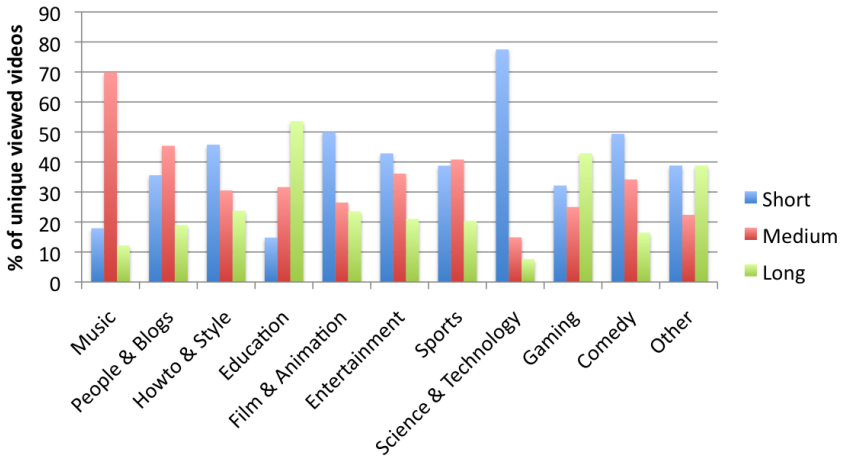


Figure C.2: Visited videos in each category grouped into short, medium, and long videos. Most of the viewed videos from Science & Technology were short videos (78%), while most of Music videos were medium videos (70%). Education had half of their viewed videos categorized as long videos.

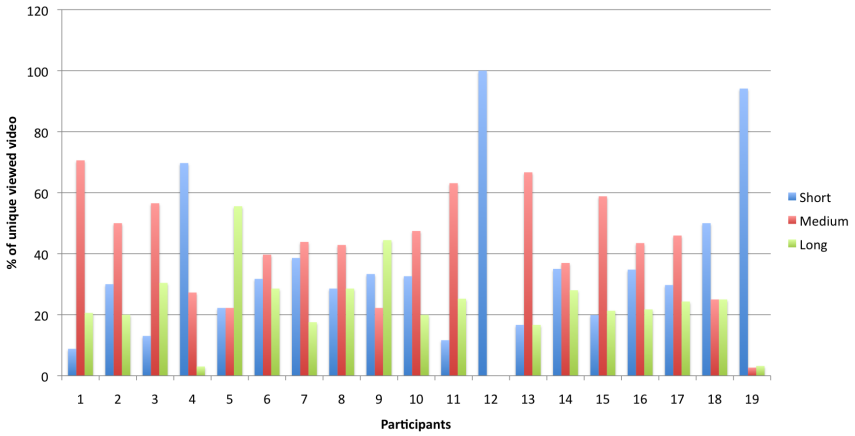


Figure C.3: The distribution of each participant viewed videos among the three duration groups (short, medium, and long). Participants 12 and 19 watched mostly short videos, while half of the videos participant 5 watched were long videos. Medium length videos were the most viewed videos for participants 1, 3, 11, 13, and 15 where they had more than half of their viewed videos categorized as medium.

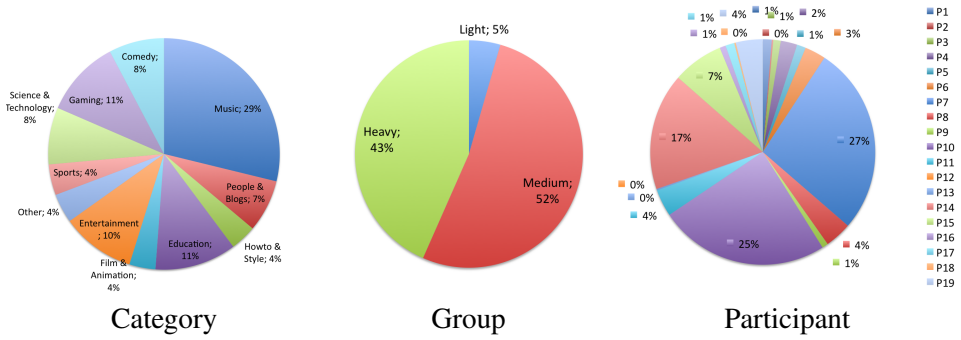


Figure C.4: Videos containing skips distributed per category, per participant, and per users group. Most of the skipped videos came from Music category (29%), participants 7 (27%) and 10 (25%), and the medium viewers (52%).

