INTERACTION SUPPORT FOR INFORMATION FINDING AND COMPARATIVE ANALYSIS IN ONLINE VIDEO

by

Jinyue Xia

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2017

Approved by:

____

Dr. David C. Wilson

____

Dr. Mary Lou Maher

____

Dr. Celine Latulipe

____

Dr. Mohamed Shehab

____

Prof. Sybil Huskey

ABSTRACT

JINYUE XIA. Interaction support for information finding and comparative analysis in online video. (Under the direction of DR. DAVID C. WILSON)

Current online video interaction is typically designed with a focus on straightforward distribution and passive consumption of individual videos. This "click play, sit back and watch" context is typical of videos for entertainment. However, there are many task scenarios that require active engagement and analysis of video content as a means to an end, such as educational material for flipped classrooms and critical review of performance — in sports, performing arts, or other spatial domains. Interface design has focused on the former context in interface interaction controls, and common interaction tools are limited to basic player affordances such as play/pause, speed change, and scrubbing. These basic interactions, however, can be cumbersome for goal-directed tasks such as contextualized communication, information finding, or comparative analysis. More active, goal-directed use of online video requires better interface designs to support these kinds of task-based interaction.

In this research, we investigate approaches that enable and support more active user interaction with online video. In order to conduct the investigation, we present an illustrative research probe that employs analytics on user engagement to facilitate interaction. For example, video segments that are re-played by many users may indicate an important concept to review in a flipped classroom video. To study this kind of interaction, we introduce metrics for measuring the *degree of user engagement* and an interface component for visualizing and interacting with user engagement data as a heatmap. To evaluate our research probe, we have conducted a series of user study experiments. The set of experiments studied the impacts on user interaction of making engagement data available across multiple levels: individual user, small groups, and large groups such as a general population. The results of the studies show that the heatmap interaction approach can provide better support for active interactions with online videos such as information finding, content understanding, and comparative analysis. More specifically in the online learning context, our research prototype supports viewers to find information more effectively and also helps them feel more confident in their learning. For the group of class instructors, the study shows that our research prototype can better support the development and refinement of teaching methods and improve course video content.

ACKNOWLEDGMENTS

First of all, I would sincerely thank my advisor Dr. David Wilson who also serves the chair of my dissertation committee. I could not complete this dissertation without his guidance and support throughout my PhD study. He has provided not only the academic advising for my research directions, but also the finanical assistantship for my PhD career. I would also like to thank my dissertation committee: Dr. Mary Lou Maher, Dr. Celine Latulipe, Dr. Mohamed Shehab, and Prof. Sybil Huskey. Each of them has given me invaluable and constructive feedback. I could not have asked for a better committee.

Next, I would like to thank my colleagues from the HCI lab. Dr. Vikash Singh, Dr. Berto Gonzalez and Stephen MacNeil. All of them helped me with the study designs and provided good suggestions for the analysis as well.

I could not complete my research without financial support. I thank UNC Charlotte for partially funding my research through the Graduate Assistant Support Plan (GASP), Dr. Wilson for his financial support throughout my PhD study, and Dr. Maher and the NatureNet project for the one-year support.

Finally, I would like to thank my family and friends who supported me through this whole process. My wife, Yujuan Qian, has been with me through the long journey. Without her support, understanding, and encouragement, I could not have succeeded.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Online sites for video distribution and sharing (e.g., YouTube[1], Vimeo[2]), have made video a ubiquitous part of our lives. Creation and consumption of video have increased in the past decade due to the wide availability of inexpensive video cameras and camera-equipped mobile phones, as well as higher bandwidth [1, 37, 38]. According to YouTube statistics up to the time of writing this dissertation, over a billion users — almost one-third of all people on the Internet — watch hundreds of millions of hours on YouTube and generate billions of views [48]. This surfeit of video has given rise to a number of different interfaces designed for users to browse digital video content for various tasks [7, 25, 14, 26, 34]. And, if even 1% of the video uploaded to YouTube each minute is relevant to user tasks of analysis and study, that would mean 4320 hours of important, data-rich video is being uploaded daily. Moreover, video is not only used for entertainment now, but it is a useful tool for people to learn [8, 43, 17, 21, 20] and accomplish a particular task with collaboration [44, 36]. For several goal-based tasks such as content re-finding from a familiar video and comparative analysis between multiple videos, user engagement with videos for these tasks tends to be more active. As with active learning [4], active watching or engagement with video implies goal driven activity that involves doing more than just viewing. A user may want to

[1] http://www.youtube.com.
[2] http://www.vimeo.com.

highlight, excerpt, annotate or merge parts of the videos of interest, in the context of their work of analysis. These higher-order interactions can be made possible, but their design will be dependent on the design of the primary player.

In contrast to the typical "click play, sit back, and watch" user interaction with video, better video interaction features are needed for users to complete the goal-directed tasks such as information finding and comparative analysis. The focused research probe in this research is to explore better user interaction with online video. We investigate measures of user engagement with video and using the measures as a means to provide better user interactions with video.

It may be useful to further clarify the task focus under investigation. While many of the research questions apply to video interaction in general, we focus on online video interaction. We also presume that the task involves the use and analysis of existing video as a means to an end, and not the creation or modification of the video itself. Thus, this research does not directly address video recording or editing tasks, even though the analysis tasks considered here may help to inform creation/editing. In addition, this research does not directly address *automated* video content analysis (segmentation, object identification, etc.); we recognize the potential benefits, but such strategies are typically very domain-specific. The focus here is on a baseline, direct user interaction with and analysis of existing video content.

The remainder of this chapter introduces our research focus area dealing with user engagement. It concludes with a summary of contributions to the research.

1.1 User Interaction With Video

Online video serves not only as entertainment but is now a tool in people's daily lives. For example, many people engage in self-study or DIY tasks by following tutorial videos — learning sports, mechanics, and even studying a programming language. An excellent example of video applications is online education. Thousands of free online courses are open and available in existing Massive Open Online Courses (MOOCs) platforms (e.g., edX[3], Coursera[4], Udacity[5]). The courses are totally free and most often are taught by professors from top universities. Additionally, online learning gives flexibility to the students to re-watch, slow down, or speed up the video at any time. Therefore, each of the platforms attracts a huge number of viewers worldwide [39]. For example, according to Coursera's about page[6], 24 million learners are reported to have used their platform. In fact, more than the application of online education, video can be a means of marketing, sports analysis, and so on.

User interaction data, collected implicitly (e.g. [20, 21, 49]) and explicitly (e.g. [34, 26, 11]), conveys user engagement with the videos from two levels: *video-as-a-whole* and *video segment*. The explicit data can be captured by many approaches such as collecting additional feedback (e.g. like and dislike) and explicit ratings from users. Usually, this kind of explicit data is gathered from a *video-as-a-whole* level because user feedback is based on the entire video. By providing the opportunity for users to rate a video, many video systems learn user engagement of the whole video from

[3]http://www.edx.org.
[4]http://www.coursera.org.
[5]https://www.udacity.com/
[6]https://about.coursera.org/

the explicit ratings. For example, a video system Time Warp Football proposed in the work [34], enables the viewers to rate the video clips. The explicit ratings can then be used to analyze the video clips and find out which clips are the most popular. Another example is that the 5-point scaled ratings are collected for the Físchlár-News-Stories recommender system of TV news stories [45]. Implicit user interaction data, including user click data with video such as play, pause and seek, is complementary to the explicit data for helping to understand user engagement with the video. In contrast to collecting explicit data from video-as-a-whole level, the implicit user activities are usually collected from *segment-level* user's click data within a video, for example, click-level events such as play, pause, seek and so on. It is straightforward to capture the events as needed at certain times for later analysis. The implicit data is commonly used for data analytics in many of above systems too. For example, both YouTube recommender system [11] and the Físchlár-News-Stories [45] TV news recommender system log user activities such as the play event to help predict personalized videos. Since the implicit data contains segment-level and most likely second-by-second user interaction data, it can help us measure and understand fine-grained user engagement with video. However, to our knowledge, there is limited research on profiling user watching patterns on video and on measuring how the users engaged with video based on *segment-level* data.

1.2 Segment-level User Engagement With Video

In this research, we focus on segment-level user interaction data analytics. If the click-level data along with the time and user information can regularly be tracked, the long-term interaction data can represent the user's watching behavior with the particular video. There are several benefits for discovering viewer engagement with video. For example, showing user engagement can potentially change the user's watching behavior [42] and make the video to be more engaging [18] from the perspective of video production. A one-minute long segment in a video with a higher number of views may warrant more attention because it might mean the section is interesting, important or confusing [18, 20]. Besides helping watch video more effectively, learning user engagement can help a video producer make high-quality videos to attract more people to watch [18]. Following are some example scenarios illustrating how engagement metering could support goal-directed tasks:

- A lecturer dedicates himself to make clear and straightforward lecture videos for the students. After a lecture, he can check the watching patterns of students on his lecture. He finds several watching peaks. While most of them are making sense because these are highly important concepts, he figures out one peak is unusual because many of the viewers slowed down the playback rate. He then finds out that his speaking speed is too fast in that specific segment. As a result, he decides to record that part again and produce a new video.

- A newly registered student in an online course platform wants to start her first class. Since professors sometimes talk about unrelated content before the lecture begins, she does not want to waste time on the beginning of the video. Instead, she wants to jump to the lecture as quickly yet accurately as possible. With current video systems, there is no clue for her to know the actual time for the start of the lecture or where the relevant information is located in the video. But if user engagement with this video is well discovered and available to her, instead of following the traditional video player's timeline, she can choose where to start to watch based the meaningful indications of user engagement.

- In the week for preparing final exams, students like to locate and review the valuable content as quickly as possible. In a flipped classroom, videos are the source for the class materials. So students might want to seek some important segments from some related videos. However, current video systems do not support this type of active video interaction very well. By providing a visualization of how previous viewers watched the video, the students in such a critical time might benefit from earlier watchers.

Generally, three processes are involved with user engagement metering: data collection, data analysis, and data visualization. To display user engagement with a video, we first need to collect user interaction data and browsing history. Second, with the captured user watching profiles, the data can be converted to a meaningful understanding of how users engaged with the video. Finally, with the analysis, the data can be visualized to the viewers such that they can notice there is a difference of watching patterns between different segments in a video.

Although it is straightforward to capture interaction data from video viewers, the data can be interpreted in various ways. There are two representative types of segment-level user engagement measurement: finding *highly-interacted* [20, 21, 49] segments and finding *highly-played* segments. A highly-interacted segment means the segment has frequently been interacted by users. The interaction comes from any video event such as play, pause, seek, and volume change. User engagement measurement based on this metric is simple and straightforward because highly-interacted can be detected if the segment contains a high number of interaction events. This metric, however, loses the context or different importance level of the events. A highly-played segment, however, represents a segment has been played frequently. Ideally, it is better to collect data about how the segment has been watched; however, it is tricky to realize a segment has been watched because the user may have the video playing in the background [18]. Assuming the majority of users look at the video without interruption, the highly-played segments can represent the highly-watched parts within a video. To measure user engagement based on the highly-played metric, the distinct interaction data needs to be re-constructed to the user's full watching sessions.

One way to extract the watching pattern is to learn from the click-level events generated by the video viewers [20, 21]; while another approach is to convert the user interaction data based on a second to a level of importance of the video segment depending on the operation from users [49] (e.g. play, pause). The first approach is a purely data-driven approach because when an interaction with the video happens at an individual segment, it will be counted as once for the computation of the total counts of events that happen in that particular part. As the number of viewers grows, the interaction pattern on the video can be discovered. It is simple and straightforward for the computation; however, a highly-interacted segment does not mean it has been highly watched by the video viewers because a user may leave a lot of pauses during the segment. A play event may also weight more than a stop event. To address this drawback, the second method gives a different weight to different operations on a video. Rather than only counting the events, they analyze the link between each operation and the level of importance of each video shot (or segment). Similar to above research [49], many works focus on a highly-interacted metric for the measurement. So, there is an absence of a formal and commonly used approach to analyzing user interaction data during the watching of a video when the used measurement metric is based on "played". In this work, we investigate the conversion from user interaction data to segment-level user engagement measurement based on how often the segment has been played.

To study how user engagement impacts video viewers' watching, it is important to reveal user engagement in a straightforward way after converting the interaction data to user engagement. The problem becomes to how to visualize user engagement with the video. Video is not the first domain to leverage user interaction history and visualize user engagement data. Many data visualization techniques have been used in showing user engagement in many areas like education [20] and art performance [23]. Line graphs are used to visualize audience engagement with a dance performance [23]. A rollercoaster wave-like timeline with interaction peaks presented in an online educational video player [20]. One widely and successfully adopted visualization technique, and we will use for our work is the heatmap. Crazyegg [9] uses heatmap to represent user's web browsing traffic so that users or the web host can recognize what part of the web page attracts more attention. Patina [30], another example, also uses dynamic heatmap overlay the software to show user interaction frequency with the software and identify where the users interact frequently and rarely. Similar to the existing applications, we will choose heatmap as a visualization tool to study segment-level user engagement with video.

1.3 Problem Space

Most of the existing research related with user engagement with video leverage explicit and implicit user interaction history and analyze interaction data on a video-as-a-whole level. In this research, we focus on the segment-level user engagement with video. In order to measure a fine-grained segment-level user engagement, the second-by-second user interaction data is used in recent research [20, 49]. However, the segment-level data is used directly for a “highly-interacted” styled user engagement measurement. So, we will investigate segment-level user engagement with video by using a metric of “highly-played” type.

For the “highly-interacted” user engagement measurement, the approach of only counting the interaction events per second has been used. The problem with this method is that it loses the context of the video playing. For example, a segment played with different playback rates may indicate the level of the importance of the segment. In this work, we propose a new approach to computing segment-level user engagement with video by converting the interaction data with contexts such as the playback rate when the interaction happens. To collect the interaction data, we set up an online video environment to track user implicit interaction data. After accumulating enough data, the new approach of computing user engagement will be applied to explore user watching behavior over videos. After the computation of user engagement, the heatmap is used to visualize user engagement data, because it has a "temperature" indication for user engagement. We will run several experiments to evaluate whether the heatmap of user engagement is helpful in goal-directed tasks such as navigation, content finding, and analysis.

1.4 Thesis Statement

We propose to address the general research question of how novel approaches to interaction with online video can better support end users for active, goal-directed use of video content. To investigate this research question, we focus on one specific probe into the space of video interaction, evaluating interaction interfaces based on user engagement analytics. To understand the impact of interaction designs in these contexts, we will conduct user studies evaluating the following thesis. *The design features of the user engagement heatmap provide better task support than standard online video tools for user interaction and active consumption of online videos in goal-directed tasks such as information finding or comparative analysis.*

1.5 Contributions

Our completed research results in the following primary research contributions.

- We have designed and developed baseline metrics for measuring user engagement with online video, as well as a user engagement heatmap component based on these metrics. The user engagement heatmap visually indicates the level of engagement across segments of video and provides visual cues to support user navigation and interaction. We propose different user engagement metrics and parameterizations based on the captured user interaction data. In addition to capturing click-level interaction events, we consider the contexts when the interaction happens, for example, the playback rate of the video player.

- By using the proposed metrics for computing user engagement, we developed two research prototypes for visualizing user engagement with video. One interface with a single heatmap aims to help future video viewers in learning such as information finding; while the second interface with multiple heatmaps is designed for the class instructors to better support comparative analysis.

- We conducted a series of user studies to evaluate how the two prototypes affect video users in using videos for active interactions. The user studies show how the presence of the user engagement heatmap as a new kind of interaction affordance impacts people's video watching, navigation, and analysis with a single video.

1.6 Dissertation overview

This dissertation is organized as follows. Chapter 2 presents background and related work on video players, video interaction and video engagement. Chapter 3 describes data collection of user interaction, investigates the approach of computing user engagement for video, and finally explores the designs for interacting with user engagement. In this chapter, we also illustrate the design goals of the research probe and how it works. In Chapter 4, we examine the research questions for the research probe presented in this work. And the detail of experiments for the evaluations of the proposed prototypes will be described in this chapter. Chapter 5 presents a pilot study and the findings. Chapter 6 and Chapter 7 describe the research findings that were discovered from a user study conducted in a lab setting and a user study finished on Amazon Mechanical Turk respectively. Both of the user studies were designed to evaluate the interface with a single heatmap. For the evaluation of the interface with multiple heatmaps, the details and results are written in Chapter 8. In the last Chapter 9, we provide a summary of the contribution and future work of this research.

CHAPTER 2: BACKGROUND

Previous work has investigated a variety of aspects related to user interaction with video. This chapter begins with a summary of research on video interaction more generally, and it goes on to detail previous work specifically related to user engagement with video and its applications for supporting user interaction.

Video interaction research has investigated interface design to help users browse video content more quickly or efficiently through different approaches. One way is to enhance the traditional video navigation techniques for specialized scrubbers [19, 29, 41, 40]. The content-aware dynamic timeline [40] overcomes traditional player scrubbing problems in (1) fast-speed skimming through a long video, (2) low-speed navigation in a small section of the video. To provide real-time scrubbing of online videos and avoid network latency, Swift [29] overlays a small and low-resolution thumbnail of the video. In fact, thumbnail overlays are widely adopted in many industrial video systems such as YouTube. Zliding [41] uses the pressure modality to fluidly and explicitly zoom to support precise navigation while sliding. Keyframe detection [16] in the video is another method that helps quick navigation for videos. Li et al. [25] developed techniques for time compression, pause removal, textual indices and shot boundary frames, and SmartSkip [14] helps video viewers to skip video segments using the traditional TV remote. Video summarization [26, 34], a thread of research to extract keyframes and shot boundaries, aims to shorten video, provide an overview of the video and support rapid navigation over the video. Olsen's work [26, 34] in football game videos provided several innovative interfaces for users to control video playback, skip to the next play, and switch camera angles by pressing a button on a game controller. Improving video navigation is one of our design goals as well, and we will investigate video navigation through visualizing collective user engagement — particularly how this can support performance for goal-directed tasks.

Video analysis research has focused on video segmentation and annotation. Goldman et al. developed a system that supports annotations in the form of thought balloons, path arrows, and video hyperlinks to allow for interactive manipulation [17]. Silver [33] is a video-editing tool which displays an explicit 3-level view of the video timeline when the user zooms down into a video segment. The Silver system is aimed at editing and rearranging video for the purposes of creating a final video artifact, but it is not a tool for video analysis. Some commercially available video systems (e.g., YouTube, Viddler[7]) support limited forms of video annotation such as text or sketch but do not address goal-directed tasks. VidCrit [37] presented a system that allowing video recordings for video annotations. Although video annotation is not the focus of this dissertation, as part of our design goals, we aim to provide better user interaction for goal-directed tasks such as video analysis.

This previous research in navigation and analysis shares a common goal to provide better user interaction with video. The focus is on single-video applications without leveraging user interaction analytics. However, user interaction leaves rich information about user watching patterns. The following sections address previous

[7] http://www.viddler.com

work-related more specifically to segment-level user engagement with video.

2.1 User Engagement With Video

Both explicit (e.g. ratings [34, 45], likes [11]) and implicit (e.g., play, pause, jump [20, 18, 21]) interaction data generated from video viewers while they are watching the videos, can be used to profile users' viewing behaviors, measure user engagement, and potentially help video producers improve their videos. For example, to measure user engagement and make video recommendations, YouTube [11] explicitly collects likes and dislikes from users and implicitly logs user activities. Físchlár-News-Stories [45], another video recommender, uses a 5-point scale ratings as the explicit data and records the play event from users as the implicit data to measure user engagement and recommend TV news stories instead of online videos. Wistia[8] logs user watching activities implicitly and generates a heatmap for the video owner to track the video's traffic.

One particular domain, online education, provides another good example. Massive Open Online Course (MOOC) platforms, such as Coursera and edX, focus on videos as a way to present concepts to large numbers of learners. And recent research has focused on understanding user interaction with MOOCs, to help learners engage with the material more easily and efficiently [8, 43, 20, 18, 21]. Kim et al. [21], for example, explained video dropouts and interaction peaks in online lecture videos and showed that interaction peaks could be found from tutorials and re-watching sessions.

Recent research focuses on analyzing what can impact user engagement with video

[8] http://www.wistia.com

or quality of user experience with video. For example, the works [13, 2, 22] show the quality of user experience depends on the quality of internet video such as video streaming and buffering. Guo et al. [18] used two proxies to measure *user engagement*: 1) the length of time that a student spends on a video, 2) the problem attempts from a student. They analyzed the data from 6.9 million video watching sessions, measured how learners engage with MOOC videos and provided a list of recommendations for lecturers and video producers. Comparing long videos, some of their findings indicate that shorter videos are more engaging, talking head is more engaging, high production value might not matter, and so on. Based on their observations, they listed recommendations for making more engaging videos. They also tried to understand user engagement by analyzing two dimensions of within-video interaction: interactivity and selectivity. The metric of interactivity is the pause event per second per student while the metric of selectivity is the standard deviation of pause events across all seconds in a video. They claimed that high interactivity and selectivity could indicate more active engagement, and they found that tutorial watching is more interactive and selective.

Analyzing the interaction data and understanding of user engagement from these online learners can potentially benefit both the lecturer and students when they use video as a lecture tool [18, 20, 21]. On the one hand, from the perspective of the video viewer, the watching peak in a video indicates the segment around the peak could be a major concept or a confusing part of the lecture. On the other hand, from the perspective of the video owner, user engagement raises awareness of video production methods to improve the quality of lecture videos. For example, shorter videos are found much more engaging, and the recommendation for making videos is that instructors should segment videos into short chunks, ideally less than 6 minutes [18] . Lecturespace [20] provides an interface using data-driven approaches to help learners understand online lecture videos effectively. J. Kim et al. [21] studied what commonly causes of interaction peaks in online lecture videos and showed that the peaks could provide a better understanding of video learning for both the lecturers and students. Thus, user engagement can be helpful for not only the viewers to interact with videos but also for the video makers to understand the watching pattern of users.

In order to measure user engagement and understand how it affects user interactions with video, the first thing is to collect user interaction data, then interpret and analyze the data, and finally provide an interface to interact with the data. Different strategies have been approached in recent research works for these three stages.

2.2 User Engagement Data

User interaction data collection has various approaches. Previous research has focused on two levels of interaction data: video-as-a-whole and segments within the video. Collecting data from a full-length video can help video makers understand the different statistics from different videos, whereas collecting data from short segments uncovers the most interesting or important clips in a video. From the video-as-a-whole perspective, user engagement is measured based on the entire video. For example, YouTube [11] takes likes and dislikes as a part of the metric for user engagement measurement. Mongy et al. [31] employ the average watching time per video, the average number commands per viewing and number of viewing sessions. On the level of segments within a video, user engagement can vary from different segments in a video due to the importance or the degree of interest of the parts. LectureSpace [20], built on the edX platform, collects the learner navigation events second-by-second including play, pause, and navigation to a specific point.

Interaction data can be captured both implicitly (e.g. [20, 21, 49]) and explicitly (e.g. [34, 26]). Implicit user interaction data includes the click based events from a basic web video player. For example, mainstream online video platforms (e.g., YouTube, Vimeo) typically provide a basic set of video player click based events: play/pause, volume change, playing progress, forward/backward navigation, and sometimes enable fullscreen. YouTube supports speed rate change which allows a user to control the playing speed of each video manually. These click events can be collected implicitly for user engagement analysis. Many existing video systems collect data explicitly by asking users for feedback. For example, the work presented by Olsen et al. [34] provides a rating metric in the video system allowing the viewers to rate the video clips for video summarization. YouTube provides thumb's up/down for users to like and dislike the video. Both implicit and explicit logs can be used to profile user watching patterns with videos.

