ANALYSIS OF USER EXPERIENCE AND BEHAVIOR IN WIRELESS STREAMING VIDEO VIEWING

By

HUA DENG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Janne Lindqvist

And approved by

New Brunswick, New Jersey

October 2020

ABSTRACT OF THE DISSERTATION

Analysis of User Experience and Behavior in Wireless Streaming Video Viewing

By HUA DENG

Dissertation Director:

Janne Lindqvist

Online streaming is one of the most popular services available over the Internet. Today video is increasingly consumed over wireless networks with their far higher packet loss rate compared with wired networks. It generates video impairments that degrade user experience.

We studied video impairments caused by packet losses with a realistic setting. We found that viewers prefer high-resolution videos with some impairments to smooth low-resolution videos. It disagrees with the HTTP adaptive streaming protocol, which sacrifices resolution for smoothness. Additionally, viewers ignore some short impairments and feel that block-artifacts impairment occurring after a freeze is acceptable. We are the first to reveal that impairment occurrence order and inter-impairment interval length influence user experience differently. These findings show the feasibility of improving user experience by reordering and balancing impairment occurrences.

In addition, we conducted experiments on observing users streaming video viewing behaviors under packet loss wireless networks. We have observed nine types of behaviors including system level behaviors and video player level behaviors. We noticed that low video quality and perceivable video impairments are key factors to

motivate users take actions and users usually choose behaviors they believe can improve viewing experience. Also, the sequence of behaviors user has taken follows some particular orders. According to these observations, we propose a novel user video watching behavior prediction model that achieves 94.7% accuracy. Moreover, we created cognitive models that explain how video quality variations and human memories play roles in users' behavior decision making. Findings show the feasibility of improving user experience by reordering and balancing impairment occurrences.

Our study results and the user behavior prediction model provide promising ideas and tools for enhancing user experience under crowded networks where video impairments are unavoidable and optimizing network resource and user management via human engineering.

ACKNOWLEDGEMENTS

I want to express my sincere appreciation for my Ph.D. advisor, dissertation committee members, colleagues, and friends. I could not have completed my Ph.D. research without their guidance, advice, and support.

First and foremost, I would like to express my gratitude to my advisor, Prof. Janne Lindqvist. From the first day of my PhD research, Janne guided me on how to form and approach a research problem, write a paper, and present work to the general public. My knowledge and research skills has developed extraordinarily with his guidance through my PhD study. I am grateful for his dedication while supervising my research and to his support for my work and decisions.

Meanwhile, I am fortunate to have collaborated with many talented colleagues. Christos Mitropoulos is one of the most reliable people and collaborators I have met. His insights on research questions have helped our projects proceed smoothly and efficiently. Thanks for help to: Xianyi Gao, Can Liu, Meghan McLean, Keyi Li, Shridatt Sugrim and Gradeigh D. Clark, who provided feedback and suggestions for the difficulties during my research. More thanks to Fengpeng Yuan, Yulong Yang for providing me suggestions and help during my early years of research.

Furthermore, I want to sincerely thank Prof. Wade Trappe, Prof. Richard Howard, Prof. Saman Aliari Zonouz, Prof. Athina Petropulu, and Prof. Antti

Ylä-Jääski for serving as the committee members for my Ph.D. qualification exam, dissertation proposal, and dissertation defense. They have provided me with valuable suggestions for improving my research and dissertation. I would also like to extend a special thanks to Prof. Richard Howard for providing insightful comments and bringing new ideas to my research experimental design.

Moreover, I want to express my appreciation to my parents and friends. I could never accomplish this work without their encouragement and support. They have made my Ph.D study colorful and productive. The experiences these years are precious memories for me!

This material is based upon work supported by the National Science Foundation under Grant Number 1546689.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Online streaming media service delivers videos and audio to end-users over computer networks and is one of the most popular network services. Nowadays, people are more likely to stream media on mobile devices through wireless networks, and, as a result, mobile video traffic accounted for 59% of total mobile data traffic in 2017 and it will reach to 79% by 2022 (Index, 2019). Meanwhile, packet loss is a major issue in wireless networks. About 20% to 25% wireless connections suffer from 5% to 10% packet loss rates (Falaki, Lymberopoulos, Mahajan, Kandula, & Estrin, 2010; Baltrunas, Elmokashfi, & Kvalbein, 2014). Specifically, large packet loss is usually observed in crowded situations including conferences and sports events (Chen, Jin, Suh, Wang, & Wei, 2012; Shafiq et al., 2013). The lost packets in data transmission generate video impairments and degrade video quality as well as the quality of viewing experience (Boyce & Gaglianello, 1998; Rui, Li, & Qiu, 2006; Reibman & Poole, 2007; Boulos, Parrein, Le Callet, & Hands, 2009).

Many works have studied the quality of experience of watching streaming videos and only considered video clips shorter than 15 seconds with single impairment (Venkataraman, Chatterjee, & Chattopadhyay, 2009; Moorthy, Choi, Bovik, & de

Veciana, 2012; Moorthy, Seshadrinathan, Soundararajan, & Bovik, 2010; Paudyal, Battisti, & Carli, 2016; Duanmu, Ma, & Wang, 2017). In contrast, people usually watch videos with the length of several minutes (Che, Ip, & Lin, 2015; Ooyala, 2018). Human short-term memory determines that impairments occurring before could have effect on people's evaluation on current video viewing experience (Atkinson & Shiffrin, 1968; Pinson & Wolf, 2003). We studied videos that people watch during their daily lives and streamed them under networks with packet losses that generated realistic video impairments. We observed the existence of series of impairments in these streaming videos and quantify the effect of sequential, as opposed to individual, impairments on the video watching experience and the experience changes during different impairments. We discussed that reordering the occurrences of different impairments and balancing the existence of long and short impairments can improve the subjective experience without increasing bandwidth use and demonstrated feasible approaches to enhance video viewing experience for users under resource-constrained networks in which completely removing video impairments for every individual user is impractical.

In real world streaming video viewing, distorted videos bring annoyance to video viewing and induce viewers take actions as responses and viewer behaviors are showing their attitudes towards video qualities. Previous works focused on user engagements and studied how engagement parameters, including video abandonment rate, video playing time, relate to different backbone network conditions, for instance, video buffering time and network flows (Dobrian et al., 2013; Shafiq et al., 2014; Krishnan & Sitaraman, 2013). These works were motivated from online video providers and gave suggestions on improving their content delivery. These user engagement related metrics are unable to capture the explicit factors that trigger users' behaviors while watching videos. Meanwhile, local network issues can also introduce video impairments and affect viewing experience. In wireless local area networks (WLAN), the access point is a networking device allowing end users to connect to a wired network. If the number of connecting requests from users is beyond the access point's handling ability, packet losses exist during data transmission and users are unable to receive stable network connections. Users are able to take actions under these conditions to improve the video viewing experience, for instance, switching to another available wireless network, changing video resolutions and refreshing the video player. In this project, we also studied user behaviors from viewers' perspectives and specify the direct influence to case different behaviors when they are watching streaming videos under packet loss wireless networks. We analyzed the relations between video impairment occurrences and people's video viewing behaviors and discussed the roles of human cognition and memory behind viewers' decision and behaviors. Additionally, we novelly proposed high accurate user behavior prediction models that provide clues of users' feelings about perceived service experiences and potential influences on the entire network resource consumption. Our study provides encouraging ideas to improve user video watching experience on mobile devices with limited network resources and opens the door to design and practice human engineering based management protocols to optimize wireless network management in bandwidth resource-constrained environments.

1.2 ORGANIZATION

Chapter 2 will describe research backgrounds and related work on video delivery over packet loss networks, video quality assessment, behavior incentives and wireless network usage patterns. Chapter 3 will introduce the study on the influence of sequences of different video impairments on people's video viewing experience. Chapter 4 will present the study on people's video viewing behaviors under packet loss networks. Chapter 5 are discussions of the presented two studies. Chapter 6 will explain the limitations of these two studies. Chapter 7 is the conclusion of the two studies.

1.3 CONTRIBUTIONS

Our current work presents the following contributions

1. We create a machine-learning based model to classify types of user behaviors and apply our model to conduct behavior prediction which achieves 94.7% accuracy. In addition, we analyze the importance of features to categorize behavior types and establish a hierarchical model that describes how human cognition works behind users' decision making and behaviors. Users behaviors not only indicate their video watching experience, but also make potential impacts on entire network resource consumption and allocation. Good user behavior prediction provides network controllers support and tools to compose, and employ network resource management protocols via human engineering.

2. We show that people prefer watching videos with block-artifacts over videos that are completely stalled by freezes. Freezes and block-artifacts are two main

types of video impairments existing in real world video streaming services under packet loss wireless networks. Block-artifacts that occur after video freezes show a strong increase in the watching experience rating. Participants felt that the video watching experience has improved during these block-artifacts impairments. Additionally, we also found that participants believed short impairments affected their watching experience differently and short impairments that occurred after a long (9.5 seconds) or a very short (0.2 seconds) impairment-free period did not have much influence on watching experience. This observation demonstrates that changing impairment occurring time can improve watching experience even if the total impairment number remains unchanged.

3. We discover that people choose high-resolution videos with occasional impairments over smooth low-resolution videos without any impairments. Some high-resolution videos streamed under higher packet loss networks could provide the same or even better watching experience compared with low resolution videos streamed under lower packet loss networks. In contrast to the current HTTP adaptive streaming protocols, choosing video smoothness over resolution under limited bandwidth conditions is not always an ideal decision from viewers' perspectives. Providing videos with good trade-off between resolution and impairment occurrence can maximize watching experience under networks with predetermined resources.

4. We notice that impairment type, length and occurring order influence people's video watching experience differently and show evidence that enhancing video

watching experience without additional network resource usage is practical. People's differing evaluations of freezes, block-artifacts and short impairments show that properly managing the bandwidth and controlling the impairment occurrences can improve video watching experience as well as the efficiency of bandwidth utility. This idea is promising in heavily utilized wireless networks where bandwidth is scarce and video impairments are unavoidable.

5. We demonstrate that people take actions because they fell video quality is low and choose types of behaviors they believe can improve the video quality. Low video quality hurts video watching experience and users usually take actions to change the current conditions. Behaviors users frequently perform include changing wireless networks, changing video resolutions, and refreshing the video player and these behaviors produce direct effect on altering video quality and watching experience. Moreover, video quality improvement can change people's decisions on performing some behaviors. Changing networks involves two consecutive actions which are showing the list of available networks and followed by connecting to a new network. We observed that users may give up connecting to another network after displaying the network list when the video recovers from an impairment. This finding tells us that video quality is the key factor affecting user behaviors and human's decision making is a dynamic process along with video quality variations.

6. We exhibit that the time users spent on deciding to take actions is depending on their previous video watching experience. More impairments users saw recently,

more quickly they took actions during the current impairment. The decision time decreases exponentially with the number of previous video impairment increment. Furthermore, we derive a mathematical expression based on ACT-R model to demonstrate how human memory plays an role in evaluating video watching experience and performing behaviors.

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 Overview of Chapter

In this chapter, we will discuss the relevant background of gesture-based authentication methods, password guessability and attacks against gesture authentications in previous work.

2.1.1 Video Streaming over Packet Loss Networks

Video streaming is sensitive to network packet losses. The packet loss causes decoding failure at the receiver end and video impairments occur when the video is playing. These impairments influence the people's video viewing experience. Korhonen (Korhonen, 2018) studied the visibility of packet loss artifacts appearing in video sequences and proposed models for combining objective features to assess the noticeability of the artifacts. Meanwhile, Frnda et al. (Frnda, Voznak, & Sevcik, 2016) used video objective methods to evaluate the quality of video delivery in many different packet loss scenarios. Liang et al. (Liang, Apostolopoulos, & Girod, 2008) studied whether the packet loss pattern is important for precisely evaluating the expected mean-squared error distortion for compressed videos and concluded that burst loss produces a larger video distortion. Also, Joskowicz and Sotelo (Joskowicz & Sotelo, 2014) created a model to predict perceived video quality for the broadcast digital television viewers under individual transport stream packet losses circumstances. Several other studies focused on wireless video transmission. Nasralla et al. (Nasralla, Hewage, & Martini, 2014) performed quality evaluation of the received 3D video sequences over different packet loss LTE networks. Chen and Wu (Chen & Wu, 2012) derived formulas for predicting transmission distortion and provided suggestions on video encoding and transmission scheme designing and He et al. (He & Xiong, 2006) proposed a control system approach to transmission distortion modeling for wireless networks. Different from these works, we generated the distorted videos by streaming under real packet loss networks and provide cognitive approaches to assess the influences of both individual and sequences of impairments on video viewing experience.