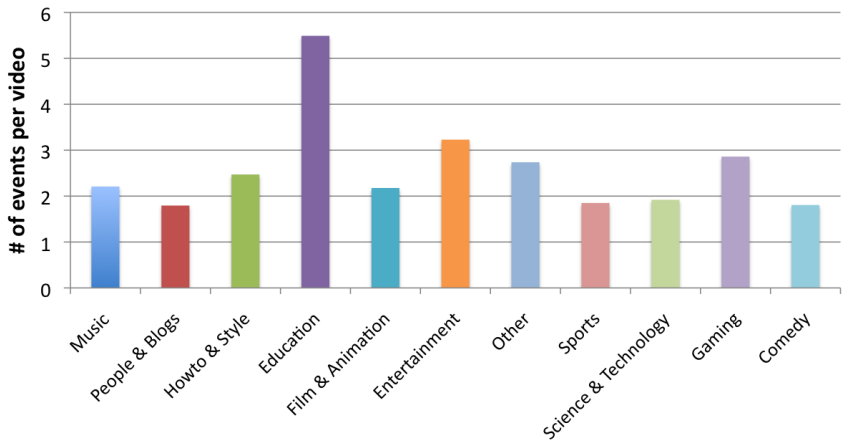


Figure C.5: For skipped videos, this is the average number of skip actions occurred per video within each category. An educational skipped video contained on average 5 skip actions indicating a search for specific information within a video.

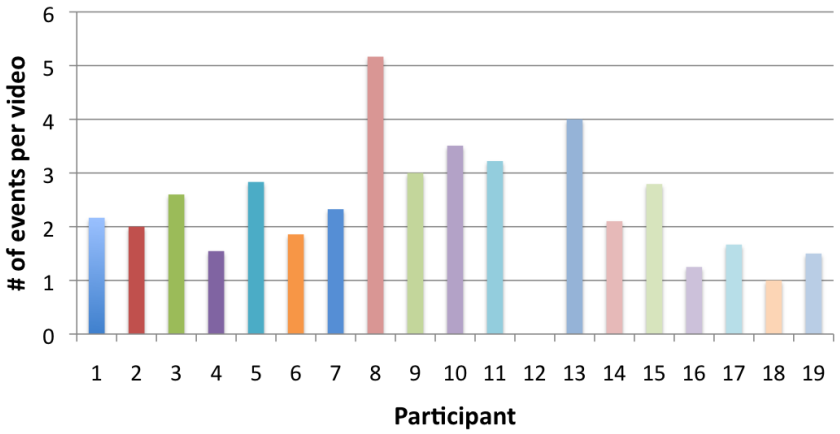


Figure C.6: Number of skips occurred per skipped video for each participant. Participant 8 and 13 had the highest average number of skips per video. While subject 12 did not skip any video.

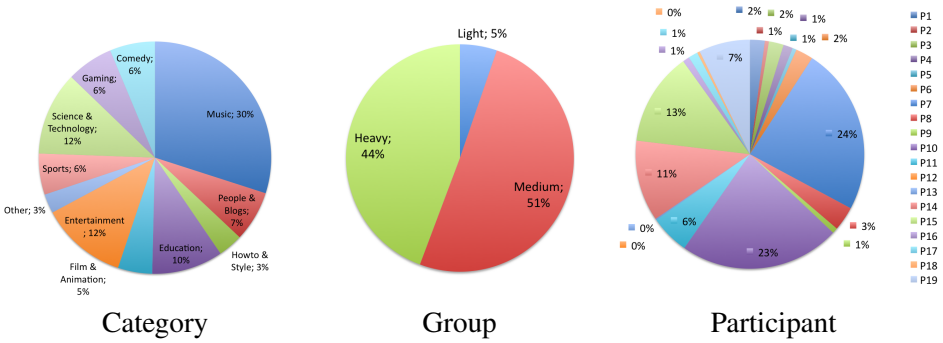


Figure C.7: Videos containing re-watch behaviour distributed per category, per participant, and per users group. Most of the videos that had re-watched portions came from Music category (30%), participants 7 (24%) and 10 (23%), and the medium viewers (51%).