2.3 User Engagement Analytics

After collecting user interaction data, it can be challenging to convert received interaction data to a meaningful understanding of user engagement. It can be difficult, for example, to ascertain whether the watching peak of a video means an important concept in a lecture or just a splendid joke made by a professor. Previous research did little work with the correlation between the interaction data and the real meaning of the data.

The approach employed in LectureSpace is purely based on the user interaction events with the video system and the approach used to measure user engagement is fairly straightforward. In the approach, the number of interaction events happening to one segment is simply counted. TwitInfo [28], an algorithm that can automatically finds peaks from the tweets, is used to determine the peaks in LectureSpace. Mongy et al. [31] used a first-order non-hidden Markov model to profile user behavior. The actions including play, pause, forward, rewind, jump, and stop are considered for the model. To make a video recommendation based on user engagement with video, the work in YouTube [11] examines explicit and implicit data. The explicit activities include rating a video, favoriting/liking a video, or subscribing to an uploader, and the implicit activities include user watching activities such as click play, change playback rate, and seek.

The interaction peaks over the video can be learned based on the counts. However, it might not be enough to understand the watching peaks correctly. It is because many other implications can be gained from user interaction that impacts the measurement of user engagement such as playback rate change from the user. When the playback rate changes, interpreting user engagement should involve the user's habitual playback rate of watching videos and the video's normal playback instead of its default playback rate. In addition to the event of playback rate change, computing user engagement can be possibly refined from other important interaction events such as volume change, screen size change, fast forwarding, and rewinding. In this research, we consider broader contextual data, such as the playback rate change for computing user engagement.

A complementary area of research that employs analysis of user interaction logs is video summarization, which focuses on editing the video into a concise summary rather than interaction with the existing video. For example, ShotRank [49] proposed to use the viewers' browsing log to measure the subjective interesting level and the importance of each video shot. In ShotRank, six event types are captured in their specific video browsing system. More interestingly, each event type has a different weight when it is used to compute the subjective importance of the shot. For example, play/pause is counted as 0.5 and jump is considered as 1.

2.4 User Engagement Interaction

After analyzing the data, there is a need to visualize the data to the video viewers and provide features to users to interact with the visualized user engagement. For example, the thumb up and down in YouTube [47] are not only are the public indicators of the video-as-a-whole user engagement (e.g. like or dislike the video) but also used for people to vote up or down. As a video owner, analytics data related to user engagement such as views and audience retention are also available in YouTube. However, it is rare to see any existing video systems provide segment-level user engagement publicly for everyone or even for a specific group other than the video owner alone.

A commercial video hosting service, Wistia [46], is an exception. The owner of the video hosted in Wistia can see heatmaps of video views. An example of Wistia heatmaps is shown in Figure 1. The two heatmaps are from two watchers individually. In the heatmaps, green indicates the video was watched once, yellow, orange, and red indicates the video section was re-watched, an empty section indicates the part of the video that was skipped. In Wistia, they provide video statistics to the author of video content.

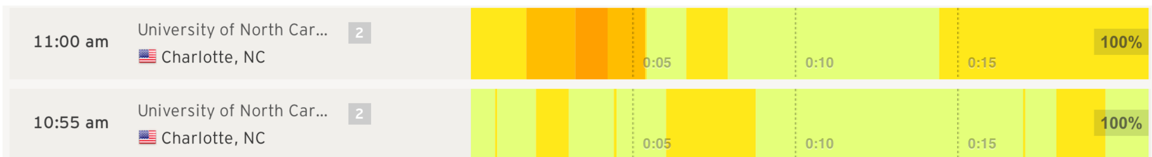


Figure 1: Wistia heatmaps

The video player in LectureSpace [20] is a non-commercial yet unreleased product. A non-liner 2D video timeline is embedded in the video player that is built on edX platform. LectureSpace employs user navigation event counts to display the collective interaction traces from all learners. The navigation events include pause and jump to a specific point. Interaction peaks, as detected by the TwitInfo algorithm [28], are highlighted in magenta and span a wider region that other points in their 2D timeline that is different from traditional video timeline. Figure 2 shows the LectureSpace 2D timeline as reported in [20], and Table 1 highlights the major differences of key features between Lecturespace and Wistia.

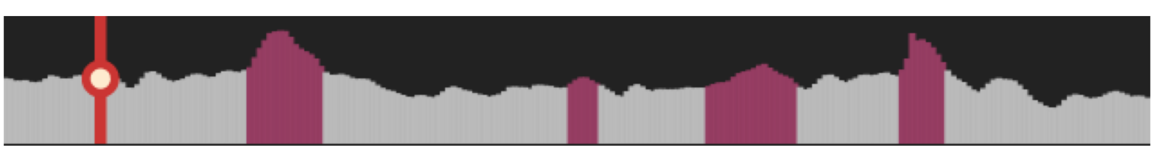


Figure 2: Lecturespace timeline

Table 1: Overview of video engagement measurement

	LectureSpace	**Wistia**
Event	play/pause/navigation	watched/re-watched/not watched
Visualization	2D with one color	1D with different colors
Who can see	everyone	video content owner
Log time unit	second-by-second	unknown

2.5 Evaluations for User Engagement Usage

Depending on the research goals for understanding and using user engagement, the evaluations can be conducted in various ways. To compare two different interfaces, a within-subjects or between-subjects user study is often conducted with a comparison experiment by using the tool with and without the newly developed features [14, 20]. For example, in the research work related with user engagement measurement [20], the researcher compared the video player with and without the developed extra features. To understand the collected quantitative data related to user engagement, experts usually are involved for the interviews [18, 23]. For example, the experts in the study of audience engagement with a dance performance are choreographers and directors [23].

More specifically, the experimental evaluation of a related work, LectureSpace [20], tries to assess the improved navigation features integrated with user engagement visualization for an online learning video player. In their within-subjects study design, participants are divided into two different groups and use the video player with and without the new features respectively. Since the new prototype aims to provide better interaction for information findings related with video navigation, the work compared the time used by the participants for completing the evaluation tasks, analyzed the perceptions of the visualized user interaction data, and tested perceived usability. In addition to the implemented prototypes, the materials used for the study in the research include surveys which helped the researchers understand the experiment results and how people like their new designs as well. That research tried to study how the visualization of user interaction with video affects video viewers from many aspects including visual search, problem search, and summarization. No significant difference was found in that research in the performance of the task of problem search from a within-subjects study, the methodology of part of the experimental evaluation, however, can be adopted because our work shares part of their design goals.

To understand the impact of video production for user engagement with MOOC videos, Guo et al. [18] conducted an empirical study with their quantitative findings from the user interaction data collected from edX courses and the qualitative feedback from the interviews involved with experts. First, they used two proxies to define *user engagement*: 1) the length of time that a student spends on a video, 2) the problem attempts from a student. They analyzed the data logged from edX coursework to measure user engagement based on the two proxies together with video properties such as video duration, type, speaking rate and lecture style. Six edX staff who were involved in video production were interviewed with the data analytics. At the end of the research, they conclude a list of recommendations for lecture video production.

Latulipe et al. 's work [23] is not directly related to video. However, it aims to understand user (audience) engagement with art performances. The audience engagement was linked with the biometric data and visualized as a line graph underneath a video player. To evaluate the collected biometric data and the line graph, they conducted an exploratory study and an empirical audience response study. The exploratory study was designed to explore the understandings of user engagement data from the experts, while the public response study was to examine the correlation between real user response and the captured biometric data.

The experimental evaluation with comparisons is helpful for the evaluations for this work because we will add extra features for a video player and want to investigate how the newly added design feature impact user interaction for performing goal-directed tasks. The methodology that used interviews from domain experts is useful as well. It can help to understand the user perceptions on a new prototype from the expert's view.

2.6 Heatmap in Data Visualization

Leveraging user interaction history is not only for video watching [46] , but is widely used in other fields such as web browsing [9, 32] and software usage [30]. Researchers from these fields also put effort on visualizing the captured interaction data for users to interact [30, 9, 32] or for their research [15, 6, 10]. Heatmap is commonly used as a visualization tool for data analytics. It intuitively shows the traffic of user interaction and provides an interactive opportunity for the users [30] as well.

One application of heatmap for data visualization is to understand user's view behaviors. We have seen that Wistia provides a heatmap for showing a user watching history of a video and it clearly shows what segments have frequently been or rarely watched. The heatmap visualizes the viewing behavior of users for a page recognition task [6] and helps the researchers understand how human attention goes when people browsing a webpage. In the study of information usage in web search [10], the heatmap visualizes the user's web browsing history from the participants' informational finding tasks for a list of search results. Both of the last two user studies for learning from user interaction with web browsing use eye-tracking as a solution to track user's watching pattern. Although the eye-tracking approach is not used in our research, the heatmap in our works serves one common purpose which is to help us understand user watching pattern on videos. We instrumented a logging server that tracks user interactions with online videos and visualize the logs. Edmonds et. al [15] presented a web in-page logging technique and claimed that it can guide designers to make design decisions for building the dynamic web. The logging mechanism tracks user interaction data with the web including every possible type of interaction: mouse click, mouse movement, key press, scroll, page leaves and window resize.

A second application of heatmap is that it provides better support for goal-directed tasks. The heatmap feature in Wistia is intended to help the video owners understand how the video has been watched by the viewers. Patina [30] collects software application usage history data and provides a dynamic heatmap overlay on the active application. It helps people quickly find commonly used menu items and complete related tasks.

2.7 Conclusion

This chapter provided a review of related work in video interaction, focusing on aspects related to user engagement analysis, visualization and the evaluation of its application. While video-as-a-whole user engagement analytics have been extensively studied, comparatively little research work focused on segment-level user engagement. In the segment-level user engagement applications, the approaches used for user engagement measurement are based on finding highly-interacted segments instead of highly-played segments. Moreover, many online video applications such as Wistia provide user engagement data but only make it available to video creators or editors rather than supporting extra interaction features for end users. To bridge the gaps between user engagement analytics and its presentation, various visualization techniques have been used in previous research such as heatmap, and line graphs. Finally, this chapter provided an overview of previous evaluation approaches for understanding user engagement measurement as a foundation for evaluating potential impacts in user interaction with video.

CHAPTER 3: ENGAGEMENT BASED USER INTERACTION

As users interact with a video, they naturally generate a substantial amount of data about their use of the video content. Re-playing a particular segment multiple times, for example, could act as a measure of interest and engagement with that specific part of the video content. In this chapter, we propose an investigation of video interaction based on user engagement data targeting users of lecture videos. First, we discuss the general design goals for visualization and interaction with videos based on analysis of user engagement data. Second, we discuss the approach of how the interaction data is collected and how the collected data is interpreted for measuring user engagement. Third, we describe the interface design — a heatmap for displaying user engagement data.

3.1 Design Goals of User Engagement Visualization

Previous work of discovering user engagement focused in the domain of online education [20, 21] aims to help online learners quickly navigate over the video and help the lecture instructors make more engaging videos. Wistia provides a heatmap based on the individual user's watching behavior on a video. The events Wistia captured from users, are “watched”, “re-watched”, and “jumped”. The heatmap generated from those events, however, is only available for the video owner. Thus, the design goal of such a heatmap is for the video owners to make more engaging videos by understanding the users' watching patterns of the videos. Potentially, it can help the video owners make more engaging videos. Although there might be different interpretations of the user interaction data, the underlying goal of learning user engagement is to help viewers quickly navigate and access engaging segments in the video. The primary design goals in our work are listed as follows:

Providing quick access to what others frequently watched. Instead of watching online videos passively with traditional controls, this goal is to provide an informative watching experience based on the watching trace of other users. This design helps users watch and navigate the video with a visual cue from others. For example, a segment talking about an important concept can potentially be a frequently watched part in a video. If the segment has more views than other places in the video after the video has been watched lots of times, the segment could be identified as a peak which could be useful for future video watchers.

Supporting easy video search/re-finding Another purpose is to help video viewers quickly locate the needed content from the video with a visual cue of user engagement, especially when the video viewers are familiar with the content. Imagine a student is preparing for a final exam, the student would like to find out the important content from a lecture to review. Assuming video viewers are more engaged in the important clips in a video, there will be a watching (user engagement) peak in the video. As a result, the peak can strongly indicate the importance of different segments in a video, which can help the student re-find the important content quickly.

Supporting context filtering Because user interaction data can be collected from everyone in the video system and different viewers have different watching patterns, it is possible to study different watching behaviors from different groups of people. The ability to show user engagement from individual viewers, different groups of people based on their demographic information, from different time ranges, even from different videos, can be more helpful than only showing an overall user engagement from all users. For example, a professor might be interested in comparing the difference of watching patterns between international and domestic students.

3.2 Data Collection and Analysis

Each operation the user made on a web video player can be captured and saved. With the web technologies, it is not difficult to learn who interacted with what video, the action the user took and a timestamp of the event. For example, a typical interaction event is that a user (who) clicked play (action) on 4:30 in a video (video) at 10 o'clock on December 12th, 2014 (timestamp). In this example, "play" is the event that is triggered by one mouse click from a user. Other than the play event, there are more interaction data can be captured from most online video players, including pause, jump, browser window/tab close, volume change and playback rate change. However, the single click-based events are the actions from users. Since data analytics in this work focuses on the segment-level data and finding the *highly-played* segments from the data, we need to capture the interaction data in a way that can represent the segment states (playing or paused).

A variety of methods may be used to measure user engagement. A straightforward approach to is to aggregate the counts of the interaction events per second in a video. LectureSpace [20] only takes play, pause, and navigation (jump) into account for the valid interaction data for user engagement. Another consideration is to use the frequency of plays as the user engagement metrics for a particular segment. The heatmap of user engagement presented in Wistia is generated based on the counts of watches, re-watches, skips that happened to the video per second. The color of the segment in the heat map darkens as the count of re-watches in a segment increases. However, skip does not have any effect on the computation for user engagement in Wistia. Either of these two methods has its own disadvantages. On the one hand, having more interactions in a segment does not necessarily mean the segment is highly engaged by the users. For example, a learner has to pause the video for answering a question embedded in the video. After completing the questions, the user continues watching the video and plays the video at the time where the video is paused. So the pause event brings two interaction events and shows no connections with the user engagement. However, the events are used for the computation. One the other hand, counting the plays of the segment is essential to compute the play frequency of the segment, but this is a less accurate way to measure user engagement because a user can leave the video to be playing without watching it. Assuming the majority of the viewers watch the video while the video is playing, the user engagement is related to the frequency of plays. And this approach is used in this work.

In addition to collecting and counting the basic operations that happened to the video player controller, the actions may be weighted differently [49]. That is, a play/pause is counted as 0.5, and a jump is considered as 1 assuming the normal weight for a play is 1. Weighting actions naturally gives rise to research questions about the context and parameterization of those actions. For example, playback rate may be considered as a context for user interaction. A question to consider might be to compare a video segment that is watched by a user under a playback rate of 0.5 versus a playback rate with 1.5. Does slower playback rate mean that user is less engaged with this segment? Thus, simply counting the number of the basic interaction events may not provide enough granularity to accurately measure user engagement.

Therefore, beyond counting interaction events or the times of watched towards each play of segment, we propose to consider the context when the user interacts with the video. Many parameters can be considered in the context of a play of a segment. The context can include the playback rate, the user's normal playback rate of watching videos, the average playback rate of a video, and the volume of the play. For example, a user may slow down the playback rate when the segment is important. In this work, we propose a model to measure user engagement from collected user interactions data and instrument an online environment to run the experiments and test the model. Many options can be considered as the context, and we investigate these options first individually and then in combination.

To prepare the study for the heatmap of user engagement, collecting enough user interaction data for generating the heatmap is the prerequisite for the study. Even though a huge number of users have been enrolled with the popular MOOC platforms, the interaction data from existing platforms has not been publicly available. To overcome the problem of lacking user interaction data for this work, we need to access an online video platform which has a decent number of audience and allows us to instrument an event logger to capture user interaction data. Video Collaboratory (VC) [44] was chosen as the video platform because of following reasons. First, VC has a good number of users who are enrolled in several courses at UNC Charlotte. In many classes, students are required to use the platform to watch lecture videos before class. Furthermore, we were approved to embed the event logger and collect user interaction data. The logger was implemented to run in the background without interrupting the normal use and the logged user click-level history is saved to a database. Note that the data collection incorporates any video events triggered by user interaction, with the exception of some events that are automatically generated by the system to support collaborative features in VC. As with the heatmap, the event logger is implemented as a generic plugin for a web video player.

We would like to clarify the usage of VC in this research, in case the readers are familiar with this video platform. It is an online video platform that allows video annotations and supports collaborative work. Many features in VC are designed for collaboration, some types of user interactions can trigger video player default events. For example, clicking one comment automatically starts playing the video despite this click event is not sent from the video player control. By default, the click events from these extra features are ignored by the event logger as these events are not used for the computation of user engagement. Therefore, the only feature we need for our work is the video player of VC. Since the event logger is designed as a general plugin for any video player, it is easy for us to integrate the event logger with the VC platform.

3.2.1 Metrics

The concept of a *viewing session* is introduced to collect user interaction data. A session begins when the user first navigates to and starts playing the video. And a session ends when the user navigates away from the video, or the web browser is closed. This may also include a timeout for inactivity. In this research, user engagement with a video comes from all watching sessions from users and a session can have multiple plays or replays. The idea of sessions helps us analyze the data from different aspects such as users, video and watching time. Choosing an appropriate data granularity is critical for an automated data collection system [24]. In terms of the length of a segment, we divide the video into segments based on a granularity of every half second.

In order to measure the degree of hotness of a segment in a video, we introduce a metric to measure the degree of engagement (DOE) for a given segment. Let $E_{ui}(segment_j)$ denote the degree of user m's engagement of the jth segment of the ith play in a session. we define the degree of engagement $E_{ui}(segment_j)$ of a user u's one single play i on a segment j as follow:

$$E_{ui}(segment_j) = \sum_{i=1}^{n}(play_i \times weight_i) \tag{1}$$

where $weight_i$ is the weight of $play_i$. When only the play-count of each segment is considered, $weight_i$ of each play is equal to 1. $play_i$ is the user's ith time play of the segment, if ith play is skipped, $play_i$ is 0, otherwise, $play_i$ is 1. n is the total played times of that video containing segment i by user u. Suppose the total number of users who have viewed this segment is m. The aggregated degree of user engagement from all users $E(segment_j)$ can be computed in equation 2.

$$E(segment_j) = \frac{\sum_{u=1}^{m} (\overline{E_{ui}}(segment_j))}{m} \tag{2}$$

where $\overline{E_{ui}}(segment_j)$ is the average degree of user u's engagement over the segment i of jth segment.

The essential part of computing the degree of engagement is the computation of $weight_i$. Two main factors should be considered to weight a play: the average playback rate of the video itself and one user's normal playback rate of watching videos. Intuitively, the playback speed is a significant factor that impacts the importance of each play. A user may slow down the video play speed if the user thinks the part is important. By default, the video's playback rate is set to 1. However, it is common that an instructor in a video lecture speaks slowly. It may be because there is no instant feedback from students when the instructor is recording his lecture and the professor wants to slow down to have most viewers can follow the lecture. Or it may be because a part of the audience is non-native speakers, which is very common in MOOCs. As a result, the students who are familiar with video content or whose native language is instructor's language may want to speed up the video playback rate. Another important factor is the user's usual watching playback speed of videos. If a user was accustomed to watching videos with a faster playback rate, the default playback rate should not be 1 for this user. Therefore the user's normal watching playback rate should be considered as a factor that influences the value of $weight_i$.

Other factors can also be considered as contexts for computing the weight such as different volume levels and the video screen sizes. We are not aware of research on comparative or collective impact of such factors. Therefore, we propose to investigate an approach for weighting each play based on the possible contexts, to understand how these factors can affect analysis of user engagement. We will first focus on the factors of video and typical user playback rate and apply them individually in the computation of the degree of user engagement. After testing them separately, we will explore how they impact the computation of user engagement in combination. We note that contextual elements such as volume may present a particular asymmetric challenge, as volume may be adjusted both within the player itself and externally at the level of the computer system. Thus explicit volume changes within the player can be considered when present, but the context is weaker as a whole.

In a typical learning environment, it is likely that a student re-watches an important part in a video and slows down the speed of video playing the part. So, presumably, a segment in a video that was played at a slower playback rate compared to the normalized playback rate is more important than a segment that was played at a faster playback rate. The computation of the weight of the importance of the play should be inversely about the speed of playback rate. For a starting step, $weight_i$ can be just computed as follows:

$$weight_i = \frac{speed_{default}}{speed_i} \tag{3}$$

where $speed_{default}$ is usually set to be 1.

Overall, the degree of user engagement $E_{ui}(segment_j)$ is proposed to compute in three different approaches. First, the simple and straightforward one is to count the number of plays of each segment, that is, the weight of each is equal to 1. Second, only the average of playback rate of $segment_j$ is considered. That is, the $speed_i$ in Equation 3 is the playback speed of the segment. Third, only the user's normal watching playback rate is considered. To obtain this value, the standard watching speed is calculated based on all of the videos the user has watched.

3.3 Prototype Design

To study user engagement of video, we developed a high-fidelity prototype that works as a universal plugin for a web video player. For the purpose of collecting data and testing, it is embedded in the video player employed in the Video Collaboratory[9] (formally Choreographer's Notebook [44]). The design goal of the prototype is to explore user watching behaviors on a video and facilitate future viewer's watching process. In recent years, using heatmap to visualize data has been very common in software usage tracking [9, 32, 30]. Adopting this idea, we use the heatmap to visualize user engagement on a video. The main reason why the heatmap is chosen for a visual cue for user engagement is that it has a strong indication of the degree of engagement of a segment in a video. For example, the darker colors represent greater engagement levels than lighter colors. Since Video Collaboratory has been used in several courses taught in Department of Software and Information Systems in

[9]http://videocollaboratory.com

UNC Charlotte, we can potentially collect interaction data from several classes and expect the heatmap of user engagement can have different color ranges after a certain number of watching sessions. For the research of this dissertation, the interaction data were collected from two undergraduate classes across two semesters and one graduate class in one semester. In total, there were approximately 200 students enrolled in the classes. For further detailed statistic data for the videos we use for the user study will be presented in the experiment Chapter 4.

3.3.1 Prototype I - Heatmap of user engagement

The illustrative heatmap of user engagement is built as an additional bar that stays on the bottom in the video and provides visual cues of video watching behavior of users by differentiating along the dimensions of color and color brightness. Figure 3 shows the initial design of the heatmap. Two specific color scales are selected to illustrate here: green and red, though any color range could be selected. The green color indicates that the segment has not been played while the red color indicates the segment has been played. Any color between the two color ranges relates different levels of user engagement. When the video is loaded the first time, the entire heatmap bar is in green color. After a number of watches over the video, the red color begins to scale with the number of views for a given segment. If a segment has a higher degree of user engagement $E(segment_j)$, the red color in that segment is darker. The threshold of the level of darkness of the red color depends on the number of views. Given a set of degree of user engagement data for each segment, $E(segment_1), E(segment_2), \ldots, E(segment_n)$, we first discard some data with extremely high (or unusual) views. For example, an unusual number of views was found in one particular spot which has more than 1000 views compared to other segments only have 200 views in average. It could happen due to the platform we used to collect data allows loops over and over again on the segment level. So the high views could be caused by a user who forgot to stop looping the video. After the data removal, we normalize the set of the engagement data, and generate the heatmap based on the normalized data. Figure 3 presents an example of the heatmap implemented for an online lecture video which was assigned to about 80 students to watch. It is clear to notice that the segments at the beginning and some other places are highly engaged by the viewers because the segments have the darker red colors in the heatmap. Having more plays at the beginning of a video makes sense due to the fact that people naturally click the play to start the video. Note that the spots with darker colors represent higher viewing rates than the segments with lighter colors. It may imply the hot spots are important content but it is not guaranteed since many other factors (e.g. confusing content) can cause a high volume of views.