2.1.2 User Subjective Assessment on Video Quality

Mean opinion score (MOS) is a subjective quality rating approach recommended by International Telecommunication Union (ITU) which provides 5-grade or 7-grade absolute category rating method to evaluate audio or video qualities (P.910, 2008). Paudyal et al. (Paudyal et al., 2016) used MOS method to study the impact of video content and network delay, jitter, packet loss rate and bandwidth on video viewing experience. Staelens et al. (Staelens et al., 2014) used continued quality evaluation methods to study the influence of video stalls on the video subjective quality. Liu et al. (Liu, Dey, Ulupinar, Luby, & Mao, 2015) studied the effects of initial delay, stall and bit rate on user perceived video quality. De Simone et al. (De Simone et al., 2009) conducted subjective assessment and created a database for H.264/AVC videos delivered over noisy channels and Trestian et al. (Trestian, Vien, Nguyen, &

Gemikonakli, 2015) found that users' video perception quality varies for different video contents and large amount of energy can be saved on wireless devices without sacrificing users' quality of experience. However, these subjective evaluation based studies used short videos less than 15 seconds long or clips with single impairment. The videos people usually watch have length of at least several minutes (Che et al., 2015; Ooyala, 2018) and the occurrences of multiple different impairments are possible when the video is streaming under packet loss networks. Moreover, Garcia et al. (Garcia et al., 2014) did a review on subjective study on HTTP adaptive streaming quality of experience and they pointed out that we still do not know the influences of combined degradation and how people deal with varying quality and long sequences. Our work addresses on these issues and studied how different video impairments affect people's video viewing experience.

2.2 Frustration Motivates Behaviors

Users' feeling towards network quality is important to understand incentives for them to make decision when accessing networks. Chen et al.(Chen, Huang, & Lei, 2006) studied the effect of network quality on real-time, interactive, online game play. They built a regression model based on the three network QoS factors to predict player risks which estimated the level of player intolerance to poor network conditions and finally compared the observed data with their model predictions. Also, Joumblatt et al. (Joumblatt, Teixeira, Chandrashekar, & Taft, 2011) introduced an end-user data collection tool, named HostView, which collects network, application and machine level data as well as gathers feedback directly from users on network performance.

They provided a feasible two-way user feedback mechanism and suggestions on end-host tracing tool design.

Impairments hurts people's video viewing experience and brings viewers dissatisfaction. Stauss et al. (Stauss, Schmidt, & Schoeler, 2005-07) pointed out that frustration is a form of strong dissatisfaction and it occurs individuals do not obtain expected goals. Similarly, Feild et al. (Feild, Allan, & Jones, 2010) shown that users may become frustrated when they have trouble finding information using web search engine. Moreover, frustration can be a behavior motivator and change subsequent behaviors. Amsel et al. (Amsel & Roussel, 1952) described frustration as a motivator for behavior and a causal factor influencing future behavior and Aula et al. (Aula, Khan, & Guan, 2010) found that people use more advanced queries when they have difficulty in finding information in a online search task. In addition, Campion et al. (Campion & Lord, 1982) mentioned that any discrepancy between goals and performance creates a corrective motivation. These works provide evidence that bad video viewing experience gives people motivations to behave and try to improve the current experience.

User engagement shows people's viewing behaviors for online video services. Dobrian et al.(Dobrian et al., 2013) studied the impact of video buffering time on viewer's length of viewing time for long and short videos and Shafiq et al. (Shafiq et al., 2014) focused on the video abandonment rate caused by network flows and transport level factors. Chen et al. (Chen, Zhou, & Chiu, 2013) found the relation between video seeking and video freezes and Krishnan(Krishnan & Sitaraman, 2013) infered causal relation between viewers' video playing time and rebuffering ratio. The works above are based on dataset of commercial online video providers and give advice to providers on content delivery. In addition, Mok et al.(Mok, Chan, Luo, & Chang, 2011) shown that pausing and reducing the screen size are two types of user-viewing activities triggered by video impairments caused by network packet losses. In our study, we consider all types of behaviors viewers can perform at client side in real-world scenarios that include video player level and system level activities. We analyze how video impairments trigger viewer behaviors as well as how viewer behaviors affect network management.

2.3 Wireless Network Usage Patterns

The wireless network usage patterns show people's habit of utilizing wireless network resources in their daily lives. Balachandran et al. (Balachandran, Voelker, Bahl, & Rangan, 2002) analyzed user behavior and network performance in public WLAN using data collected from four wireless access points (APs) at an ACM conference. They found that short-time sessions including web browsing and SSH connections were dominant and user movements were observed at the beginning and end of conference sessions. Authors also pointed out that network load and performance was directly correlated with conference schedules. Meanwhile, Afanasyev et al.(Afanasyev, Chen, Voelker, & Snoeren, 2008) studied the usage of the Google WiFi network deployed in Mountain View, CA. They observed that the aggregate usage of the Google WiFi network is composed of three distinct user populations, characterized by distinct traffic, mobility, and usage patterns. Modem users were static and always connected, and placed the highest demand on the network while hotspot users were concentrated in commercial and public areas, and had moderate mobility. Smartphone users were numerous and place very low demand on the network. Similarly, Fukuda et al.(Fukuda, Asai, & Nagami, 2015) studied the evolution of smartphone usage in the Greater Tokyo area from 2013 to 2015. They noticed that smartphone users selected appropriate network interfaces taking into account the deployment of emerging technologies, their bandwidth demand, and their economic constraints. In details, more users connected to WiFi networks and use bandwith-consuming services such as video streaming. Moreover, Wei et al.(Wei, Valler, Madhyastha, Neamtiu, & Faloutsos, 2015) presented Brofile, a device-centric approach for grouping devices and used it to study handheld device user's individual behavior. Their results shown that small fraction of users consumed most of the network bandwidth and traffics varied across devices significantly and were very bursty in time. Additionally, users patterns of appearance followed weekly and daily patterns. Our work is focusing on users' streaming video viewing behaviors in indoor WiFi settings and studying their influences on network performances.

CHAPTER 3

USER SUBJECTIVE RATING ON STREAMING VIDEOS UNDER PACKET LOSS WIRELESS NETWORKS

3.1 Overview of Chapter

We studied how people's viewing experience was influenced when they watch streaming videos under packet loss wireless networks. Additionally, we provided evidence to support the feasibility and proposed suggestions of improving the quality of video viewing experience under resource-constraint networks.

First, we design a system to emulate real-world wireless network communication and stream videos over the system with controlled packet loss configurations. Then, we conducted a user study with 96 participants to view and evaluate these impaired videos. Finally, we analyzed and compared the subjective evaluations provided by participants with video impairments occurrences and summarized conclusions on how wireless packet loss networks affect people's streaming video viewing experience.

3.2 Method

In this section, we first introduce the source videos we choose and the approaches to generate impaired processed videos for our study. Next, we describe the design of user experiments and procedures we take to conduct the experiments.

3.2.1 Source Videos

The source videos were chosen from four of the top eight popular non-music content categories on YouTube (Cheng, Dale, & Liu, 2008; Che et al., 2015). These categories include education, movie, news and sports. We selected one video from each category and downloaded it from YouTube. Copyright law in the United States regards the use of videos for research purposes as fair use (*Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code*, 2016). Each video clip has 3 different resolutions at 720p (1280 $\times$ 720), 480p (854 $\times$ 480) and 360p (640 $\times$ 360) with a frame-rate of 24 fps and a codec of H264-MPEG-4 AVC (ITU-T RECOMMENDATION, 2017). We cut the videos into 120-second segments, which are close to the length of average non-music videos on YouTube (Che et al., 2015). The news video was an interview about introducing a drawing game using machine learning techniques, the education video was a doodle cartoon describing the energy consumption required for walking and running in cold weathers, the movie was a trailer from Guardians for the Galaxy Vol.2, and the sports video was a top ten highlights clip of NBA games. The average bitrates for videos with different categories and resolutions are listed in Table 3.1. These videos are from well-known YouTube channels that have at least two million subscribers and they are representative of video contents that users might experience. In total, we gathered 12 source videos with different video specifications. Meanwhile, a 30-second long Netflix series video clip at a resolution of 720p was chosen as the training video for the experiment.

	News	Education	Movie	Sports
360p	225kbps	320kbps	406kbps	912kbps
480p	306kbps	485kbps	608kbps	1496kbps
720p	522kbps	902kbps	1068kbps	2869kbps

Table 3.1. The Average Bitrates for Videos with Different Categories and Resolutions

3.2.2 Processed Videos

We built a simple server-client network architecture with one server and one client. The source videos were stored in the server, and the client streamed the videos from the server through a configured network connection. We connected the server and client with an Ethernet cable and emulated the wireless data transmission in order to avoid the possible interference from other wireless signals. All settings were identical for different networks, including initial delays and protocols, except network packet loss rates. We used HTTP streaming over TCP and set four different levels of packet loss rates at 0%, 2.5%, 5% and 10%. These values properly cover the wireless network packet loss rates that appear in real-world scenarios (Chen et al., 2012; Shafiq et al., 2013; Falaki et al., 2010; Baltrunas et al., 2014). The data transfer delay was 30 milliseconds for each packet loss rate configuration (Venkataraman & Chatterjee, 2012). We controlled the server's egress traffic and dropped packets randomly at the network adapter using the traffic control (tc) command in the Linux OS to simulate the packet loss during data transmission. The client used VLC, an open-source media player, to stream the videos from the server and recorded the streaming video at the same time (*VideoLAN media player*, 2018). We disabled the video buffer at the client to avoid any possible video impairments caused by buffering schemes under packet loss networks. Meanwhile, WireShark (*WireShark network protocol analyzer*, 2018) was running on the client to monitor the real-time packet loss rate during the video streaming and verify the network configurations. Each of the 12 source videos was streamed through the four different networks separately and 48 processed videos were recorded and prepared for the experiment.

3.2.3 Experimental Design

We conducted a single-stimulus continuous quality evaluation (SSCQE) study with hidden references (BT.500, 2012; Pinson & Wolf, 2003) over a period of five months. We prepared 48 video clips for the experiment in total and put them into 3 groups based on video resolutions. In each group, we used the Latin-square method to arrange the clips to eliminate the possible order effects and formed 96 test sessions. For test sessions with 360p and 480p videos, we placed the corresponding 720p source videos in the sequence at random positions to act as the hidden references and the participant was unaware of the existence and position of reference videos in the test session. Each test session included 8 clips for the 720p group and 10 or 11 clips for the 360p and 480p groups. In addition, a 5-second long countdown clip was shown between consecutive test video clips in every test session and the entire test session length was less than 30 minutes to satisfy ITU requirements (BT.500, 2012).

We chose single stimulus (SS) method over double stimulus (DS) (BT.500, 2012) since SS only requires participants to watch video once and the shorter test sessions are less likely to fatigue the participants. More importantly, a SS design is closer to the real video viewing experience in people's daily activities (BT.500, 2012; P.910, 2008). The continuous scale helped us capture subjective rating changes during video impairments and was more useful for real-time quality evaluation (Pinson & Wolf, 2003).

3.2.4 Apparatus

The experiment interface was a customized HTML5 video player interface using the video.js framework (*Video.js: The Player Framework*, 2018). The interface, shown in Figure 4.1, contained a display screen at the center and a slider bar ranging from 0 to 100 on the bottom. The slider bar was divided into five equally sized regions labeled “Bad”, “Poor”, “Fair”, “Good” and “Excellent”. Participants were able to use keyboard arrows to move the cursor continuously to rate the video while watching it, and the interface sampled this rating every 0.1 seconds at the background and saved the ratings and the corresponding time-stamps as a .csv file.

The experimental device was a common off-the-shelf PC desktop computer. The interface was running on the Google Chrome v61 web browser and the experimental environments, including viewing distances and luminosity levels were set following the ITU standard requirements (P.910, 2008).