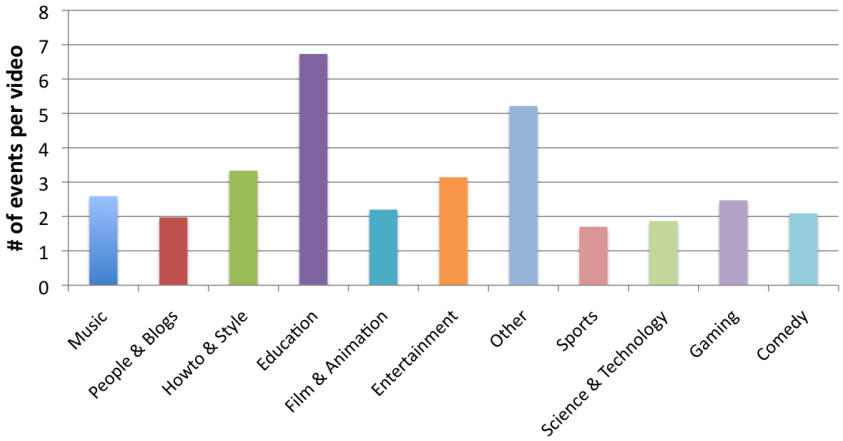


Figure C.8: Number of re-watch actions occurred per re-watched video within each category. An educational re-watched video had the highest average number of re-watch actions (seven) which confirms [70] findings.

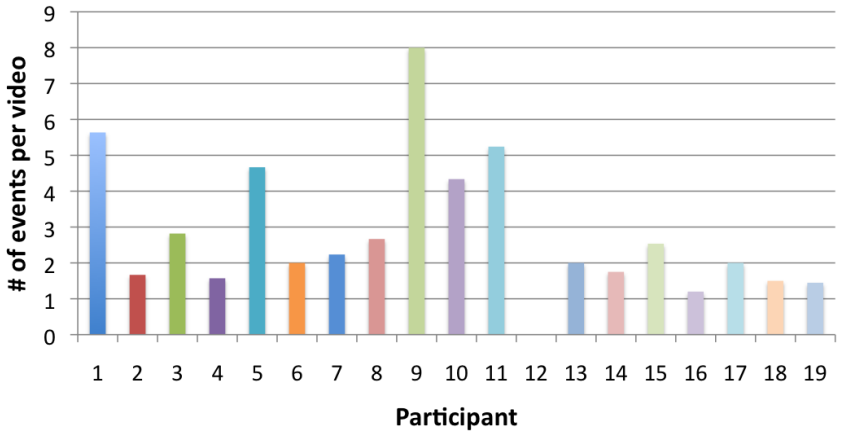


Figure C.9: Number of re-watch actions occurred per a re-watched video for each participant. Participant 9 re-watched video had eight re-watch actions on average where most of these performed in Science & Technology videos.

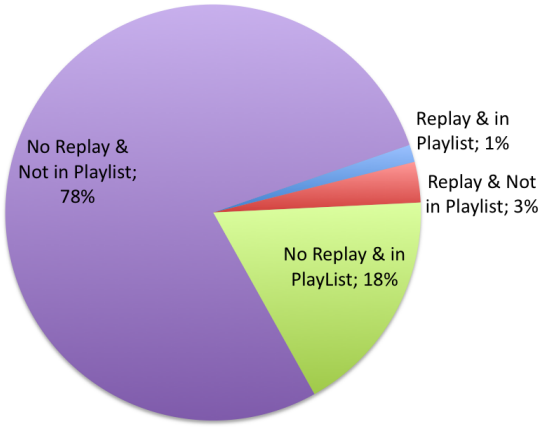


Figure C.10: Percentage of videos being replayed and whether they are in a playlist. 4.5% of videos had replay actions where only 32% of the replayed videos came from playlists while 68% of these were intentionally replayed.