Overall, the heatmap we have built for this research is designed to facilitate video viewers to be able to identify highly watched segments on a video and potentially help them more effectively watch the video and complete analysis work related to video materials. To create the heatmap, we collect user interaction data through several semesters and then follow our proposed process to compute and degree of engagement for each segment. The collected data represented as a heatmap are used to conduct our user studies to evaluate the interface and investigate how the new views are affected by the presence of previous users watching activities. Further discussions are

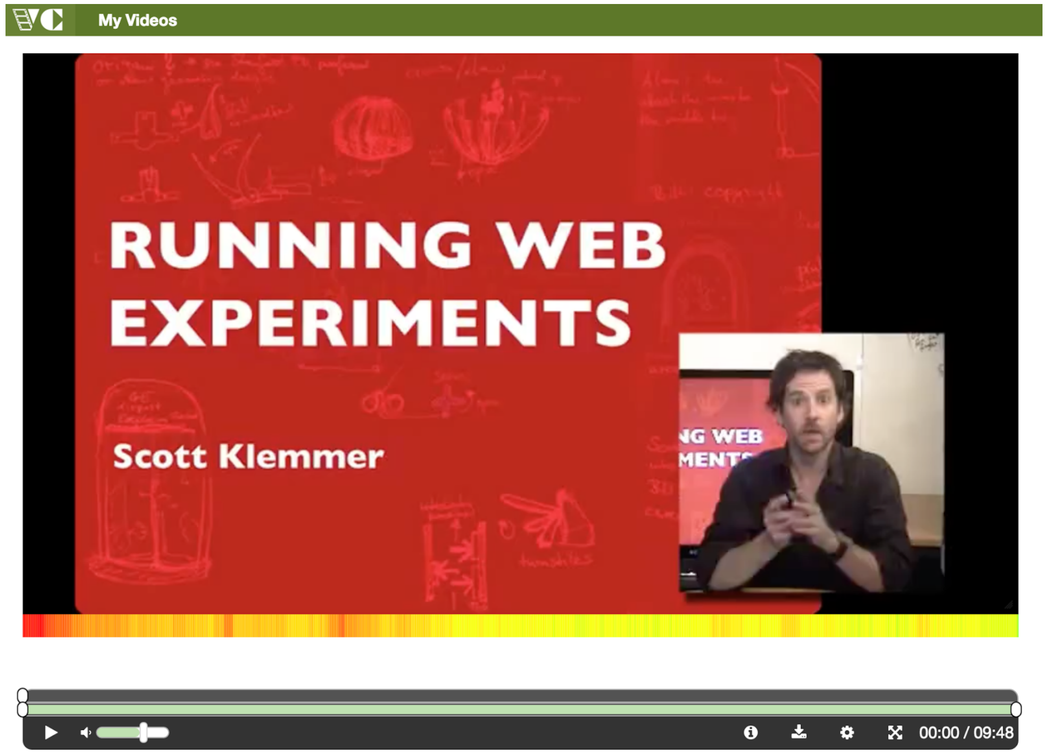


Figure 3: Heatmap of user engagement

described in later chapters (Chapter 4 to Chapter 8).

3.3.2 Prototype II - Multiple heatmaps in Video Collaboratory

The interface in Figure 3 demonstrates a single heatmap of aggregated user engagement data on a video. More than showing one heatmap all the time, users such as class professors or the lecture video producers may also be interested in comparative viewing patterns. For example, a user may want to see his personal engagement with the video and compare his watching pattern to the overall user engagement. Or the user may want to find out how the user engagement evolves as the time goes, for example, from weekly to monthly. Or a lecture video producer may want to improve the video and to find if an unclear part is causing a problem for the students with poorer performance. Then she may want to know whether a different behavior may imply different performance levels by comparing the user engagement of students with better performance and poorer performance. In other words, if it is possible to link the audience demographic data with the interaction data, multiple heatmaps can be employed for comparative analysis.

Essentially, the heatmap combinations are the results of data filtering of different inputs such as the audience demographic data (e.g. international and domestic), students class performance data (e.g. good and poor), different time (beginning and end of a semester), and so on. Figure 4 shows the interface setup. On the left of the figure, there is a video player with embedded heatmaps. On the right, a time range selector is provided for the user to set the start and end time. The design of the interface is flexible to add multiple heatmaps as many as the user wants to. The example in this figure uses two heatmaps for the purpose of demonstration. The two heatmaps, in Figure 4, are the user interaction results of the video based on two different time selections in the video system. The one on the top is generated from the watchers who had watched the video at the beginning of the time when the video was assigned to watch; while the second one represents the watching patterns from all the students over all the time. We can notice that the color evolves in areas of the heatmap over time — the earlier (top) heatmap contains many lighter colors, but these segments evolve into darker hues later (bottom) given more views on the same parts. An excellent application for this type of combination might be for comparative analysis. For instance, a professor would like to compare how students watched the video differently in the different time of the semester such as the beginning of the semester versus the time for the mid-term or final exam.

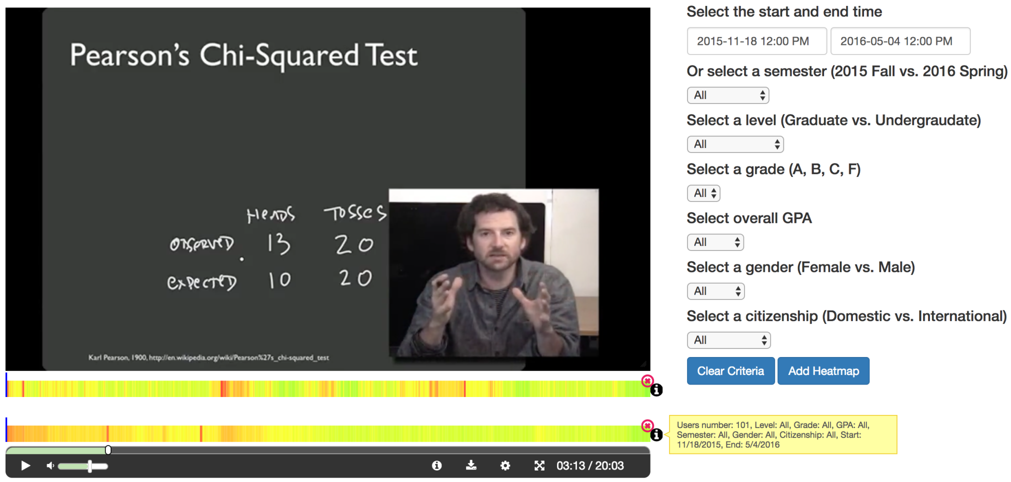


Figure 4: Multiple Heatmaps for Comparison

Similarly, many other combinations of user engagement data sources can be placed in the same way as Figure 4. For example, a heatmap with the data came from a list of students who had good performance can be placed on the top of another heatmap generated based on a group of students with poor performance (e.g. Grade or GPA). In this research, we built the interface based on the student's demographic and academic data including student's grade on the course, GPA in the semester when they registered the class, gender, and citizenship. Citizenship is one of the best ways we could get to determine the student is a domestic or international student. With the listed combinations, the user can place as many heatmaps as the user want to. There are several different dimensions to explore in comparative heatmap analysis, determined primarily by (1) the cohort of the heatmap (e.g., individual student, class section, demographics, etc.) and (2) the timeframe of the data analysis (e.g., within vs. across course offerings).

The comparative heatmap interface incorporates several additional features. First, each individual heatmap retains all of the functionality of the single heatmap which is introduced in the single heatmap interface. Second, a button with an “i” icon is a button where the user can find the selection for generating the heatmap by hovering on it. Third, a “clear criteria” is added to reset all the filters conveniently.

3.3.3 Summary of prototypes

In this section, we present two interfaces involving the heatmap that is a data visualization of user interaction with the video. The first one is a single heatmap that shows the watching activities from all of the previous uses. The aim of this heatmap is to enable future viewers to learn from the video more efficiently. For example, the heatmap might help the learners identify what parts are the most interactive. Second, we demonstrate an interface that has a high flexibility to enable multiple heatmaps at a time. The source data of the heatmaps varies from the demographic data of the previous audience to the class performance of the students. The goal of the second interface to evaluate how multiple heatmaps affect the comparative analysis. In any case, we should emphasize that the two prototypes can be used in any online video player as long as there is user interaction data available. The two interfaces provide an important opportunity to advance the understanding of the application of user interaction data analytics.

3.4 Conclusion

Many approaches for computing user engagement have been presented and are aiming to increase the accuracy of the degree of user engagement, however, some of them actually compute the degree of user interactions. Consequently, the user engagement findings are user interacted segments. To overcome the drawback, we first convert the user interaction data to every valid segment within a video and associate the playing context with the segment play. In this chapter, we proposed to compute the degree of segment-level user engagement considering the context of video playing such as the playback speed of the video and a user's normal watching speed. Moreover, two high-fidelity prototypes have been developed for the visualization of user engagement in videos. In the first interface prototype, only one heatmap of overall user engagement is available. To give more data visualization available to the users, more heatmaps can be added for comparative analysis. To evaluate the developed interfaces, the click-level data including play, pause, navigation, volume change, and playback rate will be collected. A detailed study design for the evaluation of this heatmap of user engagement and the analysis of the user study results are presented from Chapter 6 to Chapter 8.

CHAPTER 4: RESEARCH DESIGN

In the previous chapter, we introduced two heatmap interaction designs to address different types of goal-directed tasks: content understanding, information finding, and comparative analysis. Our goal in this research is to investigate how interaction approaches that apply user interaction data analytics to enable active video analysis can impact people's video watching, navigation, and analysis with a single video. To this end, we have conducted a series of user studies to evaluate our research prototype and address a set of underlying research questions. We posit that (1) providing a visualization of engagement data from all users to video viewers can potentially impact future user's viewing on a video for goal-directed tasks — especially the tasks for information finding and content understanding; and (2) providing flexibility of data filtering can help the video owner / course instructor better understand the watching behavior of the subscribers of the owner's videos. For the evaluation of the first interaction design with one single heatmap, we conducted an in-class pilot study, an in-lab user study and an online user study deployed on Amazon Mechanical Turk. To evaluate the second interaction design which enables the existence of multiple heatmaps, a user study employing a think aloud protocol was conducted with course instructors.

For the user study evaluations, the goal-oriented task context was selected as learning within an education environment. User interaction data was collected from educational videos and the participants were either college students or instructors. While different interaction behaviors may arise on different types of videos (e.g. entertainment) or different participants from a different domain, we believe the design of study itself is a good representative for task-oriented video interaction and is replicable to other video domains related with video analysis for goal-directed tasks. In this chapter, we introduce the research questions that guide the study design, hypothesis testing, and result analysis. Second, we present the hypotheses for the research questions. Finally, we describe the entire process of the user study in detail and present a pilot we conducted.

4.1 Thesis Statement

We propose to address the general research question of how novel approaches to interaction with online video can better support end users for active, goal-directed use of video content. To investigate this research question, we focus one specific probe into the space of video interaction, evaluating interaction interfaces based on user engagement analytics. To understand the impact of interaction designs in these contexts, we will conduct user studies evaluating the following thesis. *The design features of the user engagement heatmap provide better task support than standard online video tools for user interaction and active consumption of online videos in goal-directed tasks such as information finding or comparative analysis.*

4.2 Research Questions

In a typical online video learning environment, two parties are involved: learners and instructors. In order to fully evaluate our prototype, we consider both of these groups representing different interaction use cases. From the learner's perspective, they would use video as a tool to search information and expedite the learning process. From the instructor's perspective, they would use video engagement for comparative analysis among students and self-critiquing teaching materials / methods. User studies are conducted to address following research questions As a consequence of two different groups of subjects, we describe two sets of research questions based on the two groups.

- **Learners Group**
  - **RQ1** How does the heatmap of user engagement impact video learner's understanding of the lecture? (content understanding)
  - **RQ2** Is the heatmap control effective for learning tasks based on video material? (effectiveness)
  - **RQ3** How does the heatmap impact user's watching pattern? Do users use a different way to navigate the video, particularly when they search the content from the video? (watching patterns)
  - **RQ4** How do the video viewers (learners) interpret the heatmap of user engagement in a lecture video? (user perceptions)
- **Instructors Group**
  - **RQ5** By exploring the interface of multiple heatmaps, how do the class instructors or video owners interpret the heatmaps of user engagement with online videos?

– **RQ6** How does the heatmap of user engagement influence the class instructor in self-critiquing the teaching method?

4.3 Hypotheses

This section contains the hypotheses that guide our study designs and later analysis. Each hypothesis will be tested by analyzing collected user study data and the full description of the analysis is presented in the section of results analysis of each study. The hypotheses for the evaluation of the single heatmap interface are based on a comparative user study between a traditional video player and a video player embedded with a heatmap representing user engagement information.

1. User performance of learning (video content understanding) will be different between a basic video player and a video player with a heatmap. More specifically, compared to a traditional video player with basic controls, we hypothesize that,

   - $H_{1.1}$: With the video player with the heatmap, the participants will perform better on the quiz which is taken after watching the video. Let $CorrectnessRate_{Heatmap}$ denote the rate of correct answers from the study with the heatmap, and similarly let$CorrectnessRate_{NoHeatmap}$ denote the rate of correct answers from the study without the heatmap, $H_{1.1}$ is:

     $$H_{1.1}\text{:}\ CorrectnessRate_{Heatmap} > CorrectnessRate_{NoHeatmap}$$

   - $H_{1.2}$: With the heatmap, the participants feel they can understand the content better.

- $H_{1.3}$: With the heatmap, the participants are more confident to answer questions.

Testing $H_{1.1}$, $H_{1.2}$ and $H_{1.3}$ will answer the research question about how the heatmap impacts video content understandings (RQ1).

2. The video player with the heatmap is more perceivably effective to use. Specifically, the video player with the heatmap helps for learning lectures more than the video player with basic controls.

   - $H_{2.1}$: When the users are learning from lecture videos, the video player with the heatmap is more useful than the video player only with basic controls.
   - $H_{2.2}$: By using the video player with heatmap, the participants can more easily find the information needed for quiz questions that are related with the video content.

   Testing $H_{2.1}, H_{2.2}$ is to investigate perceived effectiveness of the heatmap for learning tasks based on lecture videos materials (RQ2).

3. Regarding user navigation / watching patterns, we hypothesize that:

   - $H_{3.1}$: When the heatmap is available, more users will use heatmap than other navigation methods to watch the video for the first time.
   - $H_{3.2}$: In the self-reported data, a significant proportion of participants will use heatmap for completing tasks related to video content (e.g. quiz).

An example of other navigation methods is a common pattern that a user just "sit, click, and play". Testing $H_{3.1}, H_{3.2}$ is to address the research question concerns about user watching patterns (RQ3).

4. As part of user perceptions of the heatmap, we hypothesize that:

    - $H_{4.1}$: The video player with the heatmap is more enjoyable (less frustrating) than the video player with basic controls (RQ4).

Independent variables:

- Online video player (2 treatments): a basic video player and a player with heatmap(s) of user engagement.
- Tasks: Watching the video, finding content from a video, answering questions related with video content, and completing surveys.

Dependent variables:

- Task performance: the time used for completing the tasks, the accuracy of the task performance (the quiz performance).
- Perceived effectiveness: easy to find information, useful for learning.
- Subjective satisfaction: easy of use, easy for learning, frustration.

4.4 Overview of user study design

The education domain was selected for the study of the interaction design for the following reasons. First of all, we have reasonable access to online lecture videos, user interaction data with videos, and users (students) demographic and academic data. Access to the user population and interaction data were a straightforward fit for development. Second, in education domain, the class instructors who have flipped classroom experience or use videos as a tool in class are reasonable candidate users of the multiple heatmap interface. For example, they might use the tool for comparative analysis based on different watching behaviors from different students.

The heatmap of user engagement is a data visualization of the watching behavior of users. Our user studies test how the new heatmap impacts user interactions for goal-directed tasks: content understanding, information finding, and comparative analysis. First, the single heatmap interface provides a visual cue for the learners to watch the videos. We evaluate how this additional feature of the video player affects people when they are actively learning from the video. Second, the comparative heatmap interface provides the availability of engagement data filtering designed for the video owners or the class instructors. Testing this filtering functionality helps us understand whether multiple heatmaps of user engagement can help video analysis for comparison. As a result, two groups of participants are involved in the user study: a group of students and a group of class instructors.

In the group of learners, we focus on user content understanding, information searching, perceived user ability and watching patterns. First of all, searching information from the video happens frequently when the learner watches the video for the second time and wants to skip non-important content. For example, a student wants to find a solution for an assignment or use the lecture videos for the final exam review. Thus part of the evaluation is to test whether the heatmap interface can better help the learners find needed information. This type of experiment is similar to the evaluations of problem search and visual search in work [20]. Second, being able to quickly find the accurate information can save time for the user. However, it should not compromise user's learning quality. As a result of this, in addition to testing the feature of searching information, we test whether the learners can better understand the video given the presence of the heatmap. Third, our proposed user study evaluates usability of the additional feature for showing segment-level user engagement with a video. Usability assessment, in fact, is very commonly used to evaluate video related interfaces [14, 26]. Lastly, we are interested in finding out whether the presence of the heatmap of user engagement affects their watching patterns.

In the group of instructors, the primary of purpose of the user study is to evaluate whether multiple heatmaps better support comparative analysis and help self-critique in teaching methods. When instructors review lecture videos, it is potentially useful to analyze how student performance correlates with watching activities and engagement across different groups of students. Viewing, interaction, and engagement data generated by the class students can be linked to their demographic and academic performance data for analysis. Then the multiple heatmap interface could provide a lot of flexibility for contextual filtering. That is, given our multiple heatmap interface, the professor can flexibly explore different criteria — adding and removing different heatmaps for comparison. For example, the instructor can check how the watching patterns (the heatmaps) evolve as the semester goes. After trying different combinations of heatmaps, we would like to see if the multiple heatmaps impact the teaching methods of the instructors.

For both of the two groups, a common interesting research question is to understand user perception of the presence of user engagement heatmap when the users are watching the video. The evaluation involves users surveys and interviews with the domain experts. To understand user perceptions with the visualization of user engagement, post-experiment surveys will ask related questions regarding easiness, enjoyment and so on. The domain experts in our study are the class instructors because they know the lecture content very well. Similar to the empirical studies in work for the understanding of the audience engagement measurement [10] and the understanding of user engagement on online education videos [18], interviews with domain experts are used in this study as well.

4.4.1 User interaction data for the heatmap

The automated logs can provide valuable information for the understanding of the usage patterns [24] and the interface evaluation. As we discussed in Chapter 3, user interaction data was collected through an event logger embedded in one online video platform called "Video Collaboratory". For the study purpose, we enhance our event logger to capture not only the user interaction data happened to the video but also the interaction data with the heatmap.

Our user studies employ baseline video interaction data that was collected across several semesters prior to the studies. The interaction data used for this research primarily comes from students interaction with videos in the Human Computer Interaction course at UNC Charlotte. User interaction data was collected across the 2015 Fall semester and 2016 Spring semester course offerings. After collecting the baseline interaction data, we computed measurements of user engagement and generated the heatmap. To conduct the user study for the evaluations with valid heatmaps, we first examined a random sampling of heatmaps and found that most heatmaps across different videos showed clearly different "temperature" indications as expected.

It is worth noting that participants in the user study also used a clean version of VC that only has the basic control of the video player. That means the extra features are also eliminated in the user study. Features, like adding a comment and sketching on the video, are not available in the user study. When talked about the mechanism for collecting user interaction data through VC, we mentioned that we only collect user interaction data with the video player not the from collaborative components from VC. Taken together, to run the study clearly and accurately, only the events triggered by the primary player control are used for user engagement computation and the study.

4.4.2 Single heatmap user study design

We evaluate the impact of the heatmap in at least three aspects from the users: content understanding, the effectiveness of learning tasks, watching patterns and user perceptions. The approaches we use for the evaluation are mainly: (1) a series of survey questions to understand how the heatmap impacts learning lecture videos (RQ1, 2), (2) the correctness of answers to evaluate the performance of video understanding (RQ1, 3) the frequency of heatmap usage to assess how often the heatmap was used by the participants when they watched the video (RQ3). Other than the quantitative findings from above questions, the qualitative insights about the interpretation of the interaction data will help to answer RQ4.

A within-subjects design was selected for the study, with recruited participants divided into two groups. Each group used both of two different interfaces: a primary video player and a video player integrated with the heatmap of user engagement. The advantages of the within-subjects design are that fewer participants are required, and the performance is compared to the same group of people with different conditions [24]. We select different videos for different interfaces (conditions) for the participants to perform the same task. For the purpose of minimizing the ordering effect in a within-subjects design, we counterbalance the appearance ordering of the two conditions in the two study groups. To ensure the consistency of look and feel of the two interfaces, we use the same video player layout for this study. The functionalities for the purpose of collaboration in Video Collaboratory are eliminated from the study since we do not evaluate these features in this research. The learners perform two types of tasks based on the video materials: content understanding and information finding. The tasks are commonly used for the video interface evaluations in the literature [20, 14, 12]. After they finish the tasks, they are asked to fill out surveys related with the heatmap. The following describes the designs for the two tasks.

- **Video content understanding**: This task involves the content understanding of the video and emulates a viewer to learn from a lecture video. In this task, each participant watches one video with the heatmap and one without the heatmap, then complete the quiz related to the lecture content. How the user performed in the quiz can be linked to their understanding about the video content. Another important thing is that adding the heatmap to the video player should not impact the users negatively. Therefore, through the study, we also measure user's confidence for answering the quiz questions and how they feel about how much they understand the video. Overall, the purpose of this task is to help us understand how the heatmap impacts learner's understandings on the videos. (RQ1).

- **Visual search / Content re-finding**: Re-finding or searching content from a video often happens when a video viewer who is familiar with the video content wants to find the right content from a video. For example, a learner might only want to review the important content when he is preparing for his final exam. The participants in this task should be familiar with the video content or have watched the video. To meet this requirement, we allow the participants to re-watch the video as many times as they want, particularly when they are answering the quiz questions. In this task, we place the target questions around the peak and non-peak zones. This task emulates a scenario where a user re-finds a familiar video content from a video. The evaluation of the task helps us understand how the heatmap affects user's information finding and learning effectiveness (RQ2, 3).

4.4.3 User study location

There were two places available for us to deploy the user study. First, we could conduct the user study in class as part of a week assignment to the students. In this setup, the participants are the registered students in the class. Second, we can recruit a group of participants from campus and take the study in a controlled laboratory setting.

To begin with, it is feasible to deploy the user study in a class where Video Collaboratory is used as an online learning video tool. VC has been employed in Human-Computer Interaction class every semester in our department. Moreover, the lecture videos are same, and the class requirement is consistent in every semester. The HCI class requires all of the enrolled students to watch lecture videos as part of their weekly assignment. Taking advantage of the platform, we have collected user interaction data across the 2015 fall semester and 2016 spring semester. Consequently, we could generate the heatmap based on the watching history of the students from previous semesters for user study testing in a current course offering. If the study is taken in class, we will assure the class will have the minimum interruption from our study. The main benefit of running the user study in this way is that we can gain reliable user study data from the participants because the students have to watch the videos anyway. The only extra work for the students to do is to fill out the survey for our study. The downside of this approach is that the participants can not receive direct feedback from a study supervisor in case of questions or issues in performing the study tasks.