3.2.5 Procedure

We recruited participants through flyers, email lists and online advertisements. We required them to be at least 18 years old and not have color vision deficiency. We also asked them to wear glasses or contact lenses if needed. We recruited 96 participants with ages ranging from 18 to 56 (M = 22.37 , SD= 4.53). 53 were men and 43 were women. 73 of them were pursuing an undergraduate degree, 14 of them were

Figure 3.1. The video quality assessment experiment interface. A display screen is at the center and a slider bar ranging from 0 to 100 with the minimum step size as 1 is on the bottom. The slider bar is divided into five equally sized regions labeled “Bad”, “Poor”, “Fair”, “Good” and “Excellent”.

pursuing graduate degrees, and the remaining 9 had a graduate degree. The study was approved by the Institutional Review Board (IRB) of our institution.

Each participant was assigned to one test group with a unique test session. Before the experiment, we introduced the participant to the study’s purpose, the interface functionality and the types of video impairments likely to occur in the experiment. The participant also read and signed a consent form and was encouraged to ask any clarifying questions before consenting to participate in the study. Next, the participant underwent a training session that included three training clips with different levels of video impairments to become accustomed the interface and experimental settings.

All the test videos in the session were playing sequentially and there was a 5-

second countdown video between each two consecutive clips to let the participant get prepared for the next clip. The rating cursor was set to 50 automatically at the beginning of every video clip to avoid any bias caused by the starting rating. After the experiment, participants completed an exit survey with questions to self-evaluate their performance and we used it to check the data validity before analysis. The exit survey is listed in Appendix A

The experiment workflow is illustrated in figure 3.2. The interface consists of two main parts. The video player acts as the output to play video clips. The rating slider bar is an input and participants move the slider while viewing the processed videos to provide their real-time evaluations to the video quality. When the experiment starts, the video player loads corresponding processed videos from local storage and displays them to views according to a predetermined video sequences. The size of video player maintains unchanged and video clips keep playing throughout the entire experiment. Participants were asked to finish two tasks in parallel. One task is viewing the video clip from the player and another one is moving the rating slider bar to assess their instantaneous video viewing experience. They can press the keyboard arrows to move cursor by one at a time or hold the arrows to shift to the rating position reflecting their viewing experience level immediately. The experiment interface collects subjective ratings, video type, elapsed time information every 0.1 seconds and save it in local storage for data analysis.

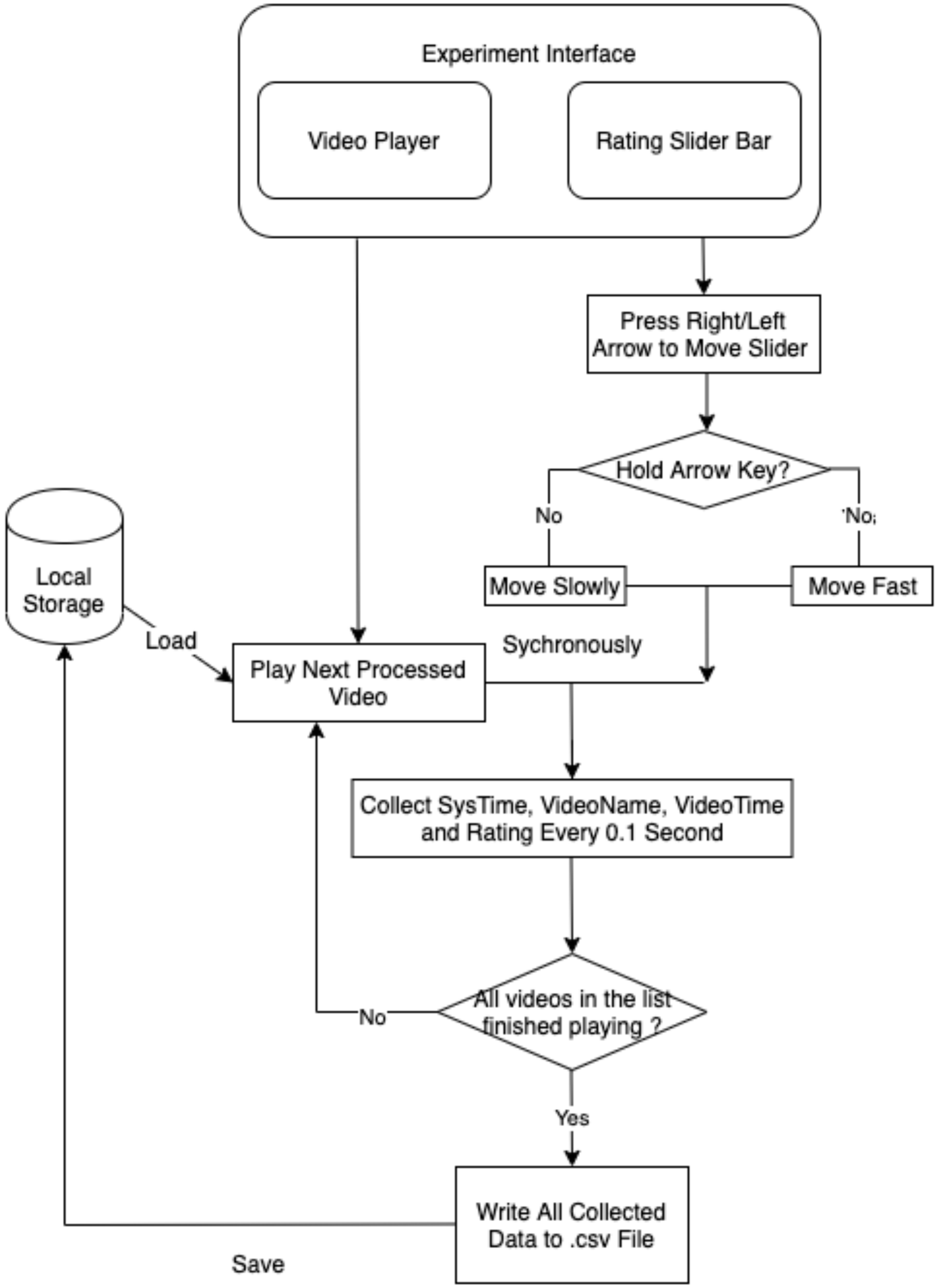


Figure 3.2. Video subjective quality evaluation experiment workflow.

3.3 Video Impairments and Ratings

In this section, we describe and analyze the types of video impairments observed in the processed videos and how video viewing experience rating changed during these impairments.

3.3.1 Video Freezes and Block-artifacts

A video impairment is a period of time during which a video is partially or completely impaired. Viewers are unable to receive the full and clear video content information when the impairment occurs. Video compression and data transmission loss are the two main factors that generate video impairments. In our study, we focused on video impairments caused by network packet losses.

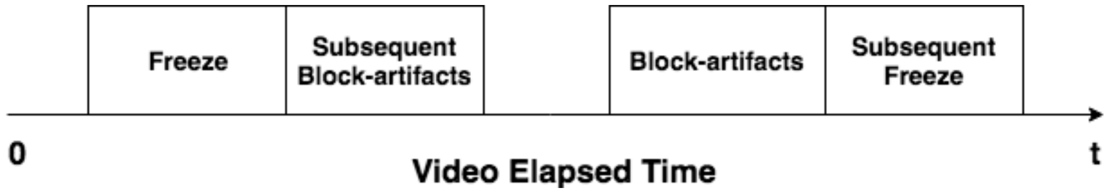
There are two main types of impairments that we observed: *freeze* and *block-artifacts*. Video freezes occur when the decoder does not receive the data for the subsequent frames due to packet loss and the video stops playing, stalling at a frame for a period of time such that video playback is no longer smooth. Another impairment is called block-artifacts. During this impairment, the video keeps playing with some block-artifacts on the frames. This occurs because of the partial frame information losses in the data transmission. Both of these two types of impairments prevent viewers from receiving complete and clear video content and degrade viewers' video viewing experience.

We have 48 videos in total and 12 of them were streaming under the 0% packet loss network. The rest 36 videos were streaming under real networks with different levels of packet loss rates and we were unable to control the lengths and types of impairments in advance, so we collected the impairments' counts and types manually and evaluated their lengths in seconds. In summary, 19 videos have at least one video impairment and 312 video impairments were observed in total. These 312 impairments include 202 video freezes and 110 block-artifacts. Almost half of the impairments have length less or equal to 2 seconds and the impairment count decreases rapidly with the increase of impairment length and 93% (291 out of 312) of the observed impairments are shorter than 7 seconds. The impairment average length is 3.4 seconds with standard deviation of 2.8 seconds. Meanwhile, freezes have longer average length (M = 3.9 seconds) and larger standard deviation (SD = 3.2 seconds) than those of block-artifacts (M= 2.4 seconds, SD = 1.3 seconds).

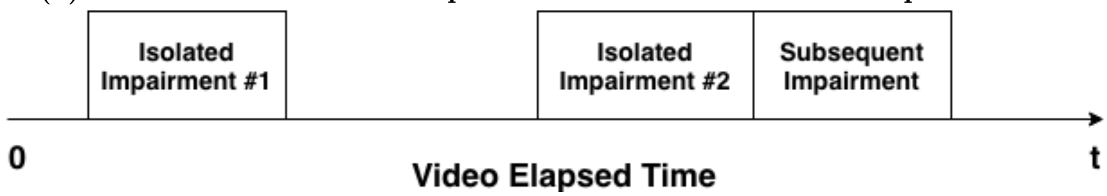
3.3.2 Subsequent and Isolated Impairments

Video impairments can occur differently over time. Some impairments occurred one after another without pause. For instance, the video froze for several seconds and then continued playing with block-artifacts or the video first played with block-artifacts and then froze for a period of time. We call such impairments as subsequent impairments. Figure 3.3a demonstrates the two types of subsequent impairments. One is a block-artifacts impairment occurring right after a freeze and another is a freeze after a block-artifacts impairment. In addition, we also observed isolated impairments and these impairments occurred after an impairment-free period. Figure 3.3b shows two cases of isolated impairments.

Among the total 312 impairments, 183 are isolated impairments and the other 129 are subsequent impairments. Isolated impairments have larger average length



(a) Demonstration of subsequent block-artifacts and subsequent freeze



(b) Demonstration of two cases of isolated impairments

Figure 3.3. Different impairments categorized by occurring times. If an impairment occurs after another impairment with different type immediately, we call it subsequent impairment. On the other hand, if an impairment occurs after an impairment-free period, we name it isolated impairment.

and standard deviation (M = 4.3 seconds, Mdn = 4 seconds, SD = 3.3 seconds) compared with those of subsequent impairments (M = 2.2 seconds, Mdn = 2 seconds, SD = 1.2 seconds). The following sections will discuss how do these impairments affect people's video viewing experience differently.

3.3.3 Video Quality Rating Process

A total of 96 volunteers participated in the experiment and each video was rated by 16 participants. We found that ratings with one sample per second still capture all rating changes during the video. Therefore, we averaged the samples within every one-second period and got a new value to represent the rating for this second and reduce the sample size (Smith et al., 1997). Then, we calculated the rating trace for each video by finding sample-wise averages over the ratings provided by different participants. In continuous quality assessment, participants moved the rating slider while watching the video. The immediate assessment can be a little bit delayed due to the participants reaction and response times (BT.500, 2012). Specifically, when the video quality changed, participants needed some time to react to those changes and alter the quality rating. Therefore, SSCQE scores were usually time-shifted to compensate for this delay in participant response, and we shifted all rating traces backwards for 1 second in our study (Winkler & Campos, 2003). The system collected rating ranges from 0 to 100, and we divided the original rating by 20 and converted it to scales ranging from 0 to 5 that mapped to the video quality regions “Bad”, “Poor”, “Fair”, “Good” and “Excellent”.