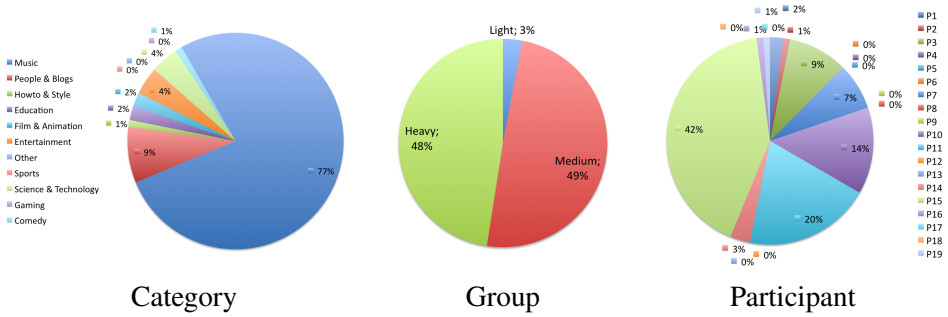


Figure C.11: Videos containing replay behaviour distributed per category, per participant, and per users group. Most of the replayed videos came from Music category (77%), participant 15 (42%), and from both heavy or medium viewers (97%).

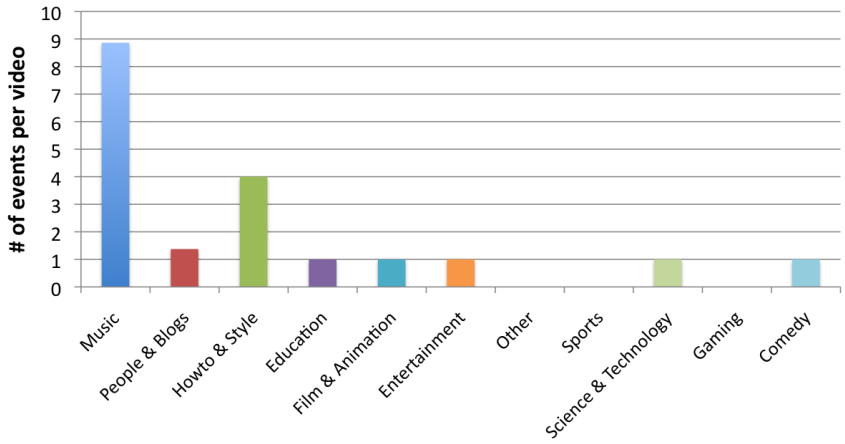


Figure C.12: Number of replay actions occurred per replayed video within each category. When a music video was replayed, then it was replayed 9 times on average.

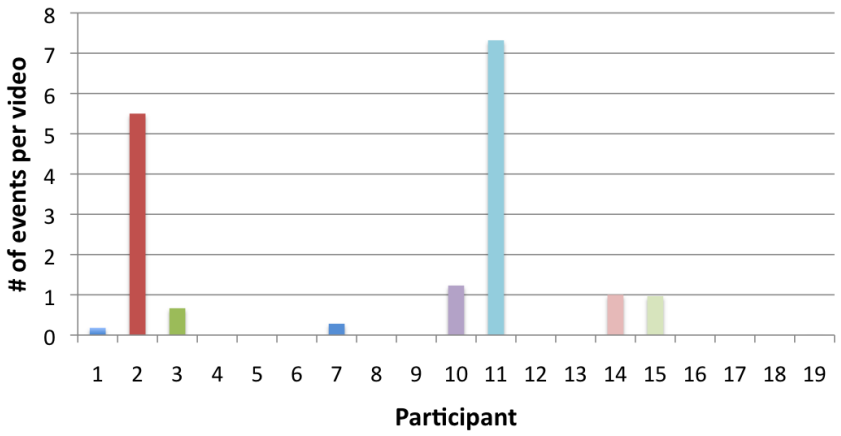
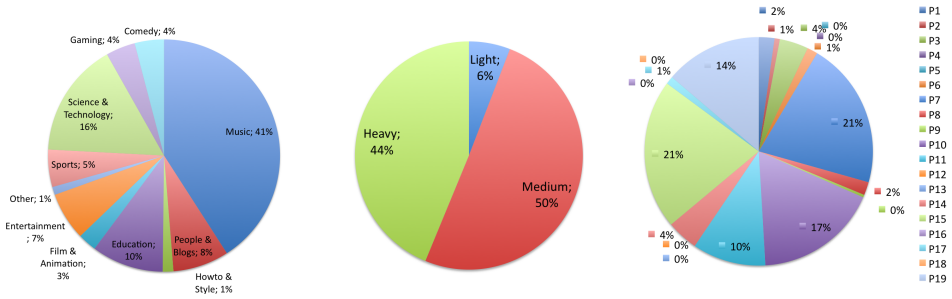


Figure C.13: Number of replay actions occurred per replayed video. When participant 11 replayed a video, then it was replayed 7 times on average.