Alternatively, we can recruit participants from campus students and gather them to a laboratory to perform the tasks. It is feasible because the selected videos for the study do not require the participants to have prior HCI knowledge. Students who have HCI background are ineligible for the participation.

4.4.4 Multiple heatmap user study design

We use this study to evaluate the impact of multiple heatmaps for the class instructors. We are interested in how the class instructors could interpret the heatmap (user perceptions) and how they could use the heatmaps as a means for comparative analysis (contextual filtering). The study adopts the "think aloud" method, and it is conducted in a semi-structured session, where the experimenter asks pre-determined questions during the session in case the participants do not think out. The experimenter followed up new topics pointed out by the participants. The study is designed to involve the class professors only and is conducted in a laboratory setting.

The responses are collected throughout the entire study. Before the study, we asked a preliminary question to obtain the respondent's first thought about different watching patterns. Doing that is to minimize the impact of our interface in making the choices for the comparison. During the study, the computer screen and participant's voice are recorded. After the study, the participants are interviewed by a set of predefined questions. All of the results collected from the study will be used for later qualitative analysis. We code the data and find the emerged pattern from the data.

4.5 Conclusion

In this chapter, we begin with the research questions that guide this research. The research questions concentrate on user's content understanding, effectiveness, watching patterns and perceptions. In general, we hypothesize the heatmap would be useful for learning from lecture videos and more people would be interested in using the heatmap instead of traditionally "click play, sit back, and watch." More fine-grained hypotheses are presented in section 4.3 and testing such hypotheses can directly answer the research questions in this work. To prepare for the hypothesis testings, we conduct a series of experiments to evaluate the interface to collect experimental data. In the first experimental method, the participants perform goal-directed tasks including learning from the lectures, understanding the content, finding information from the video and completing the quiz. The survey data from the study will help to understand user perceptions about the heatmap. All of the study data will be used for the analysis. The second experimental task for comparative analysis of user engagement will be evaluated by the class instructors. The experiment is to identify whether the multiple heatmaps provide better support for comparative analysis on user watching patterns.

CHAPTER 5: PILOT USER STUDY

In the previous chapter, we described the general user study designs for the single heatmap and multiple heatmap interfaces. In this chapter, we provide user study detail for the single heatmap interface and describe the pilot user study we conducted for the evaluation. We describe the issues that were found from the pilot study and how these were addressed.

5.1 General procedures

All participants in the study are equally divided into two groups, and each group watches the same two videos. The only difference is that the order of the heatmap appearance is counterbalanced in the second group. The following description lists the procedures for the group of participants who watch the video with heatmap first.

1. The participants fill out a survey about their demographic information and related experience about learning from online videos. The experimenter should first request the consent from the participants.

2. We briefly introduce the goals and the procedure of the study. The participants are informed that they will take a quiz after watching each video and their answers will be recorded for our analysis. The participants are allowed as much time as they would like to try with each interface.

3. After the participants are familiar with the interfaces, the experiment session

begins. After watching each video, the participants will be asked 3-4 questions that are related to the video content. In the first video session, because they experienced the heatmap interface, they will be asked specific questions about the heatmap usage. In the second watching session, the participants will watch a different video with the basic video player. All of the participant's interaction events will be captured and saved by our event logger. If we are permitted, the computer screen will be recorded.

4. When all tasks are finished, the participants fill out a post-experiment survey. This survey collects perception and user satisfaction data from the participants. The survey includes the following information: how easy of using each interface for goal-directed tasks, how easy is it for each interface to navigate, how easy is it to understand video content with each interface, and how frustrating and fun is each interface. In addition to the comparison questions between two interfaces, the survey contains several questions asking about perceptions of the user engagement heatmap through freeform comments.

The time of experiment depends on each video's duration and the time the participant spends on performing the task. It takes approximately 40 minutes to complete the entire task. After the experiment is completed, each participant is compensated with a gift card. The analysis of user study results will be in a combination of quantitative and qualitative analysis. Since it is not difficult to record the time that the participants used for the tasks and to compute the correctness rate of their answers, the focus of the analysis of the tasks that include visual search, content re-finding and content understanding will be more quantitative. The collected pre-experiment and post-experiment surveys from the students will help us understand the impact of a heatmap on a video and how the students interpret the presentation of the heatmap which can answer research question about user perceptions about heatmap.

5.2 User study materials

The following materials will be used in the user study:

1. Software: two versions of online video players. One version has the heatmap of user engagement, and the other version is a normal video player with basic controls. The base video player for both of the two versions, extracted from Video Collaboratory, has basic controls plus the speed rate change functionality. For this study, the collaboration features in VC are removed from the video player and will not be tested in our study.

2. Experiment event logger: the event logger used for collecting user interaction data plays a monitor role for tracking the click events happening in the experiment. The events will be saved in the database and provide an opportunity for finding potential correlations between the qualitative results and user interaction data. For example, because the content targets are placed around the highly and non-highly engaged segments, through the logs we can find whether the participants interact with the highly engaged segments more than other segments. During the entire process, the participant's screens were recorded to provide evidence for later verification as long as the participants agreed on the consent form.

3. Video materials: Since it is a within-subjects study design, participants will experience both of the two versions of the video player and watch two different videos. Two HCI lecture videos are selected for the study for the following reasons. First, the user interaction data, used to generate the heatmap, were collected from previous HCI classes. Second, some HCI lecture videos are very straightforward to understand for the participants who even do not have knowledge of HCI. Third, other than traditional lecture videos with talking heads, one HCI class used a few interesting YouTube videos from a public channel[10] for the lecture. Using non-traditional videos might keep the participants from being bored by the talking heads. For the final user study, one lecture video talking about mental model and one video talking about social media are selected. However, for this in-class pilot study, the video material was selected to fit with the class schedule at the time of the study.

4. A pre-experiment and a post-experiment survey. The pre-experiment survey examines the demographic information and experiment related background about the participants and the post-survey can help understand user perceived effectiveness, user perceptions with exposing the heatmap. All the original surveys are appended at the end of this dissertation (Appendix A).

In order to address potential changes in the underlying heatmap video enagement data through participant usage during the studies, the interaction data for the heatmap is fixed prior the study and remains static during the study period.

[10]https://www.youtube.com/channel/UC6-ymYjG0SU0jUWnWh9ZzEQ

Everyone in our study experiences the same heatmap interface by only using a fixed amount of interaction data to generate the heatmap. In other words, the only user interaction data used to generate the heatmap for this study is all coming from before the study starts. Any activities that occur during the user study will still be collected for the results analysis, but will not be counted for the pre-generated heatmap for the sake of the consistency of the user study.

5.3 In-class pilot user study

We conducted our pilot user study with a class in the middle of the semester so that we can find potential design flaws or the interface bugs. The benefit of doing this is that we can obtain reliable data because the study will naturally be part of the class assignment. The class has about 120 students and they are required to watch the videos before the class. The class students were offered a $5 gift card for their additional work if they signed up for our study. Nineteen students signed up and eleven of them did our user study at the end. Despite the low number of participants for this study, we did receive valuable feedback not only for our interface but also for the study design.

Since the study had to be conducted as part of the class assignment, we used the lecture videos which were part of the course material in the week of conducting the study. Table 2 summarizes the information of data source used to generate the heatmaps for the two videos.

Table 2: Heatmap data information in pilot study

	Video 1	**Video 2**
Topic	Running web experiments	Comparing rates
Duration	10 minutes	20 minutes
Watching sessions	74	59
Users	61	46

5.3.1 Participants

The participants who completed the study were eleven undergraduate students from the HCI class. To run the user study, we randomly divided all participants into two groups. Table 3 shows the information of two groups and the order of executing the tasks. Each participant received a $5 Starbucks gift card, plus an extra $5 gift card if the participant was willing to record the screen while he/she is performing the tasks and provide the screen recordings at the end of the study. The study is highly flexible for the students; they can complete the tasks at any time. The only constraint is that they need to complete the study before the class begins.

Table 3: In-class user study participants groups

	Group 1	**Group 2**
Number of participants	5	6
Task 1	Video 1 without heatmap	Video 1 with heatmap
Task 2	Video 2 with heatmap	Video 2 without heatmap

5.3.2 Lessons learned from the study

The prototype is shown in Figure 3 was exactly the same interface we used for this study. Several problems were discovered on the interface itself and the study design, although this pilot user study had relatively small amount of participants. Following is the list of issues that were found from the study.

1. At the beginning of the study, part of the explanation of the procedures of the user study, we did not explicitly inform the participants that there is a heatmap on the video. They were told that they would test two versions of the video player. We intended to do so because we thought we could get a wide variety of user reported interpretations about the heatmap. Unfortunately, doing this confused a lot of participants in this study, resulting in unexpected outcomes from the study. For example, one participant complained that she / he did not even notice the heatmap at all. The survey included the following questions to understand how the participants interpret the heatmap.

   - What do you think the different colors in the heatmap represent?
   - Why do you think the "heat" colors (e.g. the red colors in below picture) in the heatmap occurred?

   More than a half of the participants indicated that either they had no clue about the heatmap or did not notice the heatmap. Example responses include "I didn't even notice it. It wasn't labeled so I just thought it had no purpose. Didn't use it.", "There wasn't", "I don't know what a heatmap is." Only a few participants were aware of the correct meaning of the heatmap correctly. Participants mentioned that "Red indicates more views, green indicates less" and "Frequency". From the feedback we can see that this is a design flaw we made for the study. This lesson taught us that we should inform the participant that the heatmap was generated based on previous users watching activities. We should also tell the participants that darker colors in the heatmap mean higher views happened to the segment while the lighter colors mean fewer views.

2. The heatmap was not interactive. In our initial design of the heatmap, the heatmap is just a static visualized heatmap staying above the player controls. This design gives negative user experience for the participants. Because some participants complained, they did not know how to use the heatmap. After this problem was discovered, we improved our heatmap interface. Two new, more interactive features are added in the new design of the heatmap. (Figure 5). First, the blue handler, which is higher than the heatmap, functions as a video player scrubber. Second, when the mouse hovered on the heatmap, we also provide a hovering handler on the heatmap. It provides an indication of the video position of the mouse. To address the user expectation with the heatmap, we provide those two features to make the new interface more interactive.

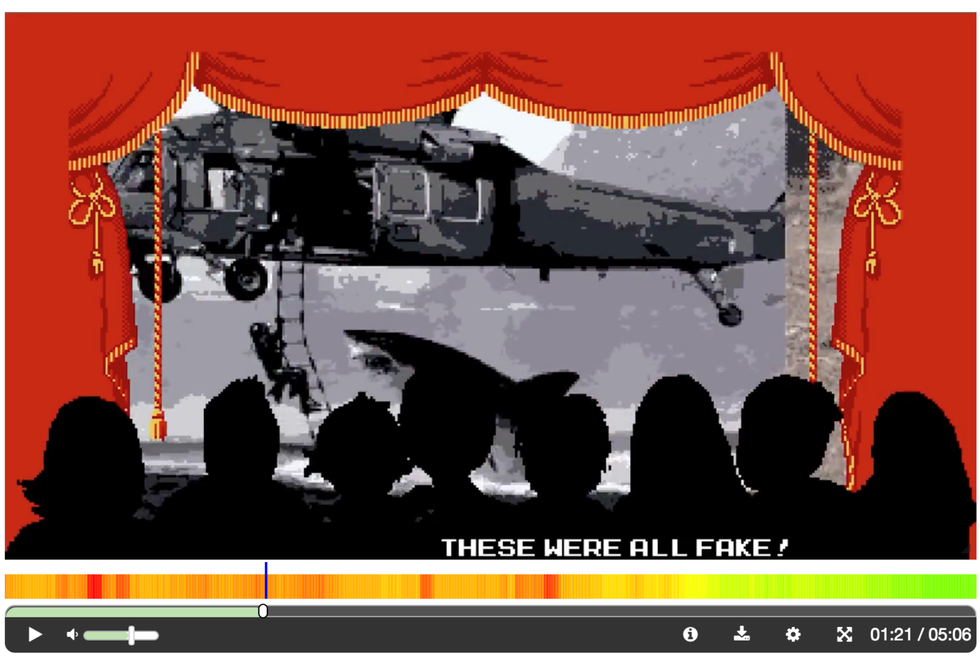


Figure 5: Improved Heatmap

Besides the above two major concerns, other general comments were collected through a couple of open-ended questions to assess the joy points and pain points of the participants while working with the heatmap. On one hand, participants enjoyed the heatmap and thought it is helpful. A lot of them stated that "The heatmap helps find important parts of the video for answering quiz questions", "It helps outline the important information," "It helps find the most relevant information" and so on. On the other hand, participants who do not like the heatmap complained the colors in the heatmap are distracting and the heatmap is useless. "It is not relaxing for the eyes", "No legend on what the colors meant on the heatmap" are the sample responses for this group of participants. The issue of missing legends for the heatmap colors can be addressed by having a clear explanation for the heatmap.

5.4 Conclusion

In this chapter, we begin with the general procedures for the evaluations for the single heatmap interface, which will be referenced in later formal user studies. Then we introduced a pilot we conducted for the preparation of the formal user studies. Although the pilot study is based on a small sample of participants, the findings suggested improvements to the user interface and the study design. Besides that, we did receive the positive feedback to continue our work. One important and interesting result shows that even some of the participants do not understand the heatmap, they still like the heatmap because it could guild them to search relevant information quickly.

CHAPTER 6: SINGLE HEATMAP USER STUDY: IN LAB

In this chapter, we present a formal user study conducted in a laboratory setting. We first describe the detail of the user study for evaluating the single heatmap interface. Second, we analyze the study results based on the research questions that are presented in the previous chapter.

6.1 Study overview

Based on what we learned from the pilot user study, we modified our study design, fixed the interface flaws and adjusted the incentives for the participation. First, the most significant change to the study is that we include an explicit instruction about the interface of heatmap and a full explanation for how it is generated. In the laboratory study, we make sure the participants fully understand the meaning and the usage of the heatmap before they perform the task. Second, what we learned from the feedback about the heatmap is that it is a lack of flexibility, so we redesigned the heatmap. The new interface, as shown in Figure 5, allows the participants to be able to interact the heatmap and control the video through the heatmap as well. Finally, the compensation was adjusted for both the in-lab user study. Unlike the in-class user study, watching the videos are required by the class, the task in the laboratory setting is an entirely new task for the participants. The study lasts approximately 35 minutes per person. Considering the predicted completion time, each participant receives a \$15 Starbucks gift card for the completion of the task. The participants are paid more than the in-class pilot study in which each participant only received a \$5 gift card. This is because the participants in the pilot study had to watch the two videos as part of the week's assignment regardless of the participation of the user study.

6.1.1 Videos and participants

The videos selected in the pilot study had to fit the planned course material at the time of the pilot. In this laboratory study we selected the videos described in section 5.2. One video is a lecture talking about mental model and the other one is about social media. Table 4 shows the information of data source that is used to generate the heatmaps for the two videos. Each participant uses the same heatmap on the same video. The full procedure is described in section 5.1

Table 4: Heatmap data information

	Video 1	**Video 2**
Topic	Mental model	Social media
Duration	15 minutes	5 minutes
Watching sessions	77	151
Users	64	96

Table 5 shows the within-subjects study design for the in-lab user study. With this study design, the participants are equally divided into two groups and each group will use both conditions in this study. As discussed earlier, the order is counterbalanced to mitigate the ordering effect of the presence of the two interfaces.

One of our research questions is to investigate how the heatmap impacts users in understanding the video content. As a result, the participants in our study are

Table 5: In-lab user study design

	Group 1	**Group 2**
Lab participants number	15	15
Task 1	video 1 without heatmap	video 1 with heatmap
Task 2	video 2 with heatmap	video 2 without heatmap

required to take a quiz after watching each video. To ensure each participant has a same background in the video lecture knowledge, we provided new videos for the participants. “New” means the participants have not watched the videos. The participants were recruited from College of Computing and Informatics, so there is a chance for some students who were enrolled in previous HCI classes in the college. To eliminate the possibility of recruiting people who might have prior knowledge of HCI that can possibly help them perform better than other participants who has no HCI background, we made an explicit requirement that the candidates who have taken HCI classes were ineligible for our study.

Regarding the participants, the study consisted of 30 participants (5 females) ranging from 18 to 34 years old. Three (3) out of the 30 participants were undergraduate students, and the rest were graduate students. About 95% of the participants used online videos for learning. The study was taken in the Usability Lab located in Woodward Hall. Based on the responses to the question asking whether the participant has watched the video, 2 of the 28 participants were not included in the result analysis of this experiment since these two indicated that they had watched the videos or some of the videos before.

6.2 In-lab User Study Results and Analysis

In this section, we analyze the in-lab study results and test our research hypotheses based on the research questions in at least four aspects: user's content understanding, user effectiveness, watching patterns and user perceptions. The study in this work is to compare the traditional online video player and a video player with the proposed heatmap. The within-subjects study design allowed each participant to experience both interfaces and then the study data can be divided into two conditions (interfaces) for the comparison. More specifically, one set of data is collected from the usage of the video player with the heatmap while the other one is collected from the interface with the core video player. The Mann-Whitney U (MWU) test is employed to test significance of the data collected from a within-subjects study [27].

6.2.1 RQ1: Content understanding

The first challenge is to examine whether the heatmap helps the users understand better on the video content (RQ1). The analysis on content understanding contains two parts. First, the participant's quiz performance is one potential method to measure understanding. Second, from the reported data on how well the participants understood the videos and how confident when they completed the quiz, another interesting analysis is to figure out whether the heatmap causes significant changes in the participants' confidence when they are performing tasks related to video materials (e.g. answering the questions on the quiz).

In total, the quiz of the two lecture videos includes seven questions (one video has four questions, and one has three questions). All of the questions are related to the

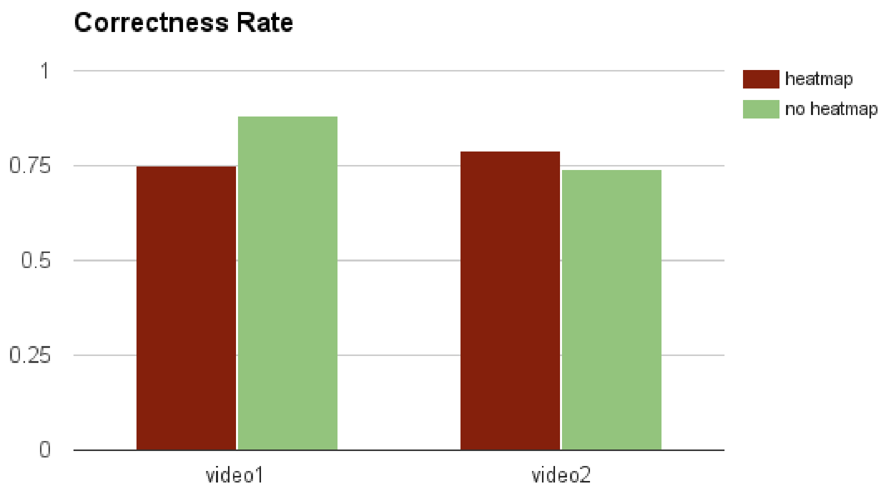


Figure 6: Correctness rate

video content. Each video has been watched under both conditions which are the player with the heatmap and the video player without the heatmap. Quiz question types are true/false and multiple choice. In the condition of the heatmap, the average correctness rate is 81%, while in the group of no heatmap, the average correctness rate is 77%. Figure 6 shows a detailed comparison for each video. The first video (obtained from YouTube and talking about social media) has 75% correctness rate which is lower than the second video (talking about mental model) with an 88% rate. The second video has more close correctness rates for the two interfaces. For this video, the heatmap version has 79% correctness versus non-heatmap version that has 74% correctness rate. According to the MWU test, none of the conditions shows a significant difference between the player with the heatmap and the player without the heatmap (Video 1: z-score = -0.74, $p > 0.05$ and Video 2: z-score = 1.67, $p > 0.05$).

Overall, we observed that the participants performances are very close in the both conditions (non-heatmap: $\mu = 5.72, \sigma = 1.2$, heatmap: $\mu = 5.36, \sigma = 1.0$) [11]. Contrary to expectations, this study did not find a significant difference for the performance in two conditions has no significant difference (MWU test: z-score = 1.16 and $p > 0.05$). Therefore, with the p-value, we can not reject the null hypothesis $H_{1.1,null}$: *the participants* ***does not*** *have a higher correctness rate of their answers to the quiz questions with the heatmap compared to a basic video player without the heatmap.*

To examine whether the heatmap impacts the participants in understanding the video content and whether the heatmap increases the participant's confidence in their answers to the question. Two Likert-type questions, scaled from 1 to 7, are used in the survey:

- How much do you feel you understood this video?
- How confident do you feel about your answers to the quiz questions?

Although the quiz performance between the two groups does not demonstrate a significant difference, participants showed higher confidence level and understanding with the heatmap from their reports. Figure 7 shows the mean value of the two sets of data.

First, regarding understanding the video content, the participants felt they understood ($\mu = 5.43, \sigma = 0.79$) with the heatmap more than learning from the purely basic video player ($\mu = 4.86, \sigma = 1.15$). However, the MWU test result does not show a significant different between two groups of the participants (z-score = -1.92 and p = 0.053). Since the p-value does not meet the threshold for statistical significance,

[11] μ and σ is the mean score and standard deviation of the data respectively.

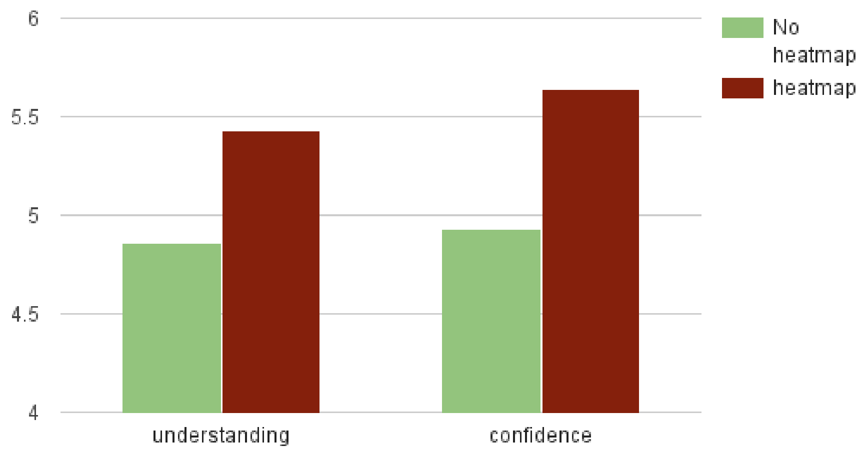


Figure 7: Content understanding and confidence level (mean, scale: 1-7)

we can not reject the null hypothesis $H_{1.2,null}$: *Participants will **not** understand the video better with the heatmap versus the primary video player.*

Second, when the participants were answering the quiz questions, they felt more confident with the heatmap ($\mu = 5.64, \sigma = 1.03$) than the basic video player ($\mu = 4.93, \sigma = 1.41$). MWU test result ($z$-score = -1.99, $p < 0.05$) statistically shows there is a significant difference between the two groups using the heatmap and without using the heatmap. Therefore, we can reject the null hypothesis of $H_{1.3}$ in favor of the alternative: *Provided the heatmap with the video player; the participants will be more confident to answer questions.*

The analysis of the results reveals that participants do not have a better quiz performance despite that they felt more confident to answer questions when they used the video player that has the heatmap. The most likely cause might be due to the fact that the quiz questions are easy to answer once they find the related segments from the videos. Both conditions have a high correctness rate, and the rates are close enough (heatmap: 77%, no heatmap: 81%). From the perspective of the feeling of understanding, participants felt they understood better when they used the video player with the heatmap, although the difference between two versions of the video player is not significant. Nonetheless, the video player with the heatmap does not show the heatmap adversely affect users in understanding the video content.