3.3.4 Network Packet Loss Rate, Bandwidth and Video Bitrate

We introduce network bandwidth and use it to evaluate available network resources in packet loss networks and simplify the following analysis and discussion. The network bandwidth shows the maximum amount of data the network connection can transfer from one point to another in a given amount of time. The bandwidth of a lossy TCP connection in networks with packet loss rate less than 30% can be approximated by Mathis equation (Mathis, Semke, Mahdavi, & Ott, 1997):

$$Bandwidth = \frac{MSS}{RTT}\frac{C}{\sqrt{p}} \tag{3.1}$$

where MSS is the maximum segment size of a TCP packet, RTT is the round trip time, $C = \sqrt{3/2}$ and p is the probability of packet loss. In our experiment, the TCP packet maximum segment size was 1460 bytes and the round trip time was 60 milliseconds. Therefore, the bandwidth of the three packet loss networks were 1473 kbps (2.5% packet loss rate), 1041 kbps (5% packet loss rate) and 736 kbps (10% packet loss rate). For fixed MSS and RTT values, network bandwidth is negatively correlated with the network packet loss rate. In the following analysis and discussion, we use network bandwidth to describe network conditions and available resources.

The video bitrate is the number of bits that are processed in a unit of time and also approximates the amount of data needed to be transferred over the network for the streaming video service. If the video is streaming under a network with a bandwidth lower than its bitrate, it is possible that some data will be lost during the transmission, resulting in video impairments. The total bandwidth is fixed for a specific network, and the video real-time bitrate varies throughout the video. Grouping videos and networks with shared properties can help us find similarities in impairment type and length and get more general conclusions. Therefore, we placed the 48 processed videos into three groups according to their bitrates and network bandwidths. Table 3.2 shows the criteria we followed in grouping the videos. B_{max} and B_{min} represent the video's maximum and minimum bitrates respectively and BW shows the bandwidth of the network through which the video was streaming. The bitrates of videos in the first group are always less than the corresponding network bandwidth, and we can therefore call them "Low-Bitrate, High-Bandwidth". The videos in the third group always have higher bitrates than the network bandwidth and can be described as "High-Bitrate, Low-Bandwidth". The rest of the videos have their real-time bitrates that intersect with the network bandwidth, and we called them "Medium-Bitrate, Medium-Bandwidth".

Table 3.2. The Video Grouping Criteria for Video Bitrate and Network Bandwidth. B_{max} and B_{min} represent the video's maximum and minimum bitrates respectively and BW shows the network bandwidth.

Group	Criteria
Low-Bitrate, High-Bandwidth	$B_{max} < BW$
Medium-Bitrate, Medium-Bandwidth	$B_{min} < BW < B_{max}$
High-Bitrate, Low-Bandwidth	$B_{min} > BW$

3.4 Results

In this section, we discuss how do our participants change the video quality ratings when they see different video impairments as well as the relations between network packet loss rates and video impairments.

3.4.1 Video Impairments Assessment

Block-Artifacts Impairment After a Freeze is Acceptable

We found that our participants believed videos playing with block-artifacts were much more acceptable than videos that stalled completely due to freezes, and they felt the video viewing experience improved when they saw the video resumes playing with some block-artifacts after a complete video freeze since the ratings increased in a majority of subsequent block-artifacts. Figure 3.4 shows the cumulative distribution functions (CDF) for both subsequent block-artifacts and freezes against the amount of rating change. More than 80% of the block-artifacts that occur after a major freeze show improved experience quality ratings. On the other hand, more than 90% of the freezes that follow a block-artifacts impairment show reduced experience quality ratings and it demonstrates that participants thought that viewing experience kept decreasing in freezes after block-artifacts impairments. The rating increase in subsequent freeze (M = -0.24, SD = 0.28) is smaller than that of subsequent block-artifact (M = 0.25, SD = 0.43) and the difference is statistically significant, U = 354, $p < .001$ (Mann-Whitney U test). It shows that users felt block-artifacts impairments after a freeze is acceptable.

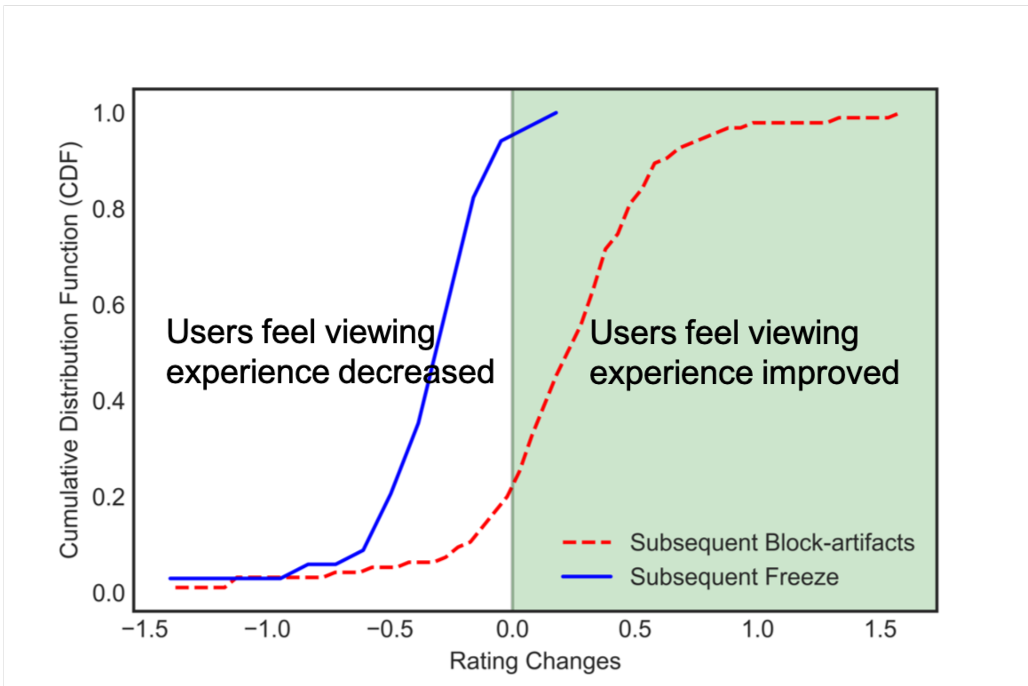


Figure 3.4. The cumulative distribution function (CDF) of subsequent block-artifacts and freezes with different rating change amounts. The positive rating change value shows that users feel video viewing experience improved during the impairment while the negative rating change value means that users' video viewing experience decreases during the impairment. The rating increases in most of the subsequent block-artifacts (> 80%) and drops in most of the subsequent freezes (> 90%). The two-sided Kolmogorov-Smirnov test result, D(34, 95) = 0.77, $p < .001$, shows the difference between two distributions is statistically significant.

Short Impairments Do not Always Degrade Viewing Experience

For impairments that led to decreased ratings , the impairment length is a key factor in determining the decreasing trend. Absolute rating drop amount is not a good metric to evaluate the video viewing experience degradation since a rating change from "Fair" to "Poor" is not equivalent to a drop from "Poor" to "Bad" (Streijl, Winkler, & Hands, 2016). We introduced a new metric called rating drop fraction, which is defined as the ratio between absolute rating drop and the maximum rating of the impairment. We picked all impairments with rating drops and compared the fraction of rating drops against the impairment lengths.

Figure 3.5 shows that impairment rating drop fraction increases when impairment has longer length and it means that participants felt that longer impairments have greater influence on video viewing experience. Meanwhile, short impairments (i.e., length $<$ 3 seconds) have wider range of rating drop fraction values and participants thought these short impairments have different effect on viewing experience. We found that rating drop fractions (M = 0.37, SD = 0.22) of short impairments (i.e., length $<$ 3s) have larger variance than that of long impairments (M = 0.71, SD = 0.15), W(1, 229) = 25.33, p $<$.001 (Levene's test).

We got the similar observation when comparing the minimum ratings within each impairment against its length. Figure 3.6 shows that the minimum rating values drop when impairment length increases and short impairment have a wider range of minimum ratings. Short impairments' (length $<$ 3s) minimum ratings (M = 1.01, SD = 0.62) have larger variance that that of long impairments (M = 0.52, SD = 0.24), W(1, 229) = 31.14, p $<$.001 (Levene's test). These observations tell us that participants have different evaluations on viewing experience during these short impairments.

We took a further look to find out why participants reported different evaluations of short impairments, especially the short impairments with small rating drop frac-

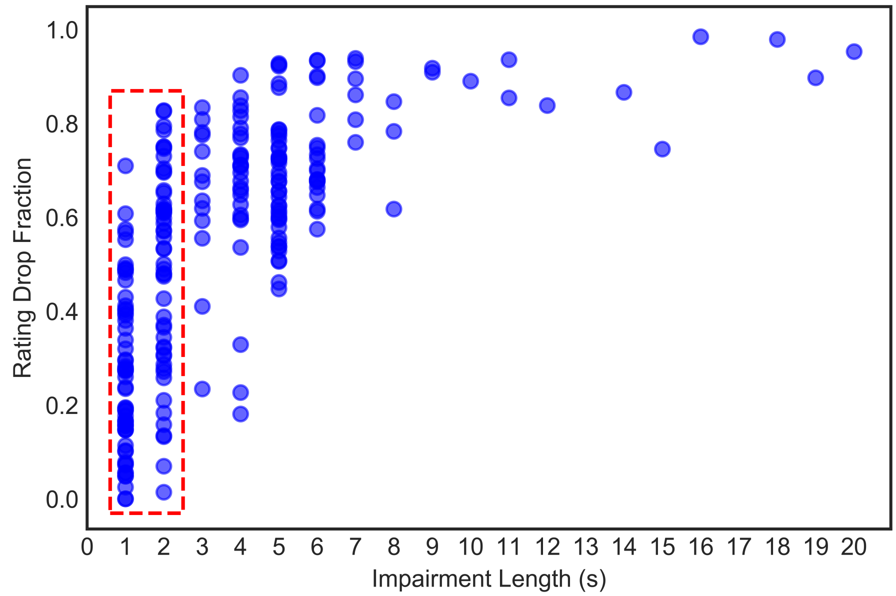


Figure 3.5. The rating drop fraction for impairments with different lengths. Rating drop fraction means the ratio between the absolute rating drop and the maximum rating for the impairment, and it measures the level of video viewing experience degradation caused by the impairment. Each point implies the rating drop fraction of one impairment. The rating drop fraction follows an climbing trend along with the increase of impairment length. However, the impairments shorter than 3 seconds have larger range of rating drop fraction values and it shows that users felt some short impairments affected their video viewing experience while some did not.

tions. We categorized these short impairments into two groups according to their minimum ratings. For impairments with a high minimum rating and a small rating drop fraction, video viewing experience was at a high level when the impairment started, and participants did not think the impairment affected their experience. On the other hand, the impairments with a low minimum rating and a small rating drop fraction told us that the viewing experience was already at a low level, and the impairment did not worsen it further. In both of these two cases, the impairments did not have much influence on participants' evaluation on the real-time video viewing

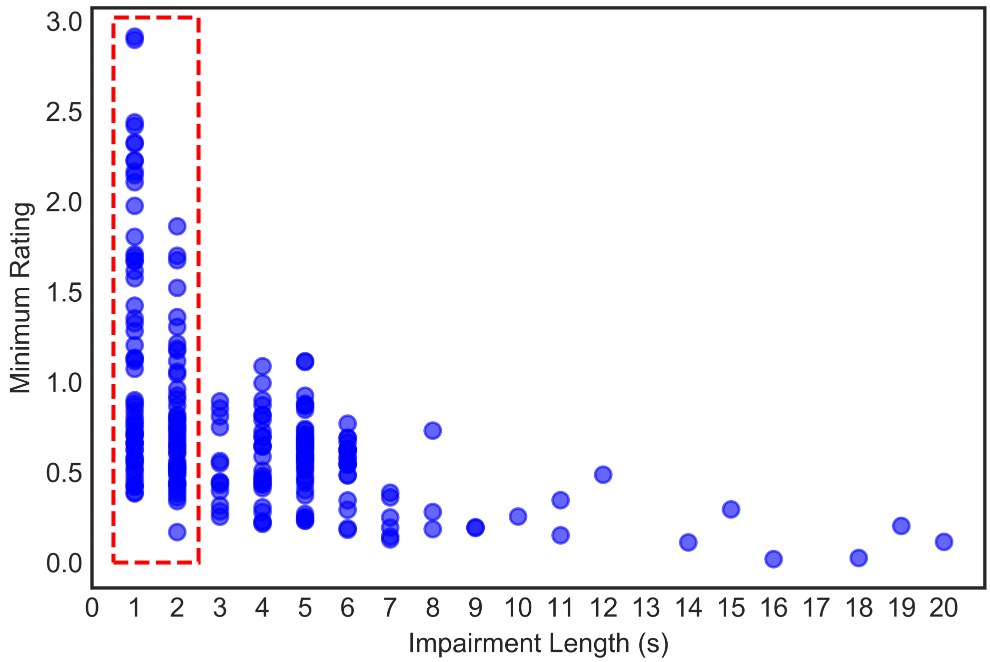


Figure 3.6. The minimum rating for impairments with different lengths. Minimum rating means the lowest rating value that is observed within every impairment. Each point implies the minimum rating of one impairment. The value gradually decreases as impairment length increases. However, impairments shorter than 3 seconds have wider range of minimum rating values and it tells us that users feel some short impairments do not influence the video viewing experience.

experience. We noticed that the average length of impairment-free period before these two groups of impairments are 9.5 seconds and 0.2 seconds, respectively. The rest of the short impairments, in which participants changed the ratings, have an average 5.8-second long prior impairment-free period. Therefore, short impairments occurring after a relatively long or a very short prior impairment-free period do not have much influence on the video viewing experience.