Category Group Participant

Figure C.14: Videos containing revisit behaviour distributed per category, per participant, and per users group. Most of the videos that were accessed again in a different session came from Music category (41%), participants 7 (21%) and 15 (21%), and the medium viewers (50%).

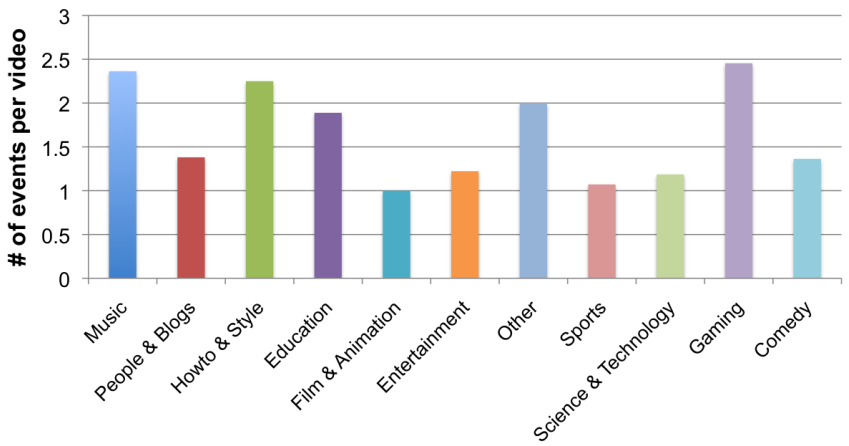


Figure C.15: Number of revisit actions occurred per revisited video within each category. A revisited How-to & Style video was visited 2.6 times on average indicating the resumption of the video content in multiple sessions.

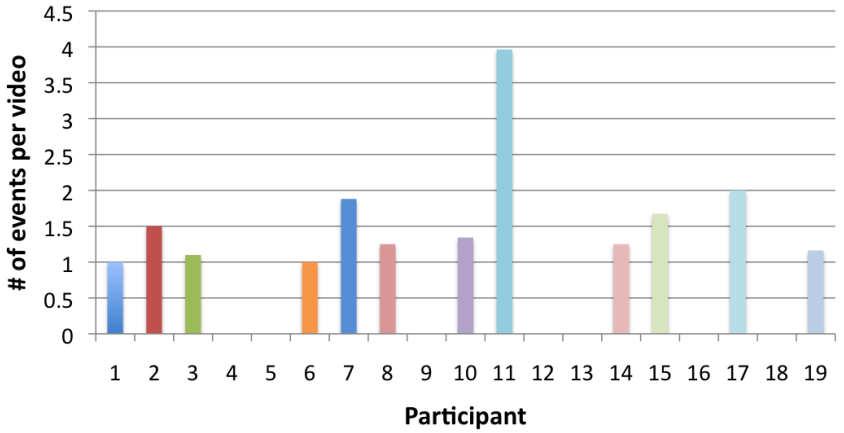
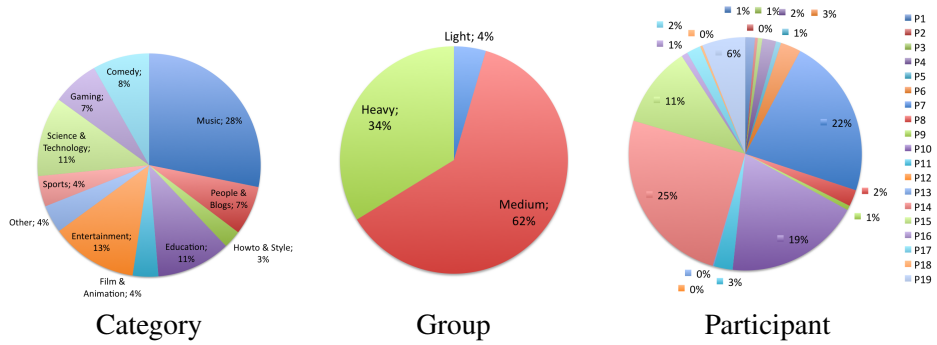
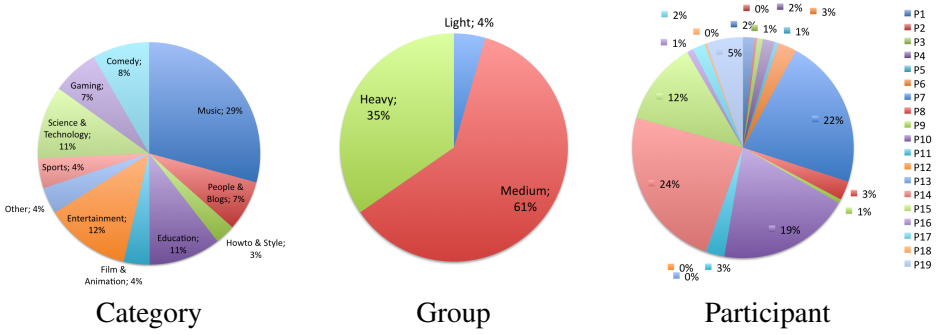