6.2.2 RQ2: Effectiveness / Perceived usability

The effectiveness is measured through the feedback from the respondents. For both conditions of the study, the participants rate how easily they could find the needed information from the video and report how useful the video player is for learning from the lecture. The responses are collected from the following 7-point Likert questions.

- How useful was the version with/without the heatmap for learning from the lectures?
- When using the version with/without the heatmap, how easily were you able to find the information needed to answer the quiz questions?

Generally speaking, the analysis reveals that participants felt the heatmap is more effective when they were completing learning tasks based on video content. Figure 8 shows the results obtained from the preliminary analysis based on users' ratings.

On the one hand, on the question of usefulness, the study found that the player with the heatmap ($\mu = 5.93, \sigma = 1.15$) is more useful for learning from the lectures than the basic version without the heatmap ($\mu = 3.93, \sigma = 1.12$). The z-score (= -5.08) and p-value (< 0.05) from MWU test indicate there is a statistically significant

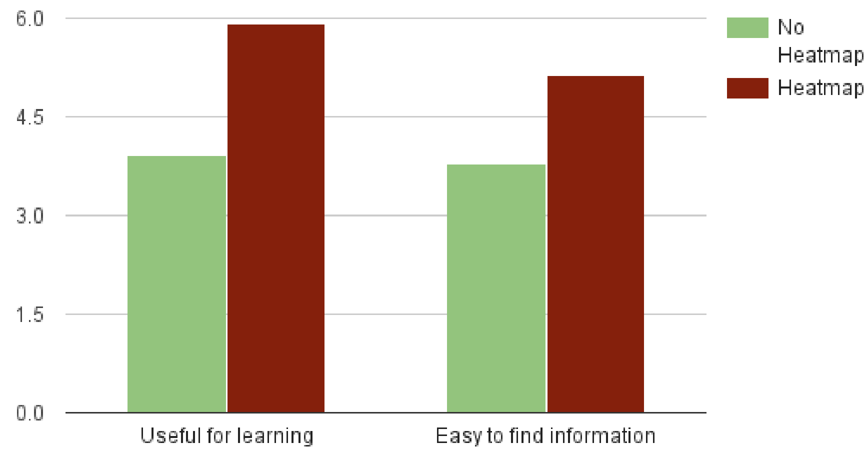


Figure 8: Useful for learning and easily find information (1-7)

difference between the two versions of the video player. Since the p-value (< 0.01) is much less than the threshold (< 0.05), for $H_{2.1}$ we can reject the null hypothesis in favor of the alternative: *The heatmap is more helpful for the participants to learn from lecture videos.*

On the other hand, by using the version of a video player with the heatmap, participants were able to more quickly locate the information that is needed for answering the quiz question. (heatmap: $\mu = 5.1, \sigma = 1.58$, non-heatmap: $\mu = 3.8, \sigma = 1.77$). MWU test analysis shows the difference between the versions is statistically significant (z= -2.84 and $p < 0.05$). With this p-value, the null hypothesis $H_{2.2,null}$ is rejected in favor of the alternative: *With the heatmap, users are able to to find more easily needed information from the video.*

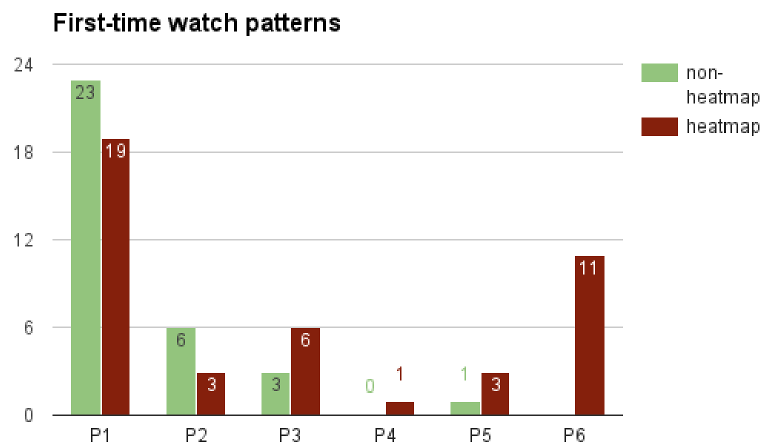
Another important finding is that 79% of those who participated in the study indicated that they prefer to use heatmap for learning lectures from videos. The rest of the participants liked to use the basic video player for learning. MWU test shows that the preference for learning lectures is statistically significant ($p < 0.01, z = 5.1962$). Additionally, we asked their preferences when they need to complete the quiz. The results show that 86% of participants would prefer to have the heatmap for completing the quiz. Similar to the preference for learning, the preference for this question is statistically significant as well ($p < 0.01, z = 5.1962$).

Taken together, the analysis of the results suggest that the video player with the heatmap is more perceivably effective than a basic one without the heatmap for learning tasks based on online lecture videos. More specifically, the heatmap is more helpful for learning and quickly locating information from the video. Furthermore, the study indicates that users are more likely to have the heatmap for learning and complete tasks based on the video material.

6.2.3 RQ3: Watching patterns

When the heatmap is presented to the users, it is interesting to understand whether users would change their interaction patterns with the heatmap. The watching patterns are analyzed through the feedback collected from the questionnaires and verified by available screen recordings. The user watching patterns are explored in two phases: first-time watch and review.

Figure 9 shows the comparison between the heatmap and non-heatmap conditions in five common watching patterns that the participants chose to watch the video for the first time. The patterns from *P1* to *P5* are listed as multiple choice answers for a question asking how to watch the video for the first time. Multiple answers can be accepted for this question. If the none of the listed patterns matched the way the participant, she/he can choose other and leave a comment. Note that the pattern *P6* is not available in the non-heatmap condition simply because there is no heatmap in the condition. The number on top of each bar in Figure 9 represents the number of participants who chose the pattern in the condition. For example, without the heatmap, there are 23 participants watched the entire video in one go without stops (*P1*).



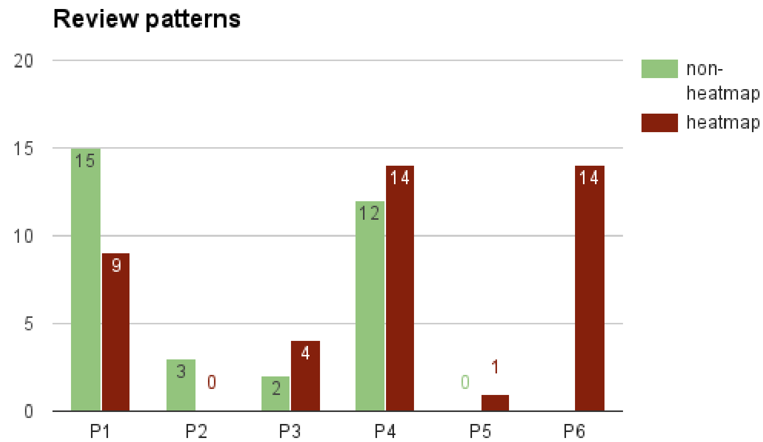
P1: Watched the full video in one go without stops.
P2: Paused, watched, and repeated the process.
P3: Watched, skipped then watched, and repeat this process.
P4: Randomly selected a start to watch and went back if needed, and repeated this process.
P5: Other.
P6: Used the heatmap to help with navigation (Only available in the heatmap condition).

Figure 9: Watching patterns in first-time watch

Figure 9 shows that *P6* has been chosen 11 times by the participants when they had the heatmap option. Due to the 11 selections of using the heatmap for help with navigation, fewer participants are observed to watch the entire video without stops.

Although the pattern *P6* is unavailable in the non-heatmap condition, it is included to figure out whether the participants' typical watching patterns have been changed due to the presence of the heatmap or to random effects. Further statistical analysis (MWU test) reveals that the change of the number of participants due to the presence of the heatmap is not significant (p-value could meet the significant level at 0.05) for each watching pattern. With this finding, the null hypothesis $H_{3.1,null}$, *when the heatmap is available; more users will use heatmap than other navigation methods to watch the video for the first time*, can not be rejected. The reason why the difference is not significant for the video's first-time watchers we think is that having the heatmap available does not mean watching the "hot" segments is sufficient to understand the content. Still, the participants need to watch the video so that they can understand what the video talks about.

In addition to the first phase of watching the video, the interaction methods happened in the phase of reviewing the video were collected for the study analysis. Figure 16 illustrates the number of participants who chose the patterns to examine the video for answering the quiz. Similar to Figure 9, the number on top of each bar means the number of participants who chose the pattern with the particular interface. For instance, the first green bar indicates 14 participants are observed to respond they did not watch the video again for completing the quiz in the non-heatmap condition. One interesting finding observed from Figure 16 is that a lot of participants just tried to find the needed information by their memories on the video in both conditions. Again, there is an additional interaction pattern (*P6*) is available for the participants when they used the heatmap interface whereas there was no such an option when the participants did not have the heatmap. From the observation, 14 participants (50% of total) reported that they used the heatmap to help with navigation. The observed number is higher than the number of participants (11) who used heatmap for the first-time watch.



P1: Did not watch the video.
P2: Watched the full video again.
P3: Tried to remember the place in the video that was related to the question, and navigated to it.
P4: Randomly selected a place to watch, and then randomly selected another one if it wasn't relevant.
P5: Other.
P6: Used the heatmap to help with navigation (Only available in the heatmap condition).

Figure 10: Reviewing patterns

From Figure 9, we see that the number of people who fully watched the video in the heatmap interface is less than the number of individuals who watched the entire video without the heatmap. To examine whether this change is significantly caused by the presence of the heatmap, like the analysis of first-time viewing patterns, every commonly used pattern for this comparison has been analyzed independently. However, the statistical analysis (MWU test) reveals that none of the changes of the number of participants due to the presence of the heatmap is significant for each individual reviewing pattern. With this finding, the null hypothesis $H_{3.2,null}$, *in the self-reported data, a significant proportion of participants will use heatmap for completing tasks related to video content (e.g. quiz)*, can not be rejected. Although no significant changes are found for conventional watching patterns between the two different conditions, the heatmap does decrease the number of people who just "click play, sit back, and watch". Indeed, half of the participants interacted with the heatmap for completing the quiz. This implies that people are willing to try to leverage the heatmap for looking for information from the video.

Overall, at least 19 participants (out of 28) reported they used the heatmap during the study either for the first-time watch, completing the quiz or both. It is interesting to note that 84% of them reported they interacted with the heatmap less than ten times and only about 7% of the participants interacted a lot (more than ten times). The percentages show that people are interested in trying the new feature even if it is not a primary means of interaction. A possible explanation for this might be that understanding the video is the primary task for the users and the new feature such as the heatmap is an auxiliary tool for learning. Or perhaps the participants just do not need to interact with the heatmap with that many times so that they can finish the quiz.

The analysis of screen recordings collected from the participants confirms previous findings from their survey feedback. When the participants watched the video for the first time, most of them did not interact with the heatmap too often. However, a common pattern was discovered when the participants reviewed the video for completing the quiz. Most participants, who indicated they interacted with the heatmap for the quiz, initially clicked the "hot" segments to see if they could find the answers. If they failed in searching answers from the "hot" segments, they would try some other places from the video. Since some "hot" segments are related to the quiz questions directly, a participant commented that *"I can navigate to important points without memorizing their location in the video."* A similar response is that *"It was easy to find stuff based on heat maps as most answers having red color"*. However, not all the quiz answers are linked to the peaks on the heatmap, thereby some participants just randomly clicked some places on the heatmap or clicked the video based on their memories like what they commented.

In summary, both participants' self-reported data and their screen recordings have identified that the heatmap does have impacts on user's watching and navigation patterns, even though comparison of self-report viewing pattern selections does not show a significant difference between the video player with the heatmap and the one without the heatmap. Whereas people do not interact the heatmap frequently when they first watch the video, people are more likely to use the heatmap as a navigation tool for actively searching information when they review the video. More interactions with the heatmap in the review stage might lead the participants to believe they can complete the video tasks more effectively, which was also found from the previous analysis.

6.2.4 RQ4: User perceptions of the heatmap

Based on the responses from many 7-point Likert scale questions asking about user perceptions and open comments from the participants, the feedback about the video player with the heatmap was very positive. Additionally, participants preferred to have the heatmap available for course work. Figure 11 illustrates the comparison between the heatmap version and the basic player from the perspectives of user experience.

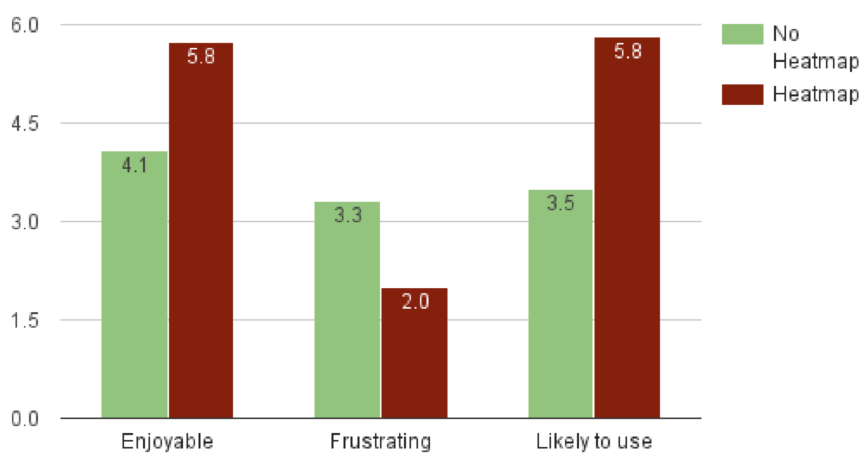


Figure 11: User experience with heatmap vs. without heatmap

The figure shows that the player with the heatmap ($\mu = 5.8, \sigma = 1.0$) is more enjoyable for learning from the lectures than the basic version without the heatmap ($\mu = 4.1, \sigma = 1.4$). MWU testing result shows that the difference is statistically significant ($z = -4.5, p < 0.01$). In the meantime, users felt less frustrated when the heatmap was available (non-heatmap: $\mu = 3.3, \sigma = 1.9$ and heatmap: $\mu =$

$2.0, \sigma = 1.3$). MWU testing result shows the difference is statistically significant as well ($z = 2.8, p < 0.01$). The matching results found from enjoyable and frustrating also verifies the participants did the survey correctly and consistently. Therefore, we can reject the null hypothesis $H_{4.1,null}$ in favor of the alternative: *The heatmap is more enjoyable (less frustrating) for the participants to learn from lecture videos.* Another finding, shown in Figure 11 and supporting to reject $H_{4.1,null}$, is that the participants prefer to have the heatmap for class work (non-heatmap: $\mu = 3.8, \sigma = 1.8$ and heatmap: $\mu = 5.2, \sigma = 1.6$). The statistical analysis (MWU testing) demonstrates that users differ in selecting the video players for class work ($z = -4.7, p < 0.01$).

Later on the survey, participants were asked to provide their feedback on most and least favorites about the heatmap. Many participants believed that the heatmap is very useful for the learning environment, which is also the part they liked most. "Important" (appeared 14 times) and "help/helpful" (appeared 11 times) are most frequent words the participants mentioned. The common praise from people who said such words is that the heatmap helps them find important information effectively. For example, participant *p1* mentioned "The importance part of the video was highlighted, which helped me to concentrate more", and participant *p2* commented that "Very helpful while studying for exams, or when you want to go through the important parts of a video quickly". One interesting comment from participant *p3* is that "It feels like I am interacting with other users." On the other hand, participants criticized the heatmap about the irrelevant highlighted parts and attention distraction because the heatmap is not a visualization about the importance of the video content. Instead, the heatmap is a visualization about the frequency of the views of the segments, and some hot spots do not necessarily represent the segments of the content are important. Due to the irrelevant hot zones, participants thought the heatmap sometimes is not accurate. For example, one participant *p4* said that "It seems like some people had replayed sections which weren't relevant" and *p5* complained "Sometimes, it is misleading." Distracting people from watching the video indeed is not the intention of the design of the heatmap. Unfortunately, it happened to several participants. A participant *p6* mentioned "while this is a very helpful idea, it might cause the user to not watch the full length of the video. This might also cause users to skip/overlook portions of the video which might be useful or informative/interesting for that particular individual". What *p6* expressed is true, and this is the shortcoming of the heatmap generated by the raw user interaction data. In this research, the heatmap is the visualization of the plays accumulated on the specific segments within the video. If more people have watched / engaged with a segment, it should warrant greater attention, but this does not necessarily indicate whether the content itself is important or not. As a result, if a person is only attracted by the hot zones, then the heatmap might have negative impacts such as "loosing learning points by following the only heatmap" said by one participant.

Overall, the study of user perceptions strengthens the idea that the heatmap is helpful for learning and enjoyable to use despite the study leads us a challenge which is how to design a heatmap to keep minimum impact in distracting the users. In fact, the participants left some good suggestions such as providing a function which is able to turn on and off the heatmap. Another participant recommended enabling annotations on the heatmap like SoundCloud [12] so that users can read the comments and then mark the heatmap based on their understandings. As stated in the previous section of study design, the video platform Video Collaboratory where user interaction data came from originally has a video annotation feature. Part of the setup of the experiment is to purely examine the heatmap; it is interesting to see the removed feature for annotation was pointed out by several participants. This particular comment made us start thinking that maybe in the future we can evaluate the heatmap together with the annotation feature in VC.

6.3 In-lab study summary

In this section, we analyzed the results collected from the user study conducted in a laboratory setting. To sum up, the in-lab study has found that the heatmap is helpful and effective for learning from lecture videos while it did not improve user performance on video related tasks such as completing a quiz. First, the evidence from the analysis suggests that the heatmap does not significantly improve user's understanding of the video content based on their quiz performance. However, the participants did finish the quiz with more confidence by using the heatmap. Second, users felt they could learn the lecture more effectively with the heatmap. The statistical results show the heatmap is significantly helpful for learning and finding information from the video. Third, many participants were observed to use the heatmap either when they were watching the video for the first time, or they were reviewing the video. However, no significant differences were found from their basic watching patterns. Finally,

[12] https://soundcloud.com/

in the aspect of user perception, users were satisfied by the heatmap. A significant proportion of participants agreed that the heatmap is useful and enjoyable for learning while a few participants argued that the heatmap could distract them from watching the video. Overall, the heatmap is beneficial and promising for the users, particularly for the learners who use the video as the learning tool.

CHAPTER 7: SINGLE HEATMAP USER STUDY: MECHANICAL TURK

Having discussed the analysis of the in-lab study in the previous chapter, this chapter moves on to a larger-scale study using Amazon Mechanical Turk. The in-lab study recruited 30 participants, and all of them are students from our college, though the video materials used in the study do not require a participant to have a high education background in computing. Compared to the previous in-lab study, the Mechanical Turk study has recruited a higher volume of participants with a variety of different backgrounds. Larger scale Mechanical Turk studies can help reduce the risk of the contaminated subject pool, dishonest responses and experimenter effects [35]. Moreover, we are interested to see how results from the in-lab study scale to a larger group of participants.

In this chapter, we present a second formal user study conducted for evaluation of the single heatmap. This user study was conducted with the Mechanical Turk with a larger pool of participants compared to the previous in-lab user study. We present details of the new study and then analyze the study results based on the research questions that are presented in the previous chapter.

7.1 Overview of Mechanical Turk user study

The user study is conducted via online recruitment and subject interaction, which is called a HIT (Human Intelligence Task) on Amazon Mechanical Turk. We deployed two separate HITs online to account for different orderings of the heatmap appearance. This setting is exactly same as the within-subjects design for previous in-lab study. Our HITs required the Turkers to meet certain qualifications to be eligible for the study. Turkers who have more approvals and higher approval rates are expected to deliver a good-quality of response. Initially, the qualification of Turkers for our study was set to be a Mechanical Turk master [13] have at least 98% approval rate and 200+ approvals, however, after a couple of days, the number of participants was lower than our expectations and some of the Turkers who finished our task complained they were paid less based on their qualifications. Therefore, the qualifying criteria were changed to a normal Mechanical Turker who is not necessary a master but at has above 95% approval rate and received at least 100 approvals. Each HIT opened up 50 slots for signing up. A criterion was set in the second HIT to make sure the participants who took the first HIT were unable to take the second one. Eventually, two HITs recruited 100 participants in total, and roughly 80% of them have reached master level.

The Mechanical Turk study is structured the same as the previous in-lab user study, but with some changes to account for the online interaction with participants. Since participants recruited from Mechanical Turk are all online and we can not give the introduction verbally, the description of the heatmap is explained in the study's task instructions. Additionally, it is not straightforward to know whether a particular Turker has previously watched one of the videos or not. To eliminate this possibility we asked a specific question on the survey to know whether they have watched the

[13]the master level is evaluated and assigned by Amazon Mechanical Turk.

video or not. If the participant indicates she/he has watched the video, her/his study data is discarded. In contrast to the in-lab user study, each Turker is paid $2 for its completion for this work. $2 seems much lower than $15, however it is still considered as a good rate in the community since the location is not limited to some places where have a high average paid rate such as the US. The literature [5] showed the evidence that Mechanical Turk can be used to obtain high-quality data inexpensively and rapidly.

7.1.1 Videos and participants

Six participants did not fully complete the task, and they were eliminated for later analysis. Additionally, the data from 10 of the remaining 94 participants were removed in the results of the experiment. Two of these participants failed the quality control questions. For example, we evaluate how enjoyable and frustrating the interface is. It is a contradiction if one individual gives the same rating for both two questions. Another two of the participants had watched the videos previously. The other six participants were removed due to inconsistent or corrupted logs. The remaining 84 participants consisted of 49 males and 35 females. The age range is much wider than the previous in-lab study. Most of the participants are between 18 and 45 years old while 11 of them indicated they are older than 55. Table 6 and Table 7 show the distribution of ethnicity and received education degree level respectively. Since location is not restricted in our HITs, the ethnicity is very diverse. More than half of the participants have the bachelor degree and above. 63% of the participants reported that they watched videos for learning.

Table 6: Mechanical Turk participants ethnicity

	Hispanic	*Asian*	*Black*	*White*
# of participants	49	16	10	9

Table 7: Mechanic Turk participants degree

	High school	*Associate*	*Bachelor*	*Master*	*Other*
# of participants	5	19	47	8	5

From the demographic data, the Mechanical Turk studies did receive more participants with a more diverse background as we expected. The following sections describe the results analysis for each research question together with hypothesis testings again.

Table 8 shows the overview of the study design. The videos used in Mechanical Turk study are same as previous in-lab study (see Table 4). Again, because of the within-subjects design, the participants are equally divided into two groups and each group will use both conditions. The procedure is exactly same as our earlier in-lab study, except more participants were involved and there is no location restriction for doing the study.