Participants Preferred High Resolution Videos with Occasional Impairments over Low-Resolution Videos without Any Impairments

We analyzed the relation between each video's average rating and its impairment ratio. Impairment ratio is defined as the proportion of total impairment length to the video length. Higher impairment ratio means more impairments were observed in the video. Figure 3.7 demonstrates 720p and 480p videos with a few impairments have higher ratings (M = 3.23, SD = 0.34) than that of 360p impairment-free videos (M = 2.35, SD = 0.30), U = 6, $p < .001$ (Mann-Whitney U test). We conclude that participants were more likely to watch high-resolution videos with occasional impairments than smooth low-resolution videos without any impairments.

For videos with the same resolution, the average rating decreased with an increasing impairment ratio, and the video viewing experience dropped if more or longer impairments were observed. Also, higher resolution videos provided better viewing experience if the videos had no impairment or similar impairment ratios.

3.4.2 Bandwidths and Video Impairments

High-Bitrate Videos Streamed under Low Bandwidth Networks Have More and Longer Impairments

We compared the number and length of impairments in videos of different groups. Figure 3.8 shows that the numbers of both freeze and block-artifacts impairments rise rapidly when the video bitrate increases and network bandwidth decreases. Specifically, the average number of impairments has a large jump from the "medium-bitrate and medium-bandwidth" group to "high-bitrate and low-bandwidth" group. Mean-

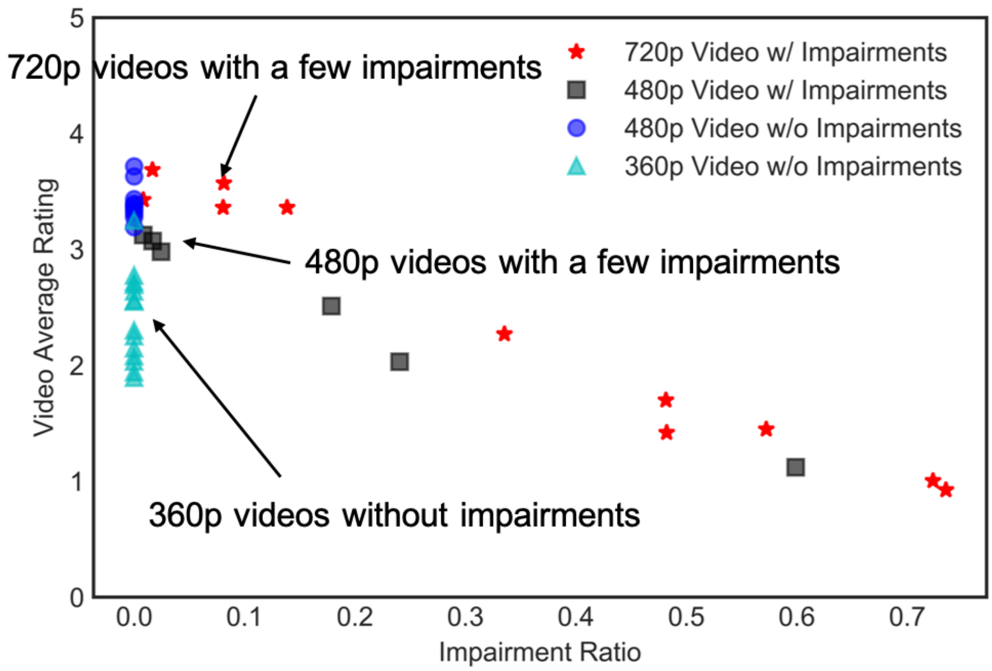


Figure 3.7. The average rating of videos with different video impairment ratios. Video average rating is calculated by averaging all sampled ratings throughout the video, and it measures the video's viewing experience on average. Impairment ratio is the ratio between total impairment length and the video length. The higher the impairment ratio is, the more impairments are observed in the video. Each data point refers to one specific video. 720p and 480p videos with a few impairments (i.e., impairment ratio < 0.2) gave users higher viewing experience than 360p impairment-free videos.

while, Figure 3.9 shows that the lengths of freeze and block-artifacts impairments also rise when high-bitrate videos are streaming under low-bandwidth networks. Numerically, high bitrate videos streamed under low bandwidth networks have larger number ($M = 46$, $SD = 9.83$) and longer length ($M = 3.7$, $SD = 3.07$) impairments and the differences from impairment number and length other videos are statistically significant ($U = 0$, $p < .001$ and $U = 9972$, $p < .01$).

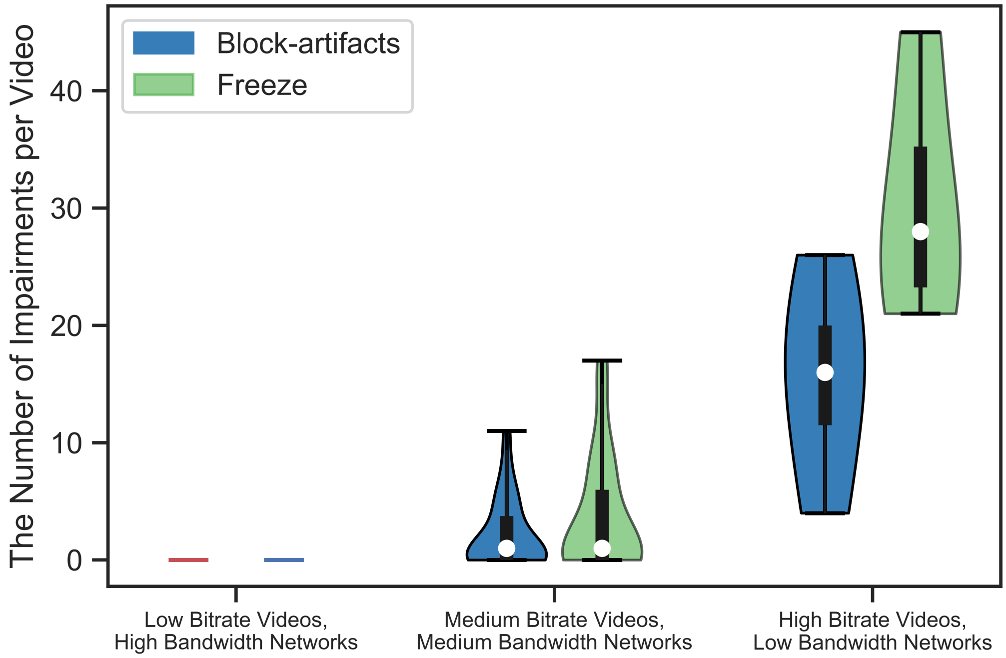


Figure 3.8. The number of impairment in videos of different groups. Violin plots show that no impairment exists in low bitrate videos streamed under high bandwidth networks and the number of impairments increase as video bitrate increases and network bandwidth decreases. Two bars in the violin plots represent the maximum and minimum values. The upper edge of the black box represents the third quartile and the lower edge represents the first quartile. The circle in the box depicts the median.

Most Subsequent Freeze Impairments Occur in High Bitrate Videos Streamed under Low Bandwidth Networks

Figure 3.10 shows that the numbers of subsequent freeze and subsequent block-artifacts impairments per video jump as the video bitrate increases and the network bandwidth decreases. The occurrences of subsequent freeze ($M = 8.0$, $SD = 8.45$) high bitrate videos streamed under low bandwidth networks are much higher than that of other videos, $U = 12$, $p < .001$. In particular, only one subsequent freeze was in videos of the “medium-bitrate and medium-bandwidth” group and subsequent freeze

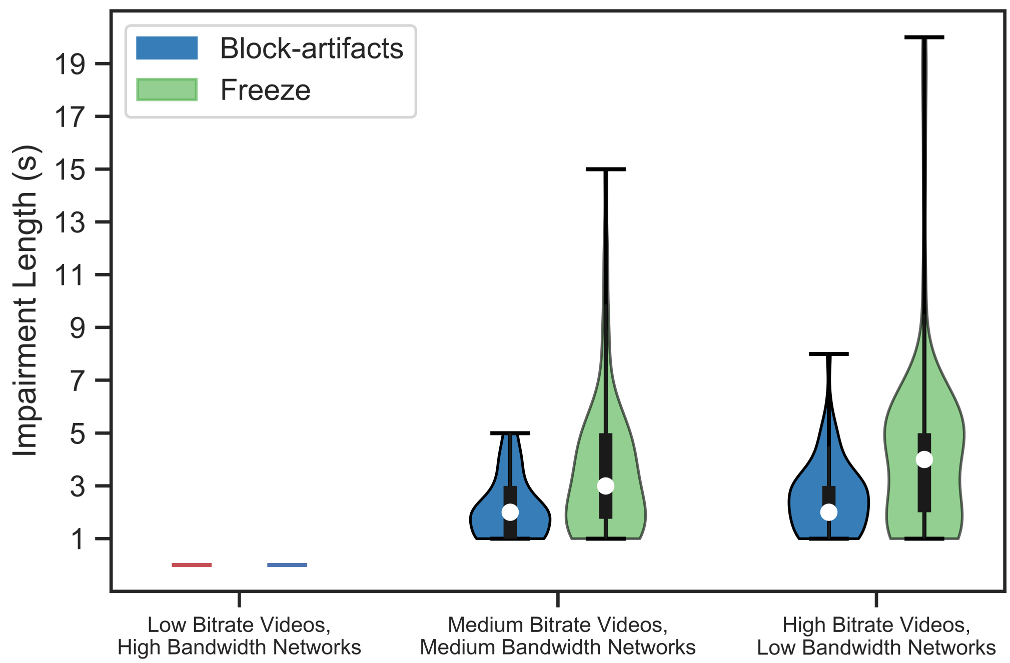


Figure 3.9. The length of impairment in videos of different groups. Violin plots show that no impairment exists in low bitrate videos streamed under high bandwidth networks and the length of impairments increase as video bitrate increases and network bandwidth decreases. Two bars in the violin plots represent the maximum and minimum values. The upper edge of the black box represents the third quartile and the lower edge represents the first quartile. The circle in the box depicts the median.

occurrence upsurges dramatically for videos in the “high-bitrate and low-bandwidth” group.

According to our observations, it is practicable for network controllers to manage the resource allocation based on video bitrates and manipulate the occurrence of different types of impairments to improve people’s video viewing experience when the impairments are unavoidable.