Figure C.16: Number of revisit actions occurred per revisited video for each participant. When participant 11 revisited a video, then it was revisited on average 4 times, while a revisited video for participant 17 was accessed twice on average



Category Group Participant

Figure C.17: Videos containing drop-off behaviour distributed per category, per participant, and per users group. Most of the videos that were accessed again in a different session came from Music category (28%), participants 14 (25%), 7 (22%) and 10 (19%), and the medium viewers (62%).



Category Group Participant

Figure C.18: Interrupted videos distributed over categories, participants, and viewer groups. Most of the videos that were accessed again in a different session came from Music category (29%), participants 14 (24%), 7 (22%) and 10 (19%), and the medium viewers (61%).

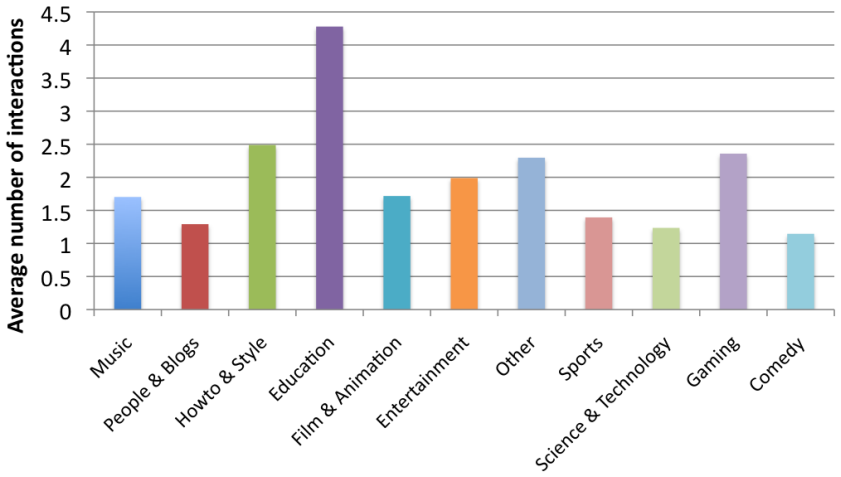


Figure C.19: Average number of interactions per video for each category. At least one interaction (i.e. either skip or re-watch) per video occurred in each category where Education videos were the most highly interacted with four interactions on average per video.

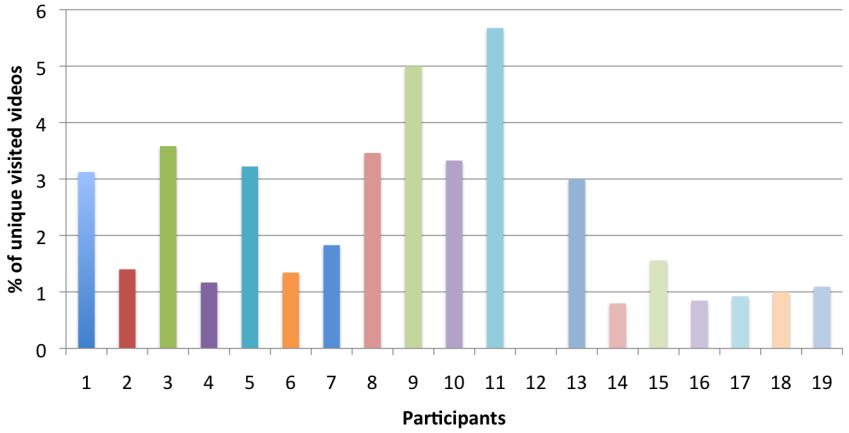


Figure C.20: Average number of interactions per video for each participant. Participants interacted at least once (i.e. either skip or re-watch) per video while viewing. Participants 9 and 11 were very active while watching where they had five interactions on average per video.