Table 8: MTurk user study design

	Group 1	**Group 2**
Participants number	50	50
Task 1	video 1 without heatmap	video 1 with heatmap
Task 2	video 2 with heatmap	video 2 without heatmap

7.2 MTurk Study Results and Analysis

Results were analyzed in the context of the same research questions for the in-lab study. Each following section is our findings to every single research question. Doing so is also helpful for us to compare the analysis we found from the previous user study.

7.2.1 RQ1: Content understanding

For the first research question related to user's content understanding, the Turkers produced very similar results as the in-lab participants did. When they had the heatmap in the video player, their quiz performance is again worse than if they just used the basic video player. Figure 12 summarizes the average correctness of each quiz after watching each video. In contrast to the second video (4 questions, mental model) has almost same correctness rate (86%), the quiz performance from the first video (3 questions, social media) is different. For the first video, the participants with the heatmap condition did worse than the basic video player. The MWU testing shows that none of interfaces significantly differs in the quiz performance (Video 1: $z = 0.79$, $p = 0.46$ and Video 2: $z = 0.52$, $p = 0.63$).

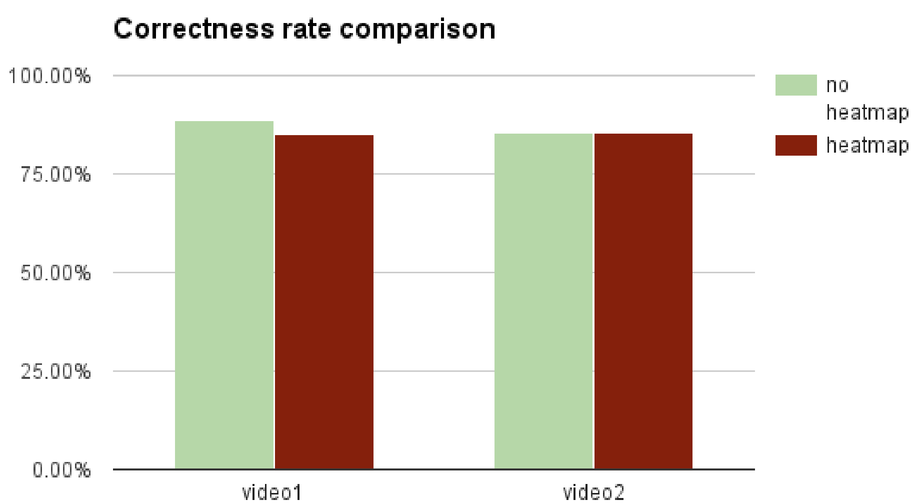


Figure 12: Correctness rate comparison

Overall, participants with the heatmap performed slightly worse in quiz than when they did not have the heatmap (non-heatmap: $\mu = 6.12$, $\sigma = 1.06$, heatmap: $\mu = 5.95$, $\sigma = 0.95$). However, based on the statistical test the quiz performance of the participants do not significantly differ in the basic online video player and the player with the heatmap (MWU test, $p = 0.31$, $z = 1.02$). Since the p-value does not meet the threshold (0.05), we can not reject the null hypothesis $H_{1.1,null}$: *the participants **does not** have a higher correctness rate of their answers to the quiz questions with the heatmap compared to a basic video player without the heatmap.* The finding matches what has been found from the previous in-lab study. A possible explanation for no significant difference might be the quiz questions were too easy and no time limit was set for the session.

In the analysis of the in-lab study results, we found that participants felt more confident in outcomes when the heatmap is available. The participants felt significantly more confident with the heatmap than the primary video player whereas there is no significant difference found in understanding. For the Mechanical Turk study, Figure 13 illustrates the difference between the two interfaces in terms of user's understanding (non-heatmap: $\mu = 5.25$, $\sigma = 1.23$, heatmap: $\mu = 5.58$, $\sigma = 1.18$) and confidence (non-heatmap: $\mu = 5.04$, $\sigma = 1.41$, heatmap: $\mu = 5.49$, $\sigma = 1.43$). The trending is basically similar to the results of the in-lab study. The statistical tests implied that the heatmap significantly helped them understand better (MWU test, $p < 0.05$, $z = -2.06$) and improved their confidence for answering the quiz (MWU test, $p < 0.05$, $z = -2.35$). With the two p values, the null hypothesis $H_{1.2}$ and $H_{1.3}$ can be rejected in favor of the alternatives: *participants will understand the video better with the heatmap versus the basic video player* and *with the heatmap, the* *participants will be more confident to answer questions* respectively.

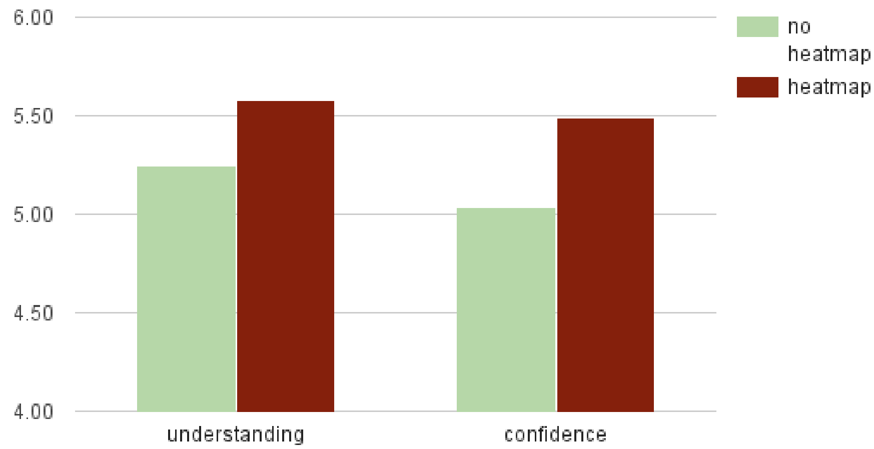


Figure 13: Content understanding and confidence level (mean, scale: 1-7)

The findings from Mechanical Turk study are similar to what have been found from the previous user study, except in the aspect of understanding. What was found in common is that participants did not perform better in the quiz with the heatmap. However, the heatmap did significantly help the participants feel more confident to answer questions and understand better about the video content. For future exploration on the research question about user's content understanding, we suggest to design the quiz with particular difficulty and set a time limit.

7.2.2 RQ2: Effectiveness / Perceived usability

In our previous study on the measurement of perceived usability, it has been shown that the heatmap was rated significantly useful for learning and easy for finding needed information. Ratings were collected from two 7-point Likert questions. The

Mechanical Turk study provided additional evidence to support the conclusion. The analysis of MTurk results informs that participants felt the heatmap is more effective when they were completing learning tasks based on video content. In Figure 14, there is a clear trend of better ratings in the heatmap condition. First, the video player with the heatmap ($\mu = 5.58, \sigma = 1.49$) is more useful for learning from the lectures than the basic version without the heatmap ($\mu = 4.61, \sigma = 1.58$). More importantly, MWU testing indicates there is a statistically significant difference between the two versions of the video player ($p < 0.01$, $z = -4.24$). Since the p-value is much less than the threshold (< 0.05), for $H_{2.1}$ we can reject the null hypothesis in favor of the alternative: *The heatmap is more helpful for the participants to learn from lecture videos*. Second, as shown right in Figure 14, by using the version of the video player with the heatmap, participants were able to more easily locate the information that is needed for answering the quiz question (non-heatmap: $\mu = 4.0, \sigma = 1.82$, heatmap: $\mu = 5.08, \sigma = 1.37,$). MWU test analysis shows that the difference between the versions is statistically significant ($z = -4.07$ and $p < 0.01$). With this p-value, the null hypothesis $H_{2.2,null}$ is rejected in favor of the alternative: *With the heatmap, users are able to more easily to find needed information from the video.*

Another important result found from Mechanical Turk study is that 74% of the participants preferred to use heatmap for learning lectures from video and 79% would prefer to have the heatmap for completing the quiz. For both cases, further MWU testing tells that the preferences are statistically significant (prefer for learning: $p < 0.01$, $z = 3.45$ prefer to use for quiz: $p < 0.01$, $z = 2.64$). The findings from these two questions in the Mechanical Turk study are consistent with the statistical analysis of

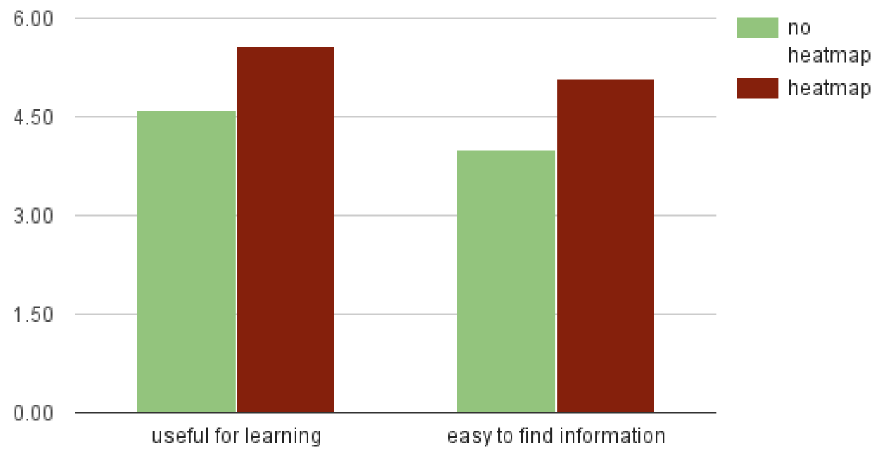


Figure 14: Useful for learning and easily find information (1-7)

the in-lab study.

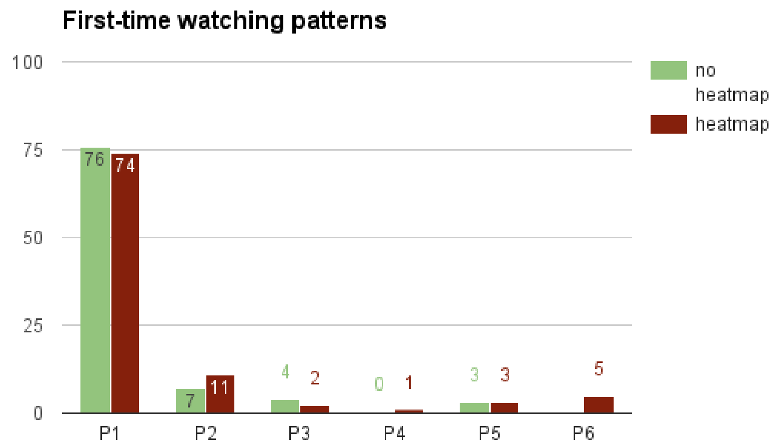
Together these results provide valuable insights into the effectiveness of the heatmap. All of the findings demonstrate that the heatmap is more perceivably effective than a basic one without the heatmap for learning tasks based on online lecture videos. For example, the heatmap supports better learning and helps to locate information from the video. Furthermore, the study indicates that users are more willing to have the heatmap for learning and complete tasks based on the video materials.

7.2.3 RQ3: Watching patterns

Previous studies evaluating user navigation patterns observed the results that many participants were choosing to use the heatmap for their first-time watch and review. However, there was no significant proportion of people who changed their basic watching patterns such as watching the entire video without stops and using the memory to locate the information from the video. It would be interesting to examine whether the heatmap can cause a significant change to the user watching patterns on a large scale of participants from Mechanical Turk community.

Figure 15 presents the comparison of a number of participants who chosen the pattern between the heatmap and non-heatmap conditions to watch the video for the first time. Five commonly used patterns were listed on the survey and multiple patterns from *P1* to *P5* can be selected. It is possible to see whether the heatmap caused changes for the number of participants who use the pattern. From the comparison, no obvious changes between two conditions were found in any individual group. Only 5 participants indicated that they used the heatmap for their first-time watch. The statistical analysis (MWU test) reveals that for each watching pattern the change of the number of participants due to the presence of the heatmap is not significant (p-value could meet the significant level at 0.05). Therefore, in this study, the null hypothesis $H_{3.1,null}$, *when the heatmap is available, more users will use heatmap than other navigation methods to watch the video for the first time*, can not be rejected. What is surprising is that less proportion of people were willing to use the heatmap for their video watching compared to the findings from the in-lab user study.

The reviewing pattern happened in the phase of finding the information to complete the quiz are examined in the same process. Figure 16 illustrates the number of participants who chose the patterns to review the video for answering the quiz. The number on top of each bar represents the number of participants who have opted for the pattern with a particular interface. From the observation, 17 participants out of 84 reported that they used the heatmap to help with navigation, which is

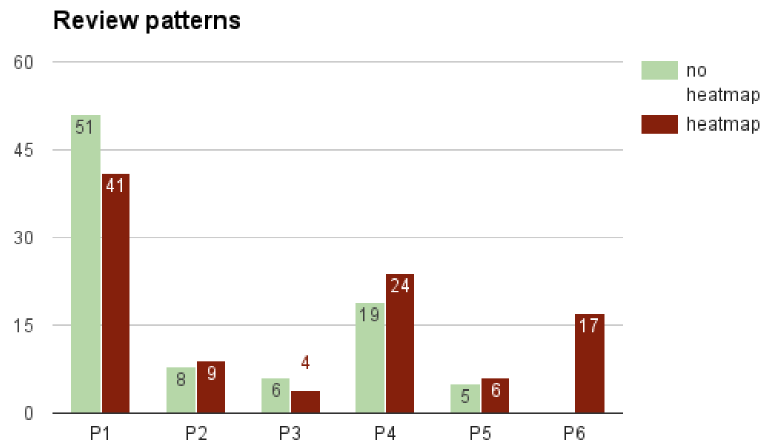


P1: Watched the full video in one go without stops.
P2: Paused, watched, and repeated the process.
P3: Watched, skipped then watched, and repeat this process.
P4: Randomly selected a start to watch and went back if needed, and repeated this process.
P5: Other.
P6: Used the heatmap to help with navigation (Only available in the heatmap condition).

Figure 15: Watching patterns in first-time watch

much higher than the number of participants (5) who used heatmap for the first-time watch. Additionally, the number of people who did not watch the video in the heatmap interface declined because more people tended to use the heatmap. The statistical analysis (MWU test) reveals that none of the changes of the number of participants due to the existing of the heatmap is significant (p-value could meet the significant level at 0.05) for each individual reviewing pattern. As a result, the null hypothesis $H_{3.2,null}$, *in the self-reported data, a significant proportion of participants will use heatmap for completing tasks related to video content (e.g. quiz)*, can not be rejected.

Our survey feedback also reveals that half of the participants (42 out of 84) used



P1: Did not watch the video.
P2: Watched the full video again.
P3: Tried to remember the place in the video that was related to the question, and navigated to it.
P4: Randomly selected a place to watch, and then randomly selected another one if it wasn't relevant.
P5: Other.
P6: Used the heatmap to help with navigation (Only available in the heatmap condition).

Figure 16: Reviewing patterns

the heatmap in general. 90% of them indicated that they used the heatmap less than ten times during the session. Different from the in-lab study that the participants provided screen recordings for us, no screen recordings were collected for the analysis for the Mechanical Turk study. However, it is still possible to discover user watching patterns from the logged user interaction data. We visualized the interaction records and compared to the original heatmap used in the study. What we found is that some hot segments shown on the original heatmap did attract a high volume of views and clicks. In the meantime, some unexpected hot segments were generated as well despite that the new hot spots were not caused by the heatmap. For example, the end of the video gained a lot of clicks because the quiz question would not be visible until the video reached the end.

In summary, the Mechanical Turk study found same results as previous in-lab user study. Compared to the changes of first-time watching patterns, there is an important number of participants who changed their reviewing patterns and used the heatmap when the heatmap was available. People focused on understanding the video content first without frequently interacting with the heatmap. However, for the purpose of searching information, more people were aware of the existence of heatmap and tried to use it as a tool. The statistical analysis does not support that the heatmap causes significant changes to people's watching patterns either for the first-time watch or later reviews.

7.2.4 RQ4: User perceptions of the heatmap

In general, the overall response to user receptions about the heatmap is very positive. The study results provide strong evidence to support that the video player with the heatmap is more enjoyable, less frustrating and more preferable. Figure 17 illustrates the comparison between the heatmap version and the basic player regarding the three aspects. Recall that the ratings were collected from several 7-point Likert questions.

The video player with the heatmap ($\mu = 5.58$, $\sigma = 1.49$) is more enjoyable for learning from the lectures than the basic version without the heatmap ($\mu = 4.61$, $\sigma = 1.58$). MWU testing result shows the difference is significant at the p = 0.05 level ($z = -3.10$, $p < 0.01$). In the meantime, a positive correlation was found between enjoyment and frustration. Users felt less frustrated when the heatmap was

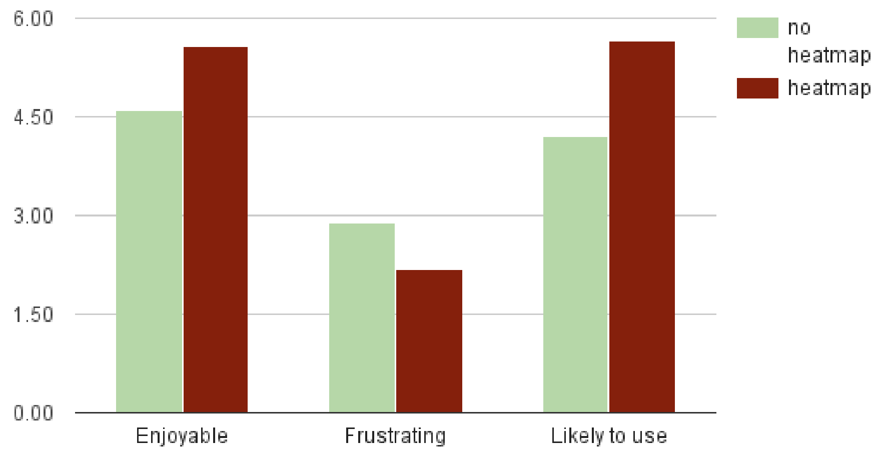


Figure 17: User experience with heatmap vs. without heatmap

available (non-heatmap: $\mu = 2.90$, $\sigma = 1.87$ and heatmap: $\mu = 2.19$, $\sigma = 1.53$). It is not surprising that the difference is statistical significant as well (MWU: $z = 2.52$, $p = 0.01$). The matching results found from enjoyable and frustrating also verifies the participants did the survey correctly and consistently. Therefore, we can reject the null hypothesis $H_{4.1,null}$ in favor of the alternative: *The heatmap is more enjoyable (less frustrating) for the participants to learn from lecture videos.* Another finding, shown in Figure 17 and supporting to reject $H_{4.1,null}$, is that the participants prefer to have the heatmap for class work (non-heatmap: $\mu = 4.02$, $\sigma = 2.02$ and heatmap: $\mu = 5.67$, $\sigma = 1.47$). Further statistical analysis (MWU testing) demonstrates that users differ in selecting the video players for class work ($z = 3.45$, $p < 0.01$).

The questionnaire at the end of the study required respondents to give information to provide their most and least favorites about the heatmap. "Important" and "help" are most frequent words mentioned by the participants from our in-lab study. These two words appeared in the Mechanical Turk study as well, although the most frequent word appears in the response is "easy / easily / easier" (34 appearances). For example, participant *p1* mentioned "Identifying the key parts of the video are very easy," and person *p2* commented that "Besides the fact that it was more interesting, it was a lot easier to see those points where others had chosen to play and replay the video. That made it easier to go back to a place that I needed to view again because I was unsure of an answer." Regarding the pain points received from the respondents, similar comments were collected as from the in-lab user study. Several negative reports have shown that the heatmap is distracting and confusing caused by the heatmap. For instance, 7 participants just clearly mentioned "the heatmap was distracting". A few participants pointed out a weakness of current heatmap that there is a correlation between the peak and the importance of the content. One participant *p3* complained "sometimes the most popular times aren't the most useful", and another respondent *p4* suggested that "It needs to be very tailored to the material that is expected to be learned. Otherwise, it could end up being a distraction." Overall, other than the feedback collected from the in-lab user study, there is no unexpected response found from this study. Besides, the positive reviews provide more evidence supporting the heatmap improves effectiveness for searching and navigation.

Two important suggestions were brought up again by some participants. First is to minimize the impact of distracting, it is better to make the heatmap toggle-able (participant *p5* said "the heatmap should be toggle-able"). Second, allowing bookmarks or annotations associated with the heatmap can help avoid confusion and increase search accuracy. For example, participant *p6* suggested that "it is good if we write individual comments on the spotted heatmap to avoid confusion". We will take these suggestions for our future work to improve the heatmap.

In summary, the analysis of the results of this part supports the idea that the heatmap is helpful for learning and enjoyable to use. Another important finding is that almost half of the participants believed the heatmap could help them navigate the video more easily. These results together with what have been found from the previous study have implications for the understanding of how effective of the heatmap and what the heatmap can be used for. On the other hand, people complained the heatmap about the relation with the importance of the content, confusion, and distraction. While receiving the critics, the collected suggestions help to improve the heatmap for the future work.

7.3 Mechanical Turk study summary

In this section, we presented a larger-scale Mechanical Turk user study and analyzed the results in comparison to the previous in-lab study. First, while this study did not confirm the participants could perform better in the quiz, it did partially substantiate that participants felt they understood the video better and felt more confident for completing the quiz when the heatmap is available. Moreover, the difference in user understanding and confidence between the two interfaces is statistically significant. The second significant finding was that the video player with the heatmap is significantly more effective than the video player with basic controls in at least two aspects: usefulness for learning and easiness for finding information. Third, this study has found that the presence of the heatmap attracted a proportion of participants to use it, despite no significant changes were observed in the user's self-reported watching and reviewing patterns. Finally, the investigation of user perceptions of the heatmap has shown that the heatmap is useful and enjoyable for learning. Nevertheless, a few of participants thought the heatmap distracted them from fully watching the video, resulting in confusions about the video. In general, therefore, it seems that the heatmap is useful and easy for the users to learn from the videos.

7.4 Conclusion

In this chapter, we presented the analysis of results collected from Mechanical Turk user studies. In contrast to a relatively small group of participants that were recruited in a controlled lab for the in-lab study, the Mechanical Turk HITs received more participants from the online community. Both of the studies produced common findings and trends. Table 9 summarizes the results of hypothesis testings based on the data collected from these two user studies. There is only one place shows a different statistical result between the two studies. That is, Mechanical Turk study reveals that the user perceived understanding of the video content is significantly better with the heatmap whereas the statistical analysis of the in-lab study results did not show a significant difference.

Table 9: Hypothesis testings

	Lab study	Mechanical Turk
$H_{1.1,null}$: quiz performance	x[a]	x
$H_{1.2,null}$: understanding	x	✓[b]
$H_{1.3,null}$: confidence	✓	✓
$H_{2.1,null}$: useful for learning	✓	✓
$H_{2.1,null}$: easy to find information	✓	✓
$H_{3.1,null}$: first-time watching patterns	x	x
$H_{3.2,null}$: reviewing patterns	x	x
$H_{4.1,null}$: enjoyable	✓	✓

a: x means the null hypothesis can not be rejected.
b: ✓ means the null hypothesis can be rejected.

CHAPTER 8: USER STUDY: MULTIPLE HEATMAPS

The interface of multiple heatmaps introduced in Chapter 3 is designed to facilitate the task of comparative analysis related to video materials. In a video learning environment (e.g. flipped classroom), for example, the instructors might want to identify how the students watched the same lecture video differently. Another example would be that the video owner might be interested in developing a better video based on user interaction analytics. With the multiple heatmap interface, we posit that providing the flexibility multiple heatmaps at once can (1) help the video owner / instructor understand better about the watching behaviors of the video viewers and (2) provide better support for comparative analysis.