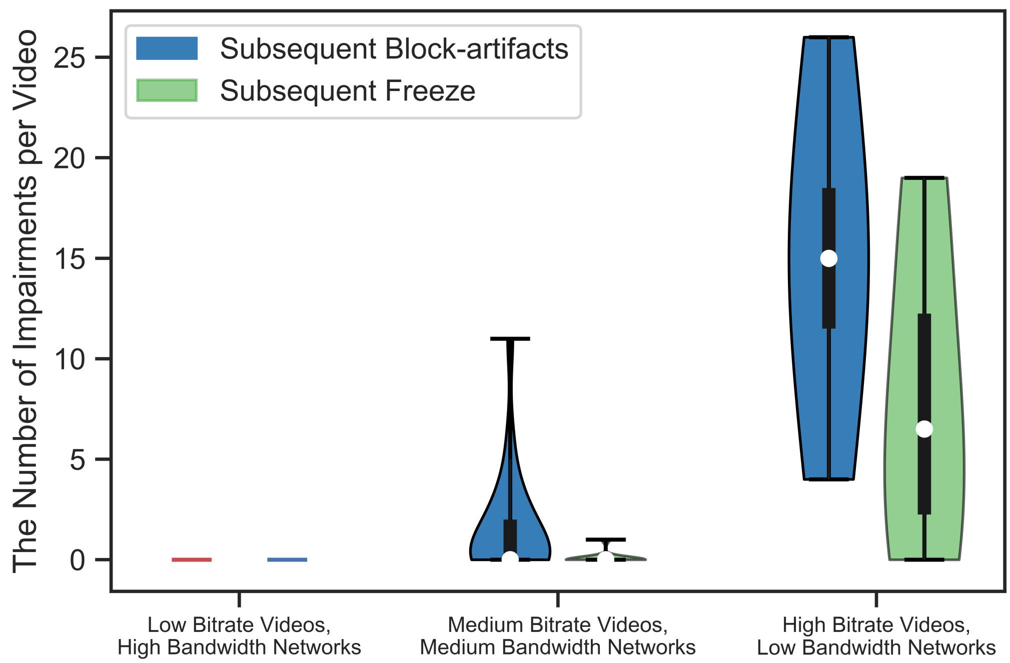


Figure 3.10. The number of subsequent block-artifacts and freeze impairments per video in videos of different groups. The violin plot shows that numbers of both subsequent block-artifacts and freezes increase rapidly from the second to the third group. Particularly, almost all of the subsequent freezes are observed in high-bitrate videos streamed under low-bandwidth networks, indicating that managing network bandwidth according to video bitrates can prevent the occurrence of subsequent freeze and enhancing video viewing experience.

3.5 Summary

We conducted a user experiment to evaluate the influence of video impairments caused by network packet losses on video viewing experience. We discovered that the block-artifacts impairments happening right after freezes are acceptable to viewers and high-resolution videos with occasional impairments are much more preferred than smooth low-resolution videos. Additionally, we found that short impairments with a relatively long or very short prior impairment-free period do not have much influence on decreasing the viewing experience. Our work is the first to study the effect of impairment occurring orders on viewers video viewing experience and demonstrate that it is feasible to improve viewing experience without additional resource usage by changing video impairments types and occurrence orders. In busy networks, the number of users exceeds the network handling capacity and it is impossible for every user to receive service with perfect quality or experience. Our study provides encouraging ideas to improve user video viewing experience with limited network resource and opens the door to looking for solutions to optimize network management in network resource-constrained environments based on user experience.

CHAPTER 4

USER VIEWING BEHAVIORS ON VIDEOS STREAMING UNDER PACKET LOSS NETWORKS

4.1 Overview of Chapter

This study follows the previous project on user subjective rating on streaming videos and analyzes how users behave when they are watching videos impaired by network packet losses. In details, the behaviors include pause video, seeking, change video resolutions, refresh video player, change video volume, make video player fullscreen, make video player normalscreen, click wireless network button to show the list of available networks and change wireless networks. These behaviors are the ones users can do and interact with video players and systems when they are watching streaming videos in daily lives. First, we introduce experimental design and the customized video player we created for the experiment. Next, we demonstrate the results of the experiment and discuss how users behaviors relate to video ratings, video impairments as well as their previous behaviors. Finally, we applied different models to predict users' video viewing behaviors and evaluated their performance.

4.2 Method

4.2.1 Experimental Design

We use the same processed videos from the user subjective rating on streaming video study. Four types of videos which are education, movies, news and sports with three different resolutions at 720p (1280 $\times$ 720), 480p (854 $\times$ 480) and 360p (640 $\times$ 360) were downloaded from YouTube. We streamed these videos under networks with three different packet loss rates at 2.5%, 5% and 10% to generate impaired videos for the experiment.

We conducted an experiment in the well-controlled laboratory environment and we deceived participants by informing them that we are doing an experiment to study how people observe and memorize video contents and details and they needed to answer ten questions related to video contents afterward. Studies show that participants unintentionally change their behaviors if they have formed an intepretation of the experiment's purpose and it introduces experimental artifacts (Rosenthal & Rosnow, 2009). We used deception and conceal the experiment objective to minimize the probability that participants perform differently from their daily lives. In addition, the informed experiment purpose makes participants get involved in the video contents and be attentive to the quality changes. Moreover, we placed a wireless network testbed in the room and kept it running through the entire experiment. We told participants that all videos were streaming from this local server and the server provided three different wireless networks. Videos streaming under these networks have different qualities and participants are able to connect to any of these three networks and switch between different networks during the experiment. However, the real story is that we streamed all videos under different wireless networks in advance and stored them locally. When participants select different networks, the player is picking and playing the corresponding video. This setup keeps video qualities under the same network settings consistent over different participants and using the local server as a part of deception increases the credibility of our story. In addition, participants can also interact with the video player in the experiment that help them have a better video viewing experience. The allowable behaviors are normal actions people can take with the player in their daily streaming video viewing. We will elaborate the types of behaviors later.

4.2.2 Apparatus

We created a customized HTML-5 video player interface using the video.js framework (*Video.js: The Player Framework*, 2018). The interface, shown in Figure 4.1, has a display screen at the center and two buttons at the top corners. Clicking the button on top-left corner refreshes the player. Top-right corner button shows a list of available wireless networks and participants are able to choose a network they want to connect during the experiment. Additionally, the display screen has integrated functions including pause videos, move video time cursor, change video resolution, fullscreen the player and adjust video volume. The experiment workflow is shown in figure 4.2. At the start of experiment, the embedded video player load the corresponding videos from local storage according to initial network condition and video setting information. During the video playing, the experiment interface is monitoring the existence of any of the nine target viewing behaviors by nine different event handlers. Once participants take a specific action, the matching event handler is triggered and the action type, execution time, video setting and network condition information before and after the action is collected by the interface and save to the local storage. At the same time, video and network configurations are updated accordingly and the video player continue playing the video with the updated settings.

The experimental device was a common off-the-shelf PC desktop computer and the interface was running on the Google Chrome v61 web browser. The experimental environments, including viewing distances and luminosity levels was set following the ITU standard requirements (P.910, 2008).

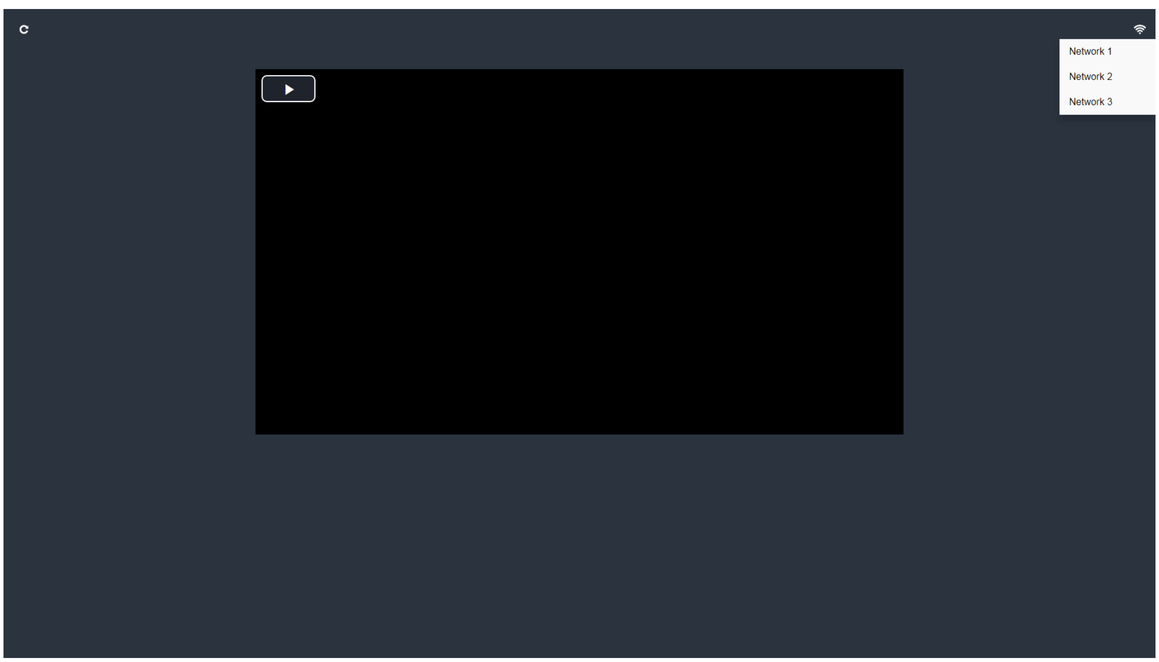


Figure 4.1. User video viewing behavior experiment interface. A display screen is at the center and two buttons are at the top corners. Player refresh button is on the top-left corner and wireless network button is on the top-right. Once the wireless network button is clicked, a list of available networks is shown.

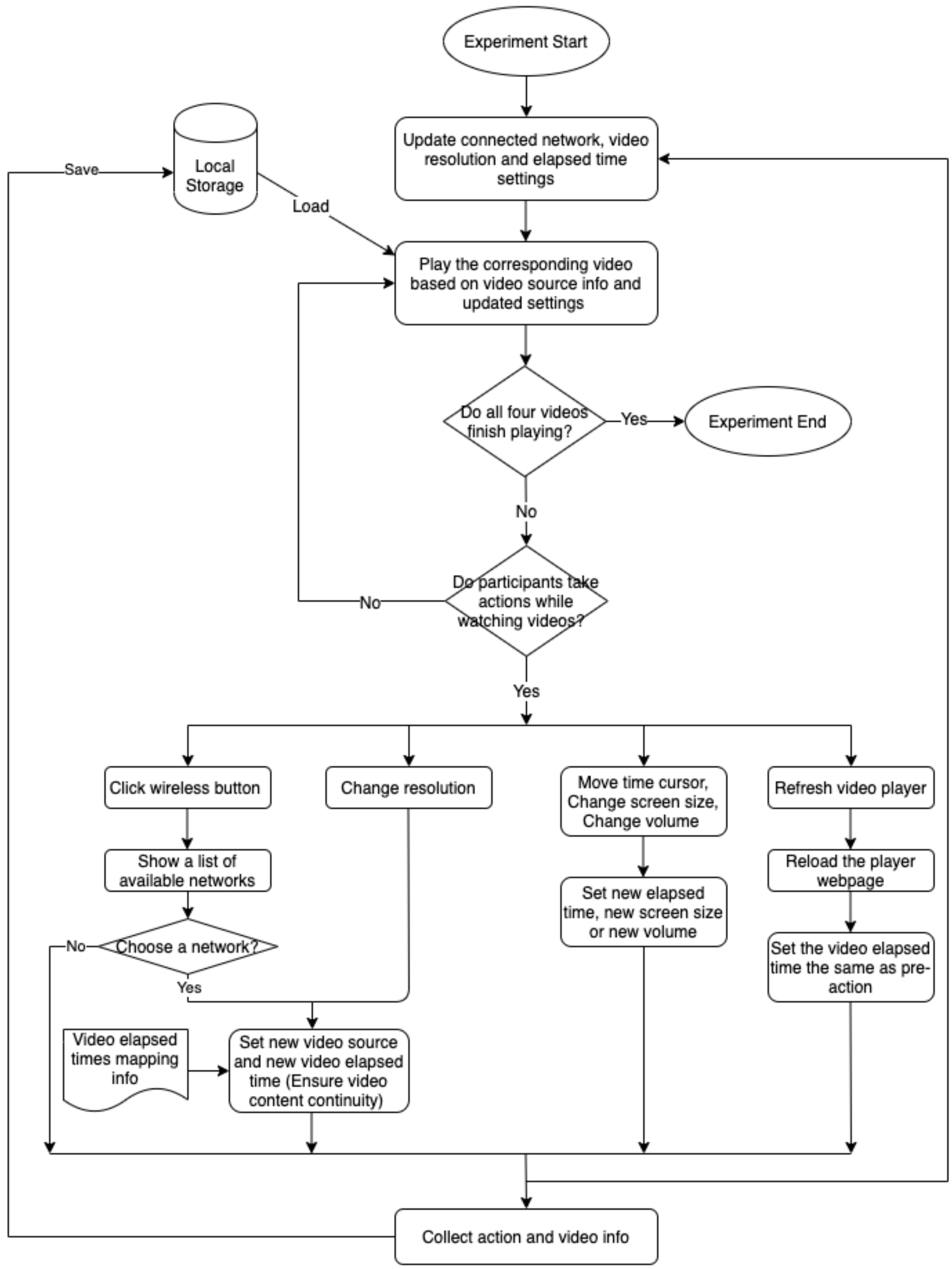


Figure 4.2. User video viewing behavior experiment workflow.