C.2 Mevie User Study

Table C.1: Participants' response when asked what they think about the history-based mobile application *Mevie*. Most participants found it useful and interesting.

P	What do you think about the app?
1	Good app, like the history, very good for looking for things you've seen in the past.
2	I think it's good, because when I was watch another video, I want to skip forward, I guess I need to remember what part I have already seen. In here I can see, what part. This is very advanced application.
3	I like it. I like that you can share just parts of the video and not the whole thing.
4	Well, I guess I'm not a fan of sharing. I can't imagine that part would be useful. I don't want Google to store my information.
5	I think it's interesting, but to be honest, I'm not a huge fan of sharing. I don't share much. But I find the history is very useful. The thing that i know which parts I've already seen. So I think that's useful.
6	I think it looks neat, quite organized. It's good.
7	I guess if you're looking at a movie, and you want to start at a specific point, or you leave it and you come back, I guess you could. And if you're looking for something specific. Yeah interesting.
8	I think it's a cool app. It's easy if you want to share stuff.
9	I think it's cool. I'm not really sharing a lot of videos. It's good maybe for, if you have videos connected to your studies, if you want to do research, It's easy, it's very good.
10	It's good. I like being able to share part of a video, good for sports videos to create highlights.

continued . . .

Continuation of Table C.1

P	What do you think about the app?
11	I like this sharing thing because sometimes you just want to share part of the video, sometimes videos are really long and you only want to share a few seconds. I think this is a very good application for your app. The history is also useful. Especially because you can see the most viewed part of each video, instead of only the beginning or end, so it is more representative for you to remember.
12	It's good. This is much easier, doesn't take as many steps
13	I think best part of it is the feature that we can share part of the selected video because usually when we watch video online, we share the whole video. But in this one, we can select the best part we want to watch, or share the best parts.
14	It's very useful because sometimes I need to replay part of the video and it's really only this part of the video I'm interested in. I always have some difficulties to find which part of the video I should go down and find where is it. It's a really good application.
15	It's interesting. I think it has some interesting applications that I haven't sen before in other apps.
16	It's very good.
17	Better than Windows Media Player.
18	It's an interesting app. Well designed, and like to use the history for finding old stuff.
19	It's good. The history can be useful for finding previously seen videos and you can share part of a video and not the entire thing.
20	it is awesome.
21	Interface is nice.

continued . . .

Continuation of Table C.1

P	What do you think about the app?
22	Pretty good.
23	Interaction is good.
24	Might be suitable for only certain kinds of video.
25	It's natural.
26	Good application, I can see its applications in gaming.

Table C.2: Participants' response when asked would they use the history-based mobile application *Mevie*. Most participants said yes and for those who said no they think it can be used if it is integrated to video websites such as YouTube and Vimeo.

P	Would you use this app?
1	Probably not. I watch video on the TV, and would fling these to the apple TV. Can imagine using it as a remote when looking for video, but watching video on a tablet is not generally something I do.
2	Yes.
3	Maybe, I don't share videos a lot. I mean I watch them, so maybe if I had an easier way to share them I would. Yeah, I would use it.
4	Well, I guess, not personally, but I guess other people would. As long it isn't shared as personal.
5	Yes, I would use history definitely. and I really like the filmstrip, with the small scenes, that I can drag to see where I am supposed to go, with the small captures.
6	Yeah I would, but usually like share video, I would share YouTube, I would send a link so everyone can watch it.

continued . . .

Continuation of Table C.2

P	Would you use this app?
7	I probably would. But not maybe not as much as I should. I probably would use it at some point, because sometimes you want to see something specific.
8	I don't know, because. when I watch things like this, I usually go on YouTube. If it were apart of YouTube, then yeah I would definitely do it. Especially if I wanted to share it. I think it's really clever, but I figure it would be hard making people use it instead of YouTube. But I think it's clever. I mean it's easy for people to be. I don't know really how, say watch a long thing and you just think if one part of it is funny, I'm going to share it, and you don't want to share something that is 10 minutes long. Yeah. And then it's really good. It's really user friendly.
9	Yeah, if I found it necessary. Maybe I would use it mostly to watch videos.
10	Yeah.
11	Yeah sure.
12	I'm not a video lover. I have seen editing video parts, those take more steps. I prefer this one if I do it.
13	Yes.
14	Yes, I would use it.
15	Yes, I would use it.
16	Yeah, I think it would be all about the database. If you could access YouTube, or other websites, Vimeo etc.
17	I don't know if I would use it as often as native players.
18	Yeah, I would use it.
19	Yes, if integrated with YouTube/Vimeo.

continued . . .