In this chapter, we begin with our research questions specifically for the multiple heatmap interface. Then we present the design and process of the user study. Finally, we analyze the data that were collected through the user study and discuss the emerged patterns we found from the study.

8.1 Research Questions

Part of the research questions that are presented in Chapter 3 are specific for the multiple heatmap interface. The study described in this chapter is designed to address these research questions.

- **RQ5** By exploring the interface of multiple heatmaps, how do the class instructors or video owners interpret the heatmaps of user engagement with online videos?

- **RQ6** How does the heatmap of user engagement influence the class instructor in self critiquing the teaching method?

8.2 User study preparations

To begin with, a fully working interface, which provides a variety of selection options for comparison, was implemented for our study. Prior to the start of the study, we obtained the demographic and academic data of the students who have watched the lecture videos that have collected interaction data. The demographic data that is used for this study includes student program level (e.g. graduate and undergraduate), grade, GPA, gender and citizenship (e.g. domestic and international). Then we combined the video's user interaction data with the demographic and academic data of the audience. With the linkage of the data, the interface is functional as what was demonstrated in the prototype (see Figure 4). For instance, a potential participant could compare different watching patterns based on the academic performance of the students.

Additionally, regarding the video selection, we randomly selected a lecture video from the lecture videos that were used in previous HCI classes. Although the video content is not directly connected to the task in this user study, we still would like to present the detail of the video information. The video is roughly 20 minutes long, and the topic of the lecture is talking about analyzing user study data and comparing the rates. From the collected user interaction data of the video, it has received 141 watching sessions from a pool of 101 students including both undergraduate and graduate students across two semesters.

Finally, criteria for selecting the subjects were as follows: the class instructors who had experience with flipped classroom or using video materials for the class. Due to the small group of the professors who can meet this requirement, the class TAs were also eligible for this study as long as they ever had a chance to prepare the class materials and used videos for the class.

8.3 Procedures

The "Think aloud" method was adopted for this particular study. The study was conducted in a semi-structured session [3], where the participants interacted freely with the interface and the experimenter asked pre-determined questions during the session and interviewed the participants at the end. Asking the questions during the session is to prevent the participants from silently participating the study. The experimenter followed up new topics pointed out by the participants. Once the participant finished exploring the interface and fully understood of the design purpose of the interface, there was a short interview with the participant. The study took the participant about 30 minutes on average. The experts, who are the class instructors / TAs in this study, were compensated a $25 Starbucks gift card. The computer's screen and the voice of the participant were recorded to support coding analysis on the results.

To minimize the impact of our interface in making the choices for the comparison, before being shown the interface the participants were asked what would they hope to find out if they were able to track and compare the different patterns from the audience. After collecting the information, the interface was available for the participant to try. By using the prototype (shown in Figure 4), the participant was able to discover, explore and compare in any combination of the following criteria from the students: temporal data across two semesters, program level, grade, GPA, gender and speaking languages. During the "think aloud" process, unless the participant spoke out, the facilitator always gave prompts such as "Why did you make this choice?", "What did you find from the comparison? Did the outcome look as what expected?". Depending on the flow of the interaction, we sometimes asked the participants "Why did you do X? Why did not do Y?". We believe the answers to those questions can allow a deeper insight into the class instructor's usage of the multiple heatmaps.

A short interview was conducted once the participants tried all the comparisons in which they were interested. Following the interview structure in [24], the interview questions included:

1. Did you meet troubles when you used the interface? If yes, what are they?

2. What comparison did you find most interesting? Why are they interesting? How may you use the comparisons?

3. What additional information do you want to know why the pattern is like it is? What are they?

4. Do you think you can use this tool to develop your future lectures / teaching preparations?

- *Follow-up:* If the videos are from a third party, how would you revise your lectures. If you were able to make your videos, what would you like to use the tool to help you?
- *Follow-up:* If the tool is not helpful at all, can you tell us why and what other tools you want to use?

5. Do you have other comments or suggestions?

Taken together with data collected from the "think aloud" process, the feedback of the interview from the instructors was analyzed to investigate: (1) whether the comparison of different watching patterns is helpful for the instructors in understanding the watching behavior of the students on the lecture videos and (2) whether the difference found from the comparison could be an indication for them to revise their current lecturing approach and lecture preparation. The analysis of the entire feedback from the instructors would help us address RQ5 and RQ6 listed at the beginning of the section. With the interface of multiple heatmaps, the class instructors are expected to understand better about how the users' video watching patterns vary across comparative data selections. The instructors are also expected to discover the potential improvements for the teaching materials after they explored and compared different combinations of patterns.

8.3.1 Study Participants

Ideally, the instructors who have taught HCI class with lecture videos are fit to the study because they know the video content pretty well and they can connect the video content with the heatmaps for comparative analysis. As a result, we started looking for those instructors by sending out emails. We also collected recommendations from the participants for finding more instructors who had similar teaching experience with using videos, so we reached out more instructors through the recommendations. Those instructors who have no HCI background are qualified for our study due to the fact that the main purpose of the study is to evaluate how people interpret the heatmaps and how people potentially use the interface. The video content is not very important for this user study.

Ten participants signed up the study, and each one received a $25 Starbucks gift card for the compensation. The ten participants consist of seven class instructors and three TAs. All of the TAs are Ph.D. students who have had TA experience for HCI class. Four out of the seven instructors directly taught the class and used the video in the class. The rest class teachers are not experts in HCI despite the fact that they used online videos for their class.

8.4 Results and Analysis

The study was not about the particular interface of multiple heatmaps, but was designed to explore how the instructors react to different heatmaps of user engagement data with online lecture videos. The experimenter explains the colors on the heatmap and how to add or remove heatmaps from the interface. The participants seemed to have no issues to understand and use the heatmaps. They were able to understand the data had been collected from students who enrolled previous HCI class and the colors on the heatmap represent the degree of views from those students. By exploring the heatmap, they could identify user engagement on different segments through the colors. We observed that all of the participants could correctly interpret the heatmap that they created every time.

A coding analysis was conducted with a qualitative method [3]. Two coders (one of whom was the interviewer) iteratively coded the data, with a joint review after each iteration. The goal of the analysis is to find out the emerged concepts expressed by the participants. The concepts are organized into three general themes:

- **Theme 1** User perceptions of multiple heatmaps. Through the analysis, we were able to identify user perceptions of the heatmaps in different aspects.
- **Theme 2** The applications of multiple heatmaps. All of the participants mentioned the potential applications of multiple heatmap comparisons.
- **Theme 3** The reasons of finding interesting comparisons. All of the participants explained why the comparison of watching patterns is interesting. Several concepts were discovered in this theme despite it is hard to claim there is a common interesting comparison among all the participants.

A couple of concepts have emerged in each theme during the coding process. The concepts were found from at least 5 participants. The detailed thematic analysis will be presented in section 8.4.2, section 8.4.3 and section 8.4.4

8.4.1 Before the study starts

The features of heatmap(s) seem to meet many participants' expectation. Before the study began, the participants were asked what to do if they were able to track different watching patterns from the students (an interview question before the participants could see the interface). *All* participants indicated that if there was such a tool they would like to compare the patterns and interpret the segments that significantly differentiate others. Participant *P10* mentioned "...The second thing would be to look at the hot spots, em.., to see either what was the most engaging but also maybe it was confusing. Then I have to what to interpret...". Even though some participants suggested that tools like a spreadsheet are enough for them to compare how the students watched the video if user interaction data is logged in the spreadsheet, we still believe that a better visualization can be a more powerful tool for the instructors to analyze student activities. Three participants wanted to have a tool that could help them find correlations between the patterns and the video content. Interestingly, four responders wanted a tool which could help them identify whether the students had watched the video or not. It makes sense that a instructor has such an interest particularly when watching the lecture video is part of the class assignment in a flipped classroom.

On the other hand, the current heatmap interface does not support some specific functions identified by participants. For example, a few participants were looking for the number of loops, pauses or plays on a segment. For instance, Participant *P3* stated "...numbers of times they stopped or looped, or went back or that kind of thing and then number of occasions that they rewatched the whole video, when they watched the video, that kind of thing,...". The user could not see these statistics because we chose not to display the numbers of the interaction events on the heatmap since the colors are the indicators for the level of engagement. However, adding numbers to the heatmap could be a fascinating study in the future work. In this research, we only focus on the design as discussed in Chapter 3.

An implication of the responses from the very first question is the possibility that the approach we adapted to display viewing patterns could potentially help the instructor with teaching methods and class preparations.

8.4.2 Theme 1 Perceptions of multiple heatmaps

When the participants were interacting with our interface during the study, everyone tried to correlate the heatmap patterns with the selected criteria. For example, if a participant saw the watching pattern from the group of international students is so different from the domestic students, the participant would link her / his previous teaching experience with the difference and try to analyze the reason why the language caused in this particular video. This theme mainly emerged when the participant was exploring our interface.

Table 10: Theme 1 Perceptions of multiple heatmaps

	Sources[a]	References[b]
Matching teaching experience	5	8
Surprised by the heatmap(s)	3	5
Uncertainty	2	3

a: Sources denotes the number of participants mentioned the concept.
b: References denotes the occurrences of the concept.

Matching teaching experience Once a specific group of students has been chosen, for example, a group of students whose grade was B for the course, the generated heatmap is a visualized watching patterns of the students from the particular group. Our coding process identified a common pattern throughout the exploration of the heatmaps. Five instructors confirmed some watching patterns match their impressions about the selected group of students. An example of a statement ascribing this match was "I think the outcome makes sense. It looks like graduate students worked more consistently versus the undergraduate students may be just click in the areas and watch." This participant was comparing the heatmap from graduate students versus undergraduate students. The heatmap could help the instructor confirm what they have experienced as well. *P4* explained "it confirms my suspicions that international students are watching the videos a lot more, ..., because of the language." when *P4* was examining the difference between domestic and international students. One more interesting finding along with concept is that the majority of the participants expressed interesting or exciting motion when they realized the patterns.

Surprised by the heatmap(s) Whereas the heatmaps confirmed the teaching experience of the participants, some heatmaps also surprised a few class instructors (3) since the output of the heatmap does not match their experience. One participant said "That surprises me. I was absolutely thinking the females are more attentive, that actually dispelled some notions I have about male and females students. Look, the males stay quite engaged with it almost all the way where the women sort of stopped earlier." Apparently, the instructor thought female students could be more hardworking than male students. However, the comparison of heatmaps reveals that male students watched more times than girls. So the heatmap in our interface can not reject the overall impression in the instructor's mind, but it could validate how the chosen students watched the selected video.

Uncertainty Rather than accepting or rejecting the patterns shown on the heatmap, sometimes, the instructors were not sure about why the pattern happened in a particular segment. So this is the third concept that was identified in this theme: participant was uncertain about what they've found from the heatmaps. An example of such a statement is that "we want to know why it could be caused by whether or not the students are mature enough. You know, the young guys are losing .., or it is because the content is designed badly suited for the graduate students, but I don't know." For some parts of the videos ended up having some unusual patterns, the participants were trying to find evidence why the pattern happened. But some participants were stuck in finding the problems. They implied that they were able to do more research if more data could be provided such as show the number of clicks, pause or looping for the segment.

In this theme about user perceptions, three concepts were identified. Together with the concepts that emerged, we also observed that the participants had a pleasure experience with the multiple heatmap comparisons. "Interesting" (120 occurrences) is the most often mentioned words when the participant saw the different heatmaps. Some other positive feedback includes "fascinating" (11 times), "fantastic" (9 times).

8.4.3 Theme 2 Applications of multiple heatmaps

Throughout the study process and the interview, all of the participants stated that they would be willing to use our multiple heatmap interface and also discussed the potential benefit they could get from the interface if it is available for use. The feedback was primarily collected from the interview question 4, in addition to the responses collected during the study when the participants were exploring the multiple heatmaps. Two primary purposes of using the multiple heatmap interface strongly emerged during the coding process. First, if the video is not the instructor's own video, many of the participants indicated that they would improve their class materials or enhance the teaching methods. Second, for some of the participants who made their videos for the class, they would like to develop the video content based on the results of comparative analysis of the multiple heatmaps. Because of the two primary applications, two sub-themes were identified for the applications of our interface: 1) Making teaching better 2) Making video better.

8.4.3.1 Theme 2.1 Making teaching better

Three concepts were identified during coding as falling within this theme. Table 11 summarizes the how many people and how many times each concept was referenced during the study.

Table 11: Theme 2.1 Change teaching

Concepts	Sources	References
Make explanations	7	11
Make quiz	3	6
Show heatmaps to students	3	3

Making explanations A majority of participants (7) talked about how they would explain the video content based on what they found out from the comparisons. Surprisingly, both the hot and cold spots gained attention from the instructors. Among the seven persons, three of them were concerned about the hot spots discovered from the heatmap since more watchings on the same segment may indicate student might have a difficult time on understanding. Therefore, it is worthy to emphasize the content again in class. On the contrary, four participants were worried about the spots that have fewer views but contain the important content from the video. Cold spots could tell the four participants that many students neglected the important part. Thus, they want to bring up those particular segments and explain them again in class. In any case, their point is that whenever they observe an unusual pattern from the comparison, they would try to figure out the reason behind the pattern and address the problem. Like *P10* said "I would probably just start with that as a baseline data and look at where the low spots were. First of all look at the overall patterns, so if I do this before class, I look at historical data and look at the overall patterns and then look at where do I have concerns, ... , I would find out what am not I happy with, with these patterns, what can I do is trying to combat them."

Making quiz This is the second concept that strongly emerged during the coding process. Rather than making explanations on both hot and cold spots in the first concept, the instructors are more willing to make quizzes on less watched segments. All of the participants, who indicated to use a quiz to make or enforce the students to learn, would like to place the quiz on the spare spots. *P2* explained: "So when I am in the class, ... , for the clicker quiz I could put more focus on the second part (where contains a lot of green areas), they could get the chance to learn what they missed.”.

Showing heatmaps to students Another common use of the multiple heatmaps is to be an example that could be shown to students directly. One feature of our interface is allowing instructors to explore and compare different watching patterns based on the students' academic performance. It would be interesting if there is a difference between good performance students and poor performance students. Indeed, a few participants observed this during the study. When *P4* realized the different watching patterns, *P4* was excited and commented: “I would actually like to show students this, I would show students, look, these are the students who are getting the A, these are the students who are getting C, ..., so I would use that as a way to inform the students about how they can doing better.” Another type of showing the heatmaps to the students is to introduce the highly-watched parts and have a deep discussion with the students about this part. For example, *P9* mentioned: “so when I'm in the class I could talk to the students about the part that was red. And enquire with them whether they thought that was something that was hard to understand.”

8.4.3.2 Theme 2.2 Making video better

Comparing different heatmaps or the difference inside a heatmap, the participants can potentially identify unusual patterns. Another theme which is part of the application of the heatmap(s) is refining the video. It was not only pointed out by the participants who made their videos for class but also was suggested by some other instructors during the study. Two concepts were referred by the participants while the interview, emerged during coding as falling within this theme, see Table 12.

Table 12: Theme 2.2: making a better video

Concepts	Sources	References
Refine the video	6	8
Shorten video / less content	4	5

Refine the video The concept was coded for a general purpose of refining the video, which includes many possible approaches. Five participants mentioned about video content modifications based on the results of comparative analysis of multiple heatmaps. Throughout the study, we observed that people were able to discover problems in the video if there exists a difference between heatmaps. As a result, a few participants suggested remaking videos to address the problems causing fewer views. For example, one participant commented "so if I was making a lecture video and I use it in one semester, and I saw that it seems to have a lot problems in confused students through the interface, then that would give me an opportunity to not just remake the video but to remake video knowing exactly what's wrong with it...". Another use of multiple heatmaps that emerged is to change the video content to motivate the students. Three instructors believed that they could incentivize students to pay attention and focus on the video after they figure out the problem. Sometimes, students do not watch the video just because they are not motivated. Like one participant expressed, " I'm not going to penalize them for not watching it but everyone needs incentives and motivation." Another approach to motivate the students on few-watched segments is to place some questions on it. "... So I try to somehow make them like several important segments, okay. I also want to embed like a quiz or like questions in the videos." said by *P5*. Apparently, this participant would like to add a question in the video and see if it could change student watching behavior.

Shorten video / less content Since the video used in the user study is a relatively long lecture video which has 20 minutes, from the heatmap many participants noticed that the students watched the video usually lost attention after the half of the video. Plus their previous teaching experience verified the observation from the heatmap. Thus, one concept, shortening the video or placing less content in a video, stood out. Four participants commonly stated that making a shorter video could help student focus more on the video content. For instance, after *P6* had realized previous students just dropped off in middle of the video, she stated "Basically it shows me that it is very very easy to see that this point somewhere people dropped off the video. Really it tells me that not to create a video more than 10 minutes long. Because a lot of people seems to drop off." While some instructors have realized that they should make short videos, they mentioned one common problem existing for online education. Unlike the traditional classroom environment, they are a lack of instant feedback from the audience. *P6* commented "One worry I have, for the online-course, is I do not have to them feedback. I do not know whether or not I'm running too fast or too slow, which concept students do not understand. This is a very bad thing, this is very, very bad thing from their educational side, ..., On the network, we usually ran faster because there is no feedback. I don't repeat.". This problem may result in too much content in a short video. So with the help of the heatmap, the instructors could get feedback from online videos, then they could adjust the information in a video. *P6* continued the solution and said " I will try to, you know, to make the video so even if I make a 12 minutes video, I probably will talk about a few things."

8.4.4 Theme 3 Why comparisons are interesting

Since the main task of the study is to compare the different heatmaps that represent the watching patterns of groups of students, all the participants experienced the comparison based on their preferred selections. On average, each participant selected at least three or four comparisons. The comparison includes all types of combinations. We initially intended to collect the most interesting comparison that the participants could identify through our interface. Throughout the interview question (question 3), participants described the most interesting comparisons and the reason why the comparison is interesting. Everyone's feedback to the question was different from each other. For example, three participants mentioned comparing the heatmaps based on gender was interesting, but some of them compared in a context of one specific semester while some ones were interested in all semesters. It is difficult to draw a conclusion for the most interesting comparison. However, we did observe an emergent theme about the reasons why they were interested in the most interesting comparison. More importantly, three concepts (see Table 13) were captured in this theme based on the participants' descriptions.

Table 13: Theme 3 Reasons for interesting comparisons

Concepts	Sources	References
Interesting results	4	4
Confirmed / disconfirmed experience	3	3
Help improve teaching	3	3

The respondents commented the reasons why they were interested in a particular comparison from two perspectives: the visual results from multiple heatmaps and the functionalities of the multiple heatmaps. The first two concepts were captured from the participants who are interested in the results of comparing multiple heatmap, whereas the last concept is about how the interesting comparison could help future teaching.

Interesting results Four participants were either excited or interested in the heatmaps when the generated heatmaps represent a huge and obvious difference. For example, a few participating instructors compared the heatmap coming from the group of students with grade A, B, C, and F. One reason for the comparison to be an interesting one is that the watching patterns shown by the comparison based on the grades showed a huge difference and caused the curiosity for the participants. Students with A or B watched the same video very differently from the students with C, or F. One participant commented that "the comparisons when there was a big difference with the most interests. Not necessarily, it was most interesting because of what I could select but most interesting because of the difference." Another reason is that some participants thought comparing the watching patterns could also help the participants develop the course or the video. If the heatmaps really help the instructors see a problem for students with grade C, the instructors are willing to modify the video or the teaching to help poor-performance students.

Confirmed / disconfirmed experience This concept emerged as part of user perceptions to the multiple heatmaps. During the coding process of the interview question 2, this pattern was confirmed again because of the watching patterns that were displayed by the heatmaps. Three participants described whatever their chosen pattern is most interesting is because it confirms their teaching experience, while there is one participant who noted that the most interesting pattern surprised him since it does not match his teaching impression on the selected group of students. An example of the statement that confirming the instructor's teaching experience is "... because it confirmed my theory of what I've observed in my class and it's good to know that I'm right, that is really what's happening".

Help improve teaching The rest of participants (3) answered the interview question in the way of considering the functionalities of comparing the watching patterns from the heatmaps. They mostly pointed out that such comparison is a kind of useful tool to develop and improve their teaching. Most of them considered the grade based comparison is the most interesting one because the different patterns could help them find out the potential problems. An example of such a statement is that "I want to know how a good student work compared with not really good ones, that could help me plan accordingly, ...I mean we don't want to watch students for fun, we want to know what is the problem so we can make it better.”

8.4.5 Discussions

While people were excited about the comparison of the heatmaps, we received a few complaints and suggestions concerning missing functions and issues. The most complained part is not really about the heatmaps itself, yet it is about some other features in our interface. For example, the information button added to our interface was designed to show the data information about the heatmap when the user is hovering over the button. The information data includes the total number of viewers, the number of females / males, etc. Although it is not the focus of this work, it is worth to point out since the function was frequently used during the study. The major concern is that the information is a little small compared to the heatmap and it disappears very quickly. Because the information is helpful for users to remember what they have selected, it is important always to display the information in a convenient way. Another possible issue is the time selector for generating the heatmaps, most of the participants only used the semester as a filter to create the heatmap and only a few participants recognized there are time selectors for setting a specific starting and end time for generating a heatmap. It was never used in the user study since none of participants knew or remembered the specified date the video was assigned to watch. However, the respondents are still willing to have it if we could provide the timeline for the video when it was assigned or other important dates such as the mid-term or final exam. For future work, therefore, it will be good to keep the time selector if the detailed timelines are available.

Regarding missing features, a few feasible and useful suggestions were received, and we pick two representative suggestions to illustrate. First, a couple of participants would like to know more statistics about the heatmap or the fine-grained data on the heatmap. Even though they could find out the overall data about the heatmap by hovering the information button; some participants believed that if we can show a detailed number on each spot of the heatmap that would be helpful as well. Participant *P6* described that "If more statistics I can look from here at this spot, how many students watched there, that would be interesting and useful." Second, some participants thought the grade is not enough. So if we can link the quiz or the homework performance with the heatmap, that would be helpful for them to figure out the potential problem of teaching. An example of such a statement is that "It will be great if we can link the video segments with the grade of the homework or the questions of the final exam. Then we can see, you know, whether or not, they really understand this.".

8.5 Conclusion

To sum up the study for evaluating our multiple heatmap interface, all of the participants believed that there is no really difficulties for using the interface and they were able to interpret the heatmaps correctly (RQ5). The study shows that the participants agreed that the heatmaps could potentially be a tool for them to discover teaching flaws in the class or on the video particularly when they found a huge difference between different watching patterns (RQ6). Despite a few of the concerns the participants described, the comparison of multiple heatmaps overall can help them develop the class material, teaching methods or produce a better lecture video, which is actually the intention of creating our interface.

CHAPTER 9: CONTRIBUTIONS

9.1 Thesis statement

Our goal was to address the general research question of how novel approaches to interaction with online video can better support end users for active, goal-directed use of video content. To investigate this research question, we focused on one probe into the design space, evaluating interaction interfaces based on user engagement analytics. In order to understand the impact of interaction designs in these contexts, we conducted user studies evaluating the following thesis. *The design features of the user engagement heatmap provide better task support than standard online video tools for user interaction and active consumption of online videos in goal-directed tasks such as information finding or comparative analysis.*

In this dissertation, we have presented a broad investigation of the problem of supporting end users for goal-directed use of video content such as information finding and comparative analysis. The heatmap of user engagement is a visualization of viewers watching patterns. To determine the colors of the heatmaps, we compute the degree of user engagement with different segments in a video and assign the degree to a range of colors. A few studies for the heatmap are proposed to evaluate how it affects user's video watching behaviors and understanding the user engagement data, while some other studies are designed for our multiple heatmap interface.