4.2.3 Procedure

Every participant watched all four categories of videos and the video order was shuffled using Latin Square design. Before the experiment, participants read and signed the informed consent form. Next, participants went through a training session with a 30-second long video and adapted the interface functionality and experimental settings. The experiment videos were playing sequentially and there was a 5-second countdown video between each two consecutive clips to let the participant get prepared for the next one. The default starting resolution for every video is 720p and the initial connected network has 10% packet loss rate. After the experiment, participants completed an exist survey with questions to self-evaluate their performance and we used it to check the data validity before analysis. The exit survey is listed in Appendix B

4.2.4 User Video Viewing Behaviors

We collected nine types of user video viewing behaviors in the experiment and we defined these behaviors as:

- **Click Wireless Network Button** (Wireless_button_clicked): This button is on the top-right corner of the experiment interface. It emulates the icon and functionality of real wireless network sign in operation systems. After clicking this button, a list of three available wireless networks is shown with names 'Network1', 'Network2' and 'Network3'. Additionally, the current connected network will have a label 'connected' after its' name. Users need to click this button first before switching to another wireless network.

- **Choose a Wireless Network from the List** (Change_network): Users choose one wireless network from the list and click its name. If the user clicked the current connected network, the system will reconnect it. If the user clicked another network, the system will switch to it and continue streaming the video over this newly connected network.

- **Move Video Time Cursor** (Seeking): Users can move the time cursor forward or backward while watching the videos or click the position they want to move to directly on the video progress bar. The video will continue playing after the cursor move. The video progress bar is on the bottom of the video player.

- **Pause Video** (Pause): While the video is playing, users can click the button on the bottom-left corner of the video player to pause the video.

- **Refresh Video Player** (Refresh_player): This button is on the top-left corner of the experiment interface and it emulates the webpage refresh when people are viewing online streaming videos. After clicking this button, the system will refresh the video player, reconnect the current connected network and continue playing the video.

- **Change Video Resolution** (Change_resolution): This button is on the bottom-right corner of the video player. After clicking this button, a list of three available video resolutions is shown with names '720p', '480p' and '360p'. Users can pick the video resolution by clicking the corresponding label and the player will continue playing the video with the chosen resolution.

- **Make Video Player Full Screen** (Fullscreen): This button is on the bottom-right corner of the video player. When the player is in normal screen mode, users can click this button to enlarge the player to full screen.

- **Make Video Player Normal Screen** (Normalscreen): This is the same button as making the player fullscreen. When the player is in full screen mode, users can click the button to make it back to normal screen.

- **Change Video Volume** (Volumechange): This button is on the bottom-corner of the video player. Users can move the volume bar left or right to decrease or increase video volume.

When participants perform any of these behaviors during the experiment, our system will collect the following information data and save it as a .csv file:

- **System Time**: Unix epoch time in seconds. It shows the system time when the behavior is performed.

- **Video Time**: The video elapsed time in seconds. It shows how many seconds the video has played when the behavior is performed.

- **Action**: The behavior type name.

- **Previous Resolution**: The video resolution before the behavior.

- **Current Resolution**: The video resolution after the behavior.

- **Previous Network**: The connected network before the behavior.

- **Current Network**: The connected network after the behavior.
- **Content**: Video Content Type (i.e., Movie, Education, News or Sports).
- **Volume Percent**: Volume Percent (of the maximum volume) after the behavior (from 0 to 100).
- **Participant ID**: Unique ID for the participant.

The system also collects mouse positions along with system time every second.

4.3 Results

4.3.1 General Behavior Information

We have observed 1496 behaviors in total with 44 experiment participants. Figure 4.8 shows the number of behaviors each participant has performed during the experiment and it ranges from 87 to 10 ($M = 34, SD = 18.3$). Additionally, figure 4.4 shows the number of observed behaviors for each type. Switching connected networks (i.e., Wireless_button_clicked and Changing_network) is the action most frequently chosen by participants in the experiment and other actions which participants believe can change their video viewing experience (i.e., Pause, Change_resolution and Refresh_player) are also performed by participant with moderate frequencies. This information tells us that participants decided to take actions when they had opinions on the video viewing experience and tried to change the current condition.

When we look at the number of behaviors in each video content category, more than half of the behaviors are in sports videos (shown in figure 4.5). Sports videos have

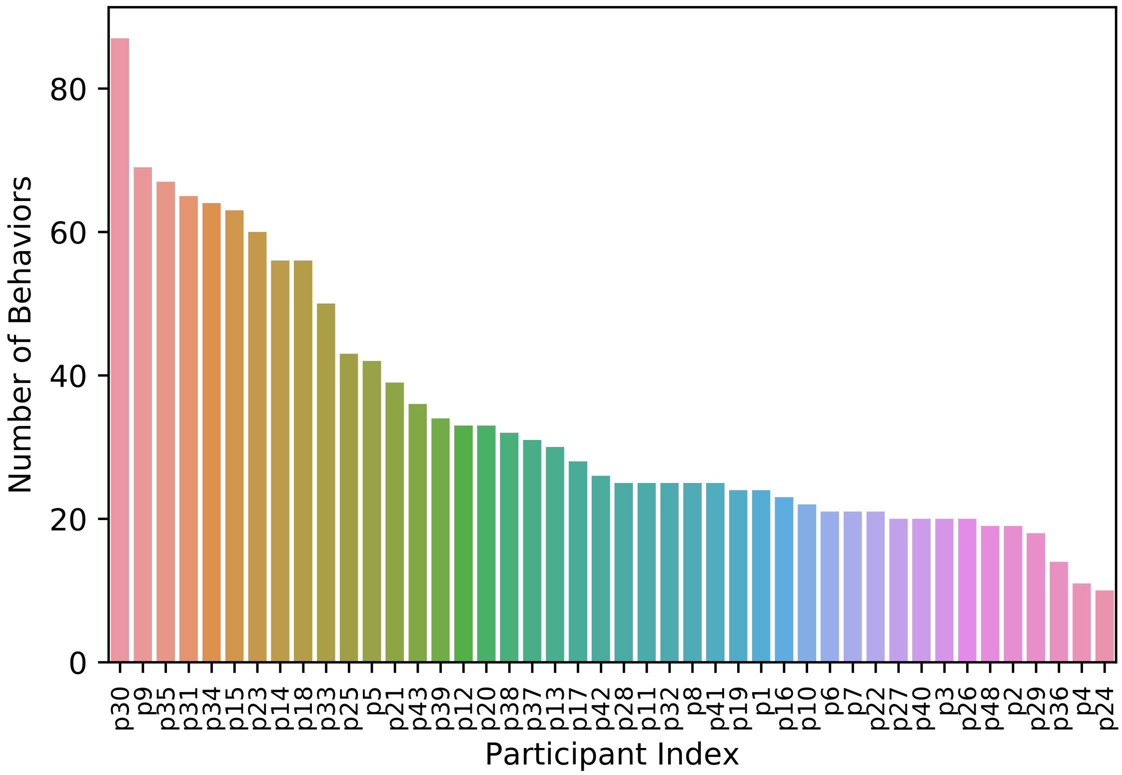


Figure 4.3. Total number of behavior observed for each participant.

higher video bitrate variations and are more sensitive to network packet losses. We notice that sports videos have more video impairments and lower ratio quality ratings. Statistically, the number of behaviors with each video category is positively correlated with the number of impairments ($r(2) = 0.99, p = .003$) and negatively correlated with the average video quality ratings ($r(2) = -0.97, p = .03$). Interestingly, the behavior numbers in different video categories do not show much correlation with the average impairment lengths ($r(2) = -0.13, p = 0.87$). We will have further discussion on this in the following sections.

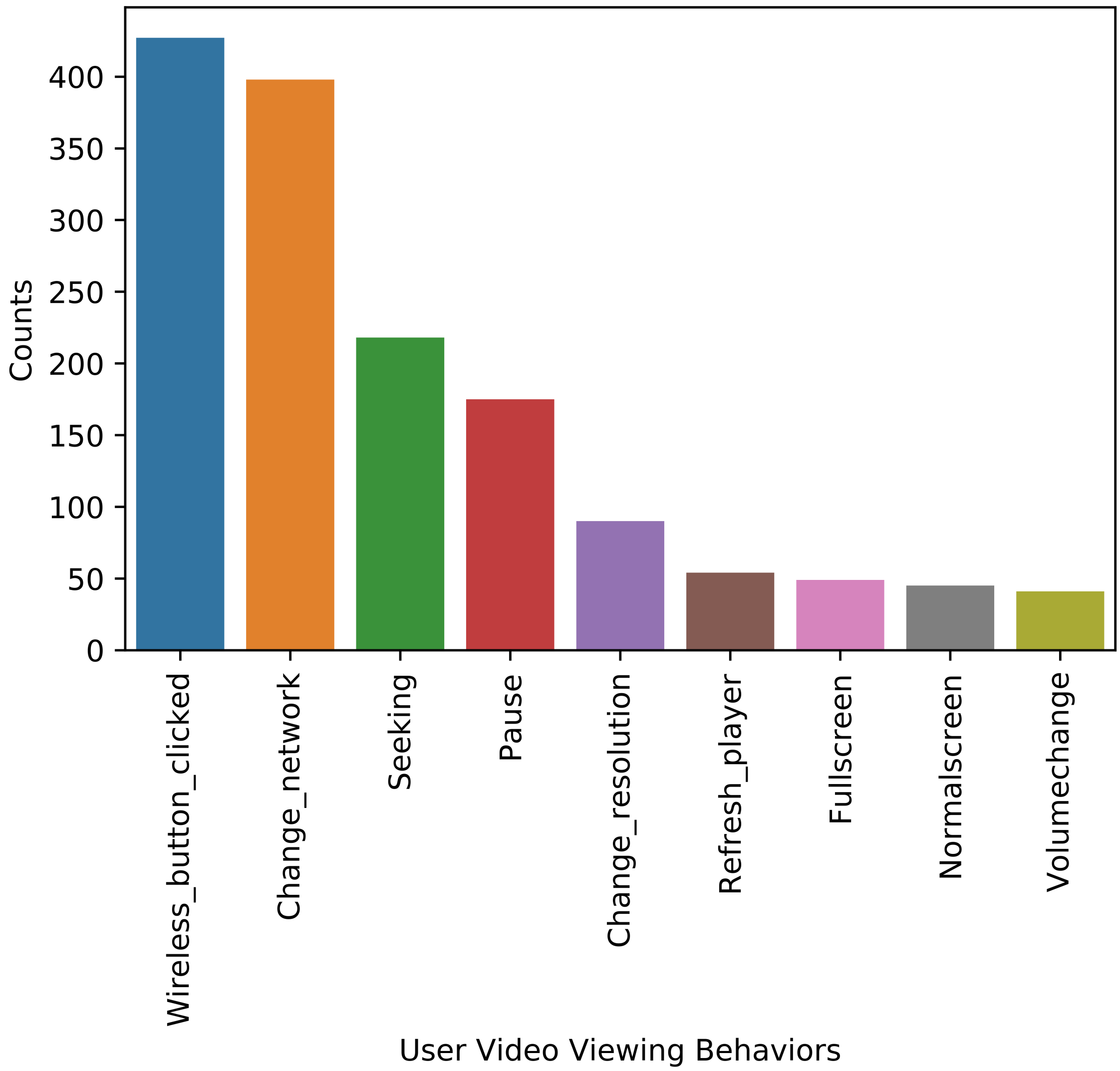


Figure 4.4. Total number of behavior observed for each type.