Continuation of Table C.2

P	Would you use this app?
20	Yes
21	For long videos and to save videos. I would not use for short videos.
22	Definitely.
23	Yes.
24	Yes.
25	Yes.
26	I would use the app, but history not so much.

Table C.3: Participants' response when asked if they would actively use the history in *Mevie*. Most participants answered yes and for those who said no they think others might use it but not themselves.

P	Would you actively use a video history?
1	I would use it for finding stuff after a few days, not right away.
2	I perhaps. I would use it looking through the history.
3	Yes, I think so.
4	As long it isn't shared and it is personal.
5	Yes, I would use history definitely.
6	Yeah, I would like to use the history, because I could see what I watched. And that's good about the app.

continued . . .

Continuation of Table C.3

P	Would you actively use a video history?
7	Yeah, the history would be really good because you can go back and check certain things or see or, depends on what you had on that you kept. I think the history would be great. That would be a a good thing. I would really like that.
8	Yeah, I would definitely do it.
9	Yeah.
10	Yeah.
11	Yes, I think so because sometimes when I have troubles and I look over the web for the videos, I like to be reminded what I saw. It's easier for me to remember what I saw if I have a timeline like this.
12	Yes, I think.
13	Yeah, I think so. Especially it shows which part we have watched, rather than just the whole video.
14	Yes, I would use it. Yesterday, I had the problem of finding a part of the video, because I may need those part of videos to watch it again.
15	The history is not really necessary for me. It's a good part of the app. It should be for some other users. Not for me. I always watch a video one or two times at most and share it.
16	I think so. I use Skype to send YouTube links. When I want to find the video again, I scroll up and find the link, but I think if it was all integrated it would be pretty useful.
17	Probably not. I know what kind of videos I have in there. It could be useful for online videos.

continued . . .

Continuation of Table C.3

P	Would you actively use a video history?
18	Yes. I would use it to find video from the same channel and find similar videos. I wouldn't use it to re-watch video though.
19	Yes, if I wanted to find something from a long time ago.
20	For long videos and when I'm trying to find something
21	For long videos.
22	I would use history, but I don't know how much of the sharing mechanism I would use. The history is very natural. Visibility of the drawer provides good affordance for something new.
23	Yes.
24	Yes.
25	Yes. it's convenient
26	Not much.

Table C.4: Participants' response when asked about any comments or suggestions to improve the design of the history-based mobile application *Mevie*.

P	Any suggestions? Such as design or layout of home screen or history?
1	No.
2	Really good.
3	No, I like the layout.
4	Have a screen with video itself. Interface is really confusing.

continued . . .

Continuation of Table C.4

P	Any suggestions? Such as design or layout of home screen or history?
5	The idea of the history drawer to the right isn't that natural. The YouTube on the right is related video. I understand that history should be on the bottom.
6	One way to share the part. It was kind of confusing. Stick with the filmstrip.
7	Nope. I quite like it.
8	Divide history into morning/evening etc.
9	Maybe a search.
10	A lot of the thumbnails look the same.
11	Annotation of history, tags.
12	No.
13	Tag videos. Comments.
14	No.
15	Bigger sizes, Fullscreen video. Colours are not very attractive use Light colours. Don't like the font sizes in the history drawer. Try to have one font size.
16	Thumbnails look similar. Blank similar colour thumbnails for example.
17	Online videos. Make people share most watched sections of videos, make suggestions based on that. Similar interests for certain videos.
18	No.
19	Oftentimes, thumbnails are similar to each other.
20	Some things are not intuitive.

continued . . .

Continuation of Table C.4

P	Any suggestions? Such as design or layout of home screen or history?
21	Needs some icons instead of labels. Colours are nice, but they're a bit typical colour highlights. Would like to have a tutorial!.
22	Size of the history buttons are too small. Long-press for history to bring the context menu. I would like to be able to share a video entirely as well.
23	Want to hide history drawer.
24	Needs Better labels, tutorial - would like immediate use, but may need tutorial.
25	To play selection from the filmstrip.
26	No