9.2 Dissertation Summary

The work of heatmap of user engagement with video is essentially the visualization of user engagement of video. It is motivated by the fact that understanding user's video watching pattern might indicate important, interesting, confusing or boring segments within a video. If the user engagement is discovered correctly, the colors in the heatmap are the indicators and provide a new way for video navigation. Compared to previous video-as-a-whole user engagement measurement, user engagement is measured based on a segment level in this research. The study of our research has shown that the interface of the heatmap of user engagement could support information finding, content understanding, and comparative analysis.

In this dissertation, we start the importance of online videos and then point out the problem space with the current prevail online video applications in Chapter 1. Chapter 2 presents a background and related work on online video players, video interaction and video engagement. The pioneer works shown in Chapter 1 and Chapter 2 are the fundamental work for this dissertation. Chapter 3 describes the preparation for this research. We present the data collection of user interaction, investigates the approach of computing user engagement for video, and finally explores the designs for interacting with user engagement. In this chapter, we also illustrate the design goals of the research probe and how the presented interface works. In Chapter 4, we examine the research questions for the research probe presented in this work. Additionally, we describe the detail of experiments for the evaluations of the proposed prototypes. Chapter 5 shows a pilot user study which was conducted in class. After receiving the feedback from the pilot study, we modified the user interface and user study design for a user study conducted in a lab setting. We describe the research findings that were discovered from the lab study in Chapter 6. The results from the user study finished on Amazon Mechanical Turk are presented in Chapter 7. Both the in-lab and Mechanical Turk user study were designed to evaluate the interface with a single heatmap. For the evaluation of the interface with multiple heatmaps, the details and results are written in Chapter 8. Instead of focusing on the viewer's side, this study is designed to evaluate how the multiple heatmap interface impacts the class instructors in teaching.

9.3 Research Contributions

The primary research goal of this dissertation is to enable and support more active user interaction with online videos, particularly information finding, content understanding, and comparative analysis. The contributions of this research work contain two major parts.

1. The first part is that we have examined the approach of computing the *degree of user engagement* based on different metrics and its application for the visualization of user engagement with video. The heatmap of user engagement with video proposed in this work is the representation of video watching patterns and the indication of the points of interests in videos. We present two different interfaces with the heatmap to investigate how the heatmap impact user interactions with video. First, we demonstrate an online video player integrated with a single heatmap. The design goal of the first interface is to provide bet-

ter user interaction for users to perform goal-directed tasks such as information finding from the video and content understanding. Second, we present an online video player with multiple heatmaps, in which users could add and remove the heatmaps as many as they want. The heatmap is generated based on previous audience's demographic and academic data. The design purpose of the interface is to support comparative analysis better, more specifically, to support the class instructors to compare different watching patterns from various groups of students.

2. The study results show that our heatmap interfaces could better support video interactions for information finding, content understanding, and comparative analysis. The research results are coming from the studies we have conducted for the two heatmap interfaces. The evaluations include a series of user studies for accomplishing goal-directed tasks such as comparative analysis and information finding in the education domain. We have conducted two formal user studies from a lab setting to an online community (Amazon Mechanical Turk) for the evaluation of the single heatmap interface. The evaluation primarily focuses on how the heatmap changes new users watching patterns, how the heatmap helps understand the content and look for information from the video. For the second interface with multiple heatmaps, we conducted a user study with 10 class instructors to evaluate how the multiple heatmap interface impacts the class instructors in teaching or making online videos. From the user study for evaluating the multiple heatmaps, the research questions to be answered include

how it affects the professor's self-critiquing the teaching method and how it is interpreted by the class instructors.

All of the research findings from the in-lab and Amazon Mechanical Turk studies were discussed in Chapter 6 and Chapter 7 respectively. In Chapter 8, we listed the analysis results from the user study for the multiple heatmap interface. In the following sections, we conclude what have been found from the analysis. Similar to the descriptions of the studies, we present the research findings from the two user groups.

9.3.1 Learners Group

- *RQ1 How does the heatmap of user engagement impact video learner's understanding of the lecture? (content understanding)*

  In both the lab and Mech Turk studies, the heatmap does not significantly improve user's understanding of the video content based on their quiz performance. However, both studies have shown that the participants did finish the quiz significantly more confident when they use the video play with the heatmap ($p < 0.05$). The Mech Turk study also reveals that the participants could feel understand significantly better with the heatmap.

- *RQ2 Is the heatmap control effective for learning tasks based on video material? (effectiveness)*

  With the heatmap, users from both studies were able to find the needed information from the video quickly. Moreover, the heatmap is more helpful for learning. Compared to the video player with basic controls, both the lab study and Mechanical Turk study show that the difference regarding effectiveness is statistically significant.

- *RQ3 How does the heatmap impact user's watching pattern? Do users use a different way to navigate the video, particularly when they search the content from the video? (watching patterns)*

  The heatmap has impacts on user's watching and navigation patterns despite there is no significant difference found between the video player with the heatmap and the one without the heatmap. Both studies have identified that people usually do not interact the heatmap frequently when they first watch the video, but they are more likely to use the heatmap as a navigation tool for actively searching information when they review the video.

- *RQ4 How do the video viewers (learners) interpret the heatmap of user engagement in a lecture video? (user perceptions)*

  Overall, the heatmap is significantly more enjoyable and less frustrating to use in both studies. The result of the evaluation of user perceptions also strengthens the idea that the heatmap is helpful for learning. The process of analyzing the data related to this research question leads us a challenge to design a heatmap to keep without distracting the users.

9.3.2 Instructors Group

- *RQ5 By exploring the interface of multiple heatmaps, how do the class instructors or video owners interpret the heatmaps of user engagement with online videos?*

  There was no issue for users to understand and interact with the heatmaps. All users in the study were able to correctly interpret the colors from the heatmap and correlate with their previous teaching experiences. The difference of watching patterns, visualized by the heatmaps, either matches the participants teaching experience or surprised the participants. In any case, the professors could learn from the difference and use it as a means to develop teaching.

- *RQ6 How does the heatmap of user engagement influence the class instructor in self-critiquing the teaching methods?*

  The study shows that the comparative analysis through the heatmaps can potentially be a tool for the professor to discover teaching flaws in the class or on the video particularly when they found a huge difference between different watching patterns. First, if the class instructor did not create the lecture video, the concepts emerged regarding the applications of the multiple heatmaps include making better explanations, making the quiz to attract students attentions, and just showing the different patterns to the students. Second, if the instructors could be able to make their videos, they are willing to refine the video shorten the video length, depending on the findings from the comparison of different heatmaps.

9.4 Future Work

All the studies designed for the single heatmap evaluation were completed in the short term. The participants only had a chance to interact one video with the heatmap in our study. It would be interesting to study how the heatmap impact users in learning and information finding in a longer time span. For example, one way is to use the between-subjects design to evaluate the heatmap in one entire semester. We can give the access to the heatmap to one section of the class while the other section of the class can use the regular video player for watching the class videos. Doing this could give more opportunities to use the heatmap for different videos so that it can trade off the impact of the video content. At the end of the study, it is possible to compare the difference in performance between the two sections of classes. Another benefit of doing is that we can receive more reliable data from students as watching the video is part of the assignment, which was discussed this in our pilot study conducted in class.

Although all of the work that has been done in the dissertation focuses on a specific domain: the online education domain for the heatmap of user engagement, we believe the results apply to other domains related to video applications. More specifically, the interface with the heatmap of user engagement supports more active user interaction with the video for goal-directed tasks in a different domain. In future investigations, it might be possible to use a different video domain to evaluate our heatmap interface. Plus it would also be interesting to expanding the results to other applicable domains. We can use the exactly same study design to conduct the user study and investigate the same research questions.

A further study with more focus on supporting annotations in our interface is also suggested. From the user studies, we observed that many participants would like to annotate on the heatmap since the annotations could help them memorize the points of interests and eliminate inaccurate hot spots. It would be worthy to study how the heatmap could be more helpful and effective for learning with adding the feature of supporting annotations.

Overall this research has shown that the heatmap interaction is a promising approach to enhance user interactions with online videos in a representative domain for education. The results of this research have implications for future practice. We envision that video interaction data analytics can better support goal-directed tasks, such as information finding, content understanding, and comparative analysis, in various domains of video applications.

REFERENCES

[1] R. M. Baecker, D. Fono, and P. Wolf. Towards a video collaboratory, 2006.

[2] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. A quest for an internet video quality-of-experience metric. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 97–102. ACM, 2012.

[3] A. Blandford. Semi-structured qualitative studies. Interaction Design Foundation, 2013.

[4] C. C. Bonwell and J. A. Eison. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC, 1991.

[5] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[6] G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 21–30. ACM, 2009.

[7] M. G. Christel, M. A. Smith, C. R. Taylor, and D. B. Winkler. Evolving video skims into useful multimedia abstractions. In *Proc. of the SIGCHI conference on Human factors in computing systems*, 1998.

[8] D. Coetzee, A. Fox, M. A. Hearst, and B. Hartmann. Should your mooc forum use a reputation system? In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 1176–1187. ACM, 2014.

[9] Crazyegg. Crazyegg. `http://www.crazyegg.com/`.

[10] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416. ACM, 2007.

[11] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[12] W. Ding and G. Marchionini. A study on video browsing strategies. 1998.

[13] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang. Understanding the impact of video quality on user engagement. *ACM SIGCOMM Computer Communication Review*, 41(4):362–373, 2011.

[14] S. M. Drucker, A. Glatzer, S. De Mar, and C. Wong. SmartSkip: Consumer level browsing and skipping of digital video content. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2002.

[15] A. Edmonds, R. W. White, D. Morris, and S. M. Drucker. Instrumenting the dynamic web. *Journal of Web Engineering*, 6(3):243, 2007.

[16] A. Girgensohn and J. Boreczky. Time-constrained keyframe selection technique. In *Multimedia Computing and Systems, 1999. IEEE International Conference on*, volume 1, pages 756–761. IEEE, 1999.

[17] 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*, pages 3–12. ACM, 2008.

[18] P. J. Guo, J. Kim, and R. Rubin. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 41–50. ACM, 2014.

[19] W. Hürst. Interactive audio-visual video browsing. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 675–678. ACM, 2006.

[20] 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*, pages 563–572. ACM, 2014.

[21] 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*, pages 31–40. ACM, 2014.

[22] S. S. Krishnan and R. K. Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking (TON)*, 21(6):2001–2014, 2013.

[23] C. Latulipe, E. A. Carroll, and D. Lottridge. Love, hate, arousal and engagement: exploring audience responses to performing arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1845–1854. ACM, 2011.

[24] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010.

[25] F. C. Li, A. Gupta, E. Sanocki, L.-w. He, and Y. Rui. Browsing digital video. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2000.

[26] S. G. Lynn, D. R. Olsen, Jr., and B. G. Partridge. Time warp football. In *Proceedings of the Seventh European Conference on European Interactive Television Conference*, EuroITV '09, pages 77–86. ACM, 2009.

[27] I. S. MacKenzie. *Human-computer interaction: An empirical research perspective*. Newnes, 2012.

[28] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 227–236. ACM, 2011.

[29] J. Matejka, T. Grossman, and G. Fitzmaurice. Swift: reducing the effects of latency in online video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 637–646. ACM, 2012.

[30] J. Matejka, T. Grossman, and G. Fitzmaurice. Patina: Dynamic heatmaps for visualizing application usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3227–3236. ACM, 2013.

[31] S. Mongy, F. Bouali, and C. Djeraba. Analyzing user?s behavior on a video database. In *Multimedia data mining and knowledge discovery*, pages 458–471. Springer, 2007.

[32] Mozilla. `https://heatmap.mozillalabs.com/`.

[33] B. A. Myers, J. P. Casares, S. Stevens, L. Dabbish, D. Yocum, and A. Corbett. A multi-view intelligent editor for digital video libraries. In *Proc. of the ACM / IEEE-CS joint conference on Digital libraries*, 2001.

[34] D. R. Olsen and B. Moon. Video summarization based on user interaction. In *Proceedings of the 9th International Interactive Conference on Interactive Television*, EuroITV '11. ACM, 2011.

[35] G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5):411–419, 2010.

[36] P. Parnes, K. Synnes, and D. Schefstrom. mstar: Enabling collaborative applications on the internet. *Internet Computing, IEEE*, 4(5):32–39, 2000.

[37] A. Pavel, D. B. Goldman, B. Hartmann, and M. Agrawala. Vidcrit: Video-based asynchronous video review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 517–528. ACM, 2016.

[38] R. Pea and R. Lindgren. Video collaboratories for research and education: An analysis of collaboration design patterns. *IEEE Transactions on Learning Technologies*, 1:235–247, 2008.

[39] L. W. Perna, A. Ruby, R. F. Boruch, N. Wang, J. Scull, S. Ahmad, and C. Evans. Moving through moocs: Understanding the progression of users in massive open online courses. *Educational Researcher*, 43(9):421–432, 2014.

[40] S. Pongnumkul, J. Wang, G. Ramos, and M. Cohen. Content-aware dynamic timeline for video browsing. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 139–142. ACM, 2010.

[41] G. Ramos and R. Balakrishnan. Zliding: fluid zooming and sliding for high precision parameter manipulation. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 143–152. ACM, 2005.

[42] K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and J. M. Jose. Video browsing interfaces and applications: a review. *Journal of Photonics for Energy*, 2010.

[43] D. Seaton, I. Chuang, P. Mitros, D. Pritchard, et al. Who does what in a massive open online course? *EducationXPress*, 2013(9):1–1, 2013.

[44] V. Singh, C. Latulipe, E. Carroll, and D. Lottridge. The choreographer's notebook: a video annotation system for dancers and choreographers. In *Proc. of the 8th ACM Conference on Creativity and Cognition*, 2011.

[45] A. F. Smeaton, C. Gurrin, H. Lee, K. McDonald, N. Murphy, N. E. O'Connor, D. O'Sullivan, B. Smyth, and D. C. Wilson. The físchlár-news-stories system: personalised access to an archive of tv news. 2004.

[46] Wistia. Wistia Heatmap. `https://wistia.com/`.

[47] Youtube. `http://www.youtube.com/`.

[48] YouTube. Youtube statistics page. `http://www.youtube.com/yt/press/statistics.html` Accessed Jan 22, 2017.

[49] 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*, pages 382–391. ACM, 2003.

APPENDIX A: Experimental Materials for Single Heatmap User Study

Demographic Survey

What is your gender? *

- ◯ Male
- ◯ Female

What is your age range? *

- ◯ 18-24
- ◯ 25-34
- ◯ 35-44
- ◯ 45-54
- ◯ 55+

What is your ethnicity? (Check all that apply) *

- ◯ American Indian or Alaskan Native
- ◯ Asian or Pacific Islander
- ◯ Black or African American
- ◯ Hispanic or Latino
- ◯ White or Caucasian
- ◯ Other: ____________

Are you an undergraduate or graduate student? *

- ◯ Undergraduate
- ◯ Gruduate
- ◯ Other:

Do you watch lecture videos for learning? *

- ◯ Yes
- ◯ No

Have you heard Video Collaboratory? *

- ◯ Yes
- ◯ No

Quiz question for first selected video lecture

The user mental model is developed through interacting with the system. *

- ◯ True
- ◯ False

Which is the term for when a user has the right model of how the system works, but they accidentally do the wrong thing? *

- ◯ Overestimation
- ◯ Blunder
- ◯ Slip
- ◯ Mistake

A design should be as different as possible from previous representations, for the sake of innovation *

- ◯ True
- ◯ False

The mental models of people incite them to use a new technologies as they would use the old physical system *

- ◯ True
- ◯ False

Post survey after watching the video without heatmap

Have you watched this video before? *

- ◯ Yes
- ◯ No

How much do you feel you understood this video? *

	1	2	3	4	5	6	7	
Not at all	◯	◯	◯	◯	◯	◯	◯	Completely understand

How confident do you feel about your answers to the quiz questions? *

	1	2	3	4	5	6	7	
Not confident	◯	◯	◯	◯	◯	◯	◯	Very confident

Check all that apply. Which of the following(s) describe the way you watched the video for the first time? *

- ☐ Watched the full video in one go without stops.
- ☐ Paused, watched, and repeated the process
- ☐ Watched, skipped then watched, and repeat this process.
- ☐ Randomly selected a start to watch and went back if needed, and repeated this process
- ☐ Other: ____________

Check all that apply. Which of the following(s) describe how you used with video when answering the quiz questions? *

- ☐ Answered the quiz without reviewing the video
- ☐ Watched the full video again
- ☐ Tried to remember the place in the video that was related to the question, and navigated to it.
- ☐ Randomly selected a place to watch, and then randomly selected another one if it wasn't relevant.
- ☐ Other:

Quiz question for second selected video lecture

Why is the line between reality, marketing, and news impossible to discern? *

- ◯ Hundreds of news channels compete for viewers
- ◯ Media and advertising operate on the same wavelength
- ◯ People prefer the copy of reality.
- ◯ All of the above

In terms of news and advertising, people willingly choose which of the following? *

- ◯ The truth
- ◯ Deception
- ◯ Trumpisms
- ◯ News bloopers

Who said that the proliferation of images makes the media untrustworthy? *

- ◯ Jean Baudrillard
- ◯ Stanley Kubrick
- ◯ Michel Foucault
- ◯ Paul Mccartney

Post survey after watching the video with heatmap

Have you watched this video before? *

- Yes
- No

How much do you feel you understood this video? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Completely understand

How many times did you hover or click on the heatmap for video navigation? *

- Not at all
- A little (less than 10 times)
- A lot (more than 10 times)
- Other:

How confident do you feel about your answers to the quiz questions? *

	1	2	3	4	5	6	7	
Not confident	○	○	○	○	○	○	○	Very confident

Check all that apply. Which of the following(s) best describe the way you watched the video for first time? *

- [ ] Watched the full video in one go without stops.
- [ ] Paused, watched, and repeated the process
- [ ] Watched, skipped then watched, and repeat this process.

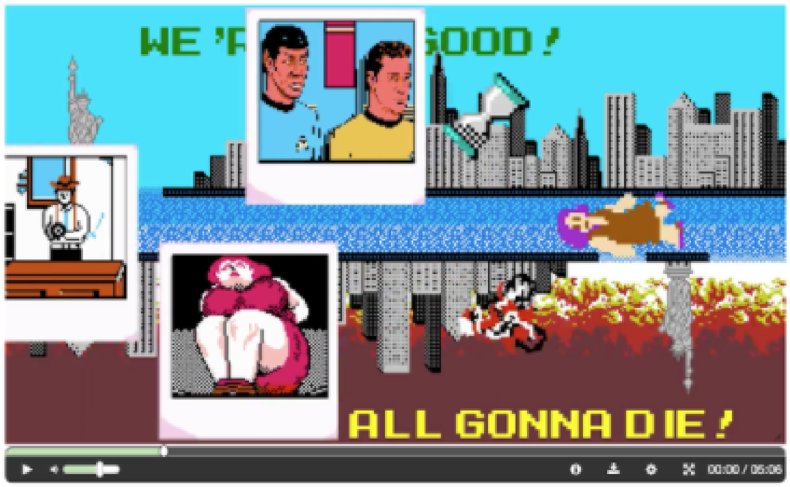
- [ ] Used the heatmap to help with navigation
- [ ] Other:

Check all that apply. Which of the following(s) best describe how you used with video when answering the quiz questions? *

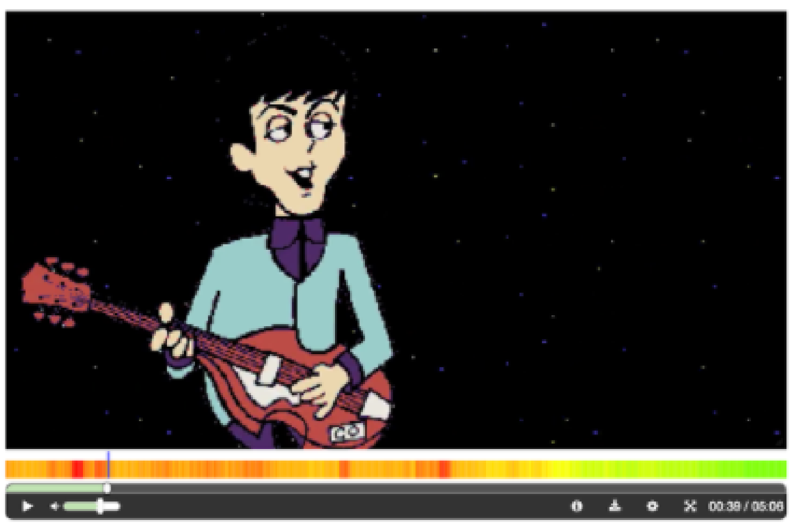
- [ ] Did not watch the video
- [ ] Watched the full video again
- [ ] Tried to remember the place in the video that was related to the question, and navigated to it.
- [ ] Randomly selected a place to watch, and then randomly selected another one if it wasn't relevant.
- [ ] Used the heatmap to help with navigation
- [ ] Other: ______________

Final survey after the study

Just a reminder, the following is a screenshot of VC version without heatmap



Just a reminder, the following is is a screenshot of VC version with heatmap



How useful was the version WITHOUT the heatmap for learning from the lectures? *

	1	2	3	4	5	6	7	
Not useful	○	○	○	○	○	○	○	Very useful

How useful was the version WITH the heatmap for learning from the lectures? *

	1	2	3	4	5	6	7	
Not useful	○	○	○	○	○	○	○	Very useful

How much did you enjoy the version WITHOUT the heatmap? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	A lot

How much did you enjoy the version WITH the heatmap? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	A lot

How frustrating was the version WITHOUT the heatmap? *

	1	2	3	4	5	6	7	
Not frustrating	○	○	○	○	○	○	○	Very frustrating

How frustrating was the version WITH the heatmap for learning from the lectures? *

	1	2	3	4	5	6	7	
Not frustrating	◯	◯	◯	◯	◯	◯	◯	Very frustrating

When using the version WITHOUT the heatmap, how easily were you able to find the information needed to answer the quiz questions? *

	1	2	3	4	5	6	7	
Not easily	◯	◯	◯	◯	◯	◯	◯	Very easily

When using the version WITH the heatmap, how easily were you able to find the information needed to answer the quiz questions? *

	1	2	3	4	5	6	7	
Not easily	◯	◯	◯	◯	◯	◯	◯	Very easily

Which version of Video Collaboratory did you prefer to use for learning from the lectures? *

◯ The version WITHOUT heatmap

◯ The version WITH heatmap

Which version of Video Collaboratory did you prefer to use for completing the quiz? *

◯ The version WITHOUT heatmap

◯ The version WITH heatmap

How likely would you be to use the version WITHOUT the heatmap in your classes/work? *

	1	2	3	4	5	6	7	
Not likely	○	○	○	○	○	○	○	Very likely

How likely would you be to use the version WITH the heatmap in your classes/work? *

	1	2	3	4	5	6	7	
Not likely	○	○	○	○	○	○	○	Very likely

* What do you like best about the Video Collaboratory version WITH heatmap? (Joy points)

Your answer

* What do you like least about the Video Collaboratory version WITH heatmap? (Pain points)

Your answer

* What would you like to improve in the Video Collaboratory version WITH heatmap?

Your answer

* Do you have any other comments or suggestions?

Your answer