4.3.2 Video Quality and Impairments in Pre-Behavior Windows

Studies show that human short-term memory shows that what people saw before also have non-trivial influence on their current behaviors (Atkinson & Shiffrin, 1968). Furthermore, Pinson et al. (Pinson & Wolf, 2003) found that the memory length is about nine seconds when people are watching videos. Therefore, we define a *pre-behavior window* and set the window length as nine seconds. We include all video

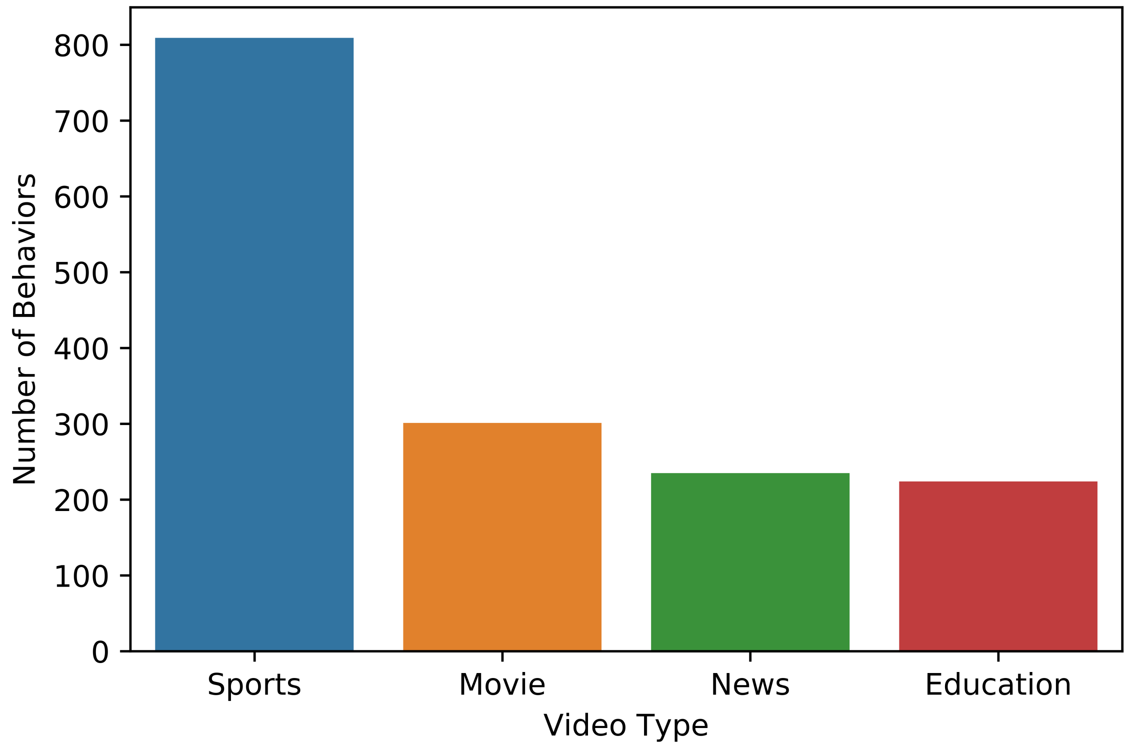


Figure 4.5. Total number of behavior observed for each video content category.

quality rating information and impairment occurrences in the pre-behavior window of each user behavior in the following analysis and discussion.

4.3.3 User Behaviors and Video Subjective Quality Ratings

We use the video subjective quality rating data we collected in the first project as the quality of video viewing experience metric and study how different video viewing behaviors relate to the quality ratings.

Participants Take Actions Because Video Quality is Low and Choose Actions They Believe Can Improve the Video Quality.

We compared the video quality ratings in all pre-behavior windows with the ratings in the rest of videos. Figure 4.6 shows the rating distributions. We notice that the ratings in pre-behaviors windows are much lower ($M = 1.7$, $SD = 1.1$) than the ratings in the rest of videos ($M = 2.7, SD = 0.9$) and the rating difference is statistically significant, $U = 6.8e6, p < .001$ (Mann–Whitney U test). This observation tells us that people took action during the video because they felt the quality is bad and influence their viewing experience.

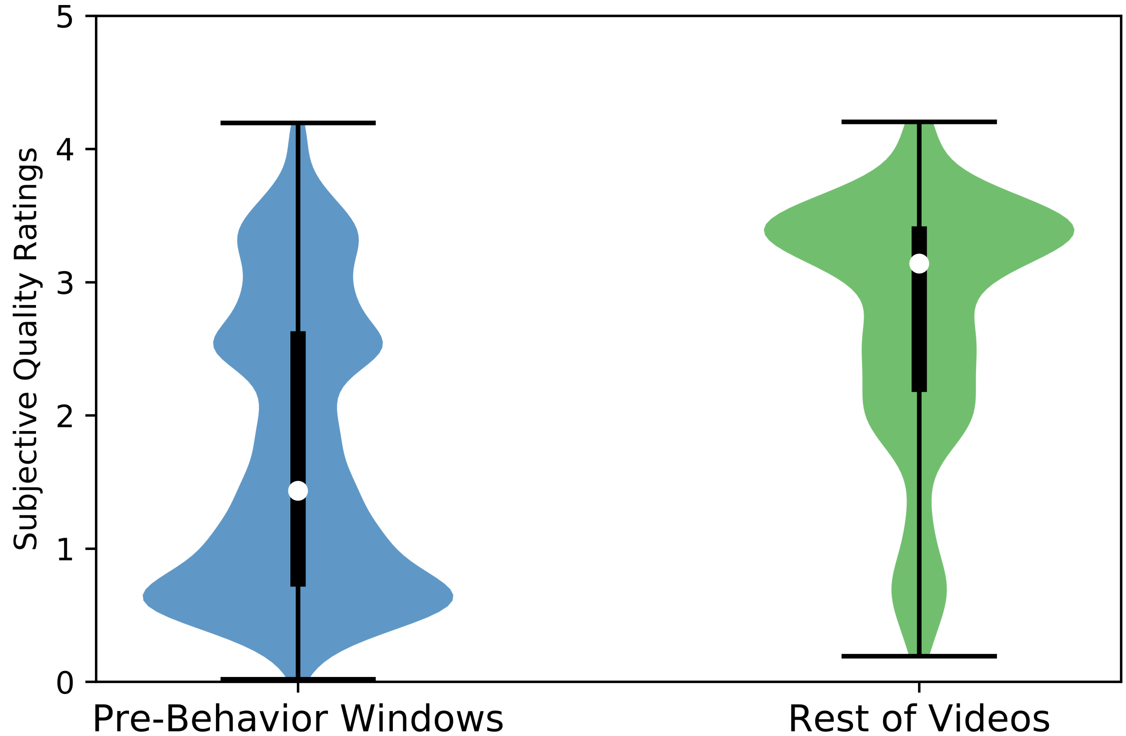


Figure 4.6. Video quality ratings in pre-behavior windows and in the rest of videos. The violin plot shows that video quality ratings within pre-behavior windows are much lower. Two bars in the violin plots represent the maximum and minimum values. The upper edge of the black box represents the third quartile and the lower edge represents the first quartile. The circle in the box depicts the median.

Meanwhile, we recognize that behaviors can be clustered into two groups according to their pre-behavior video quality ratings. Figure 4.7 depicts that pre-behavior video quality ratings lie around low values ($M = 1.6, SD = 1.0$) for behaviors including clicking wireless network button, changing network, changing video resolution, refreshing video player and pausing videos and the behaviors including seeking, changing volume, making video player full screen or normal screen have relative high pre-behavior video quality ratings ($M = 2.2, SD = 1.1$). The rating difference between two groups of behaviors is statistically significant, $U = 1.1e7, p < .001$ (Mann–Whitney U test). It is interesting that the behaviors with high pre-behavior ratings do not have much help on altering video quality and, on the other hand, behaviors with low pre-behavior ratings are able to adjust video qualities and improve video viewing experience effectively.

4.3.4 User Network Switching Behaviors

Different networks have different packet loss rates. Switching networks is an effective way to change the current network condition and improve video viewing experience. Users have to take two consecutive actions to switch to another network. In detail, users need to click wireless network button first to show a list of available networks and then choose a network from the list to complete the network switching. It means that changing network behavior is always after a wireless network button click.

Users Usually Choose to Change Networks First When They Have Bad Video Viewing Experience.

We investigated the first behavior participants performed in each video clip. Figure 4.8 shows that participants usually chose to click wireless button first when they decided to take actions and its occurrences overwhelm other types of behaviors. Since wireless button click is the first step to change networks, we can see that changing wireless networks are participants' first choice when they are not satisfied with the

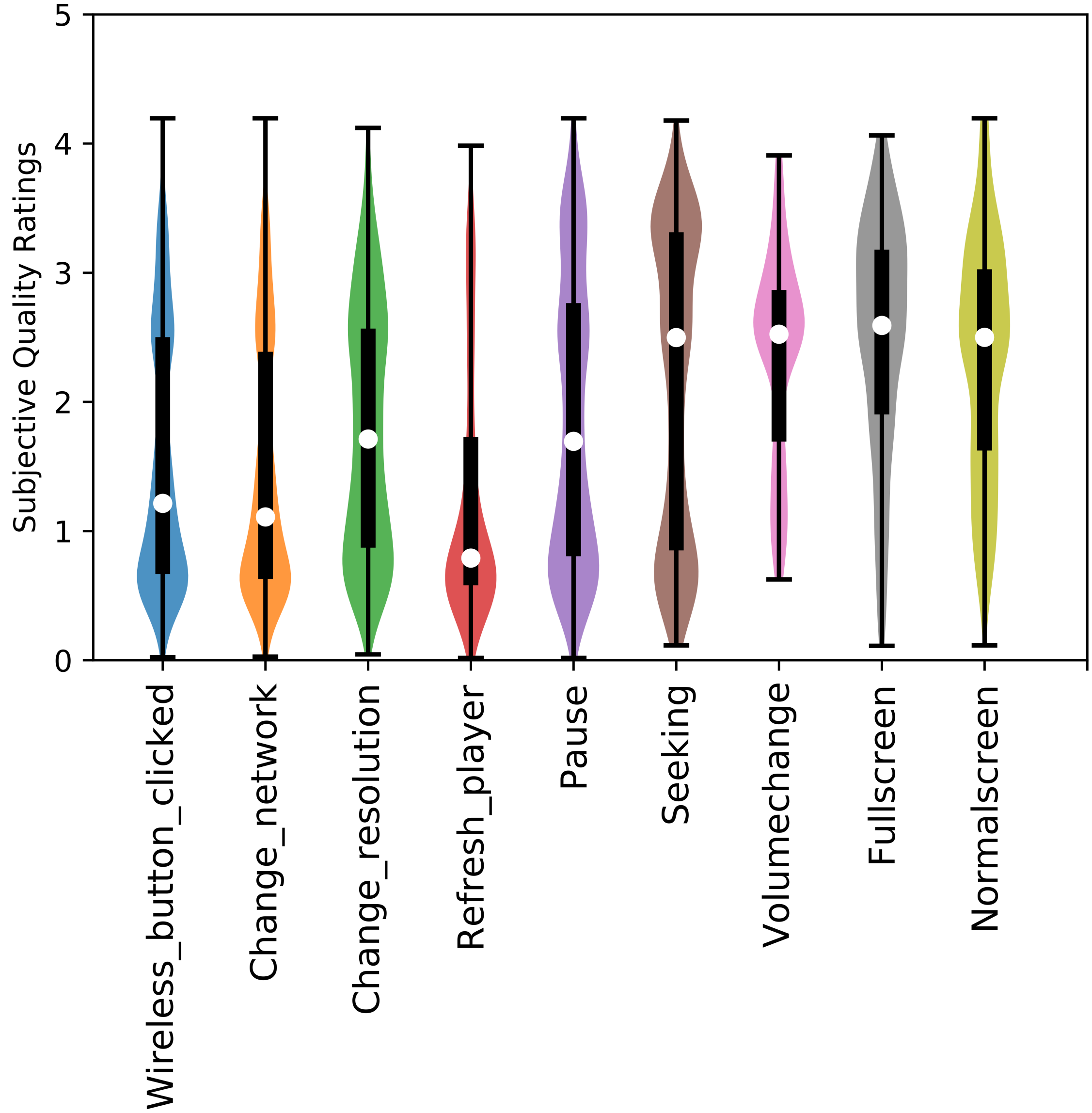


Figure 4.7. Video quality ratings within the pre-behavior window of each behavior type. The violin plot shows that video quality ratings before wireless button click, changing network, changing resolution, refreshing player and pausing video are much lower compared to other types of behaviors. Two bars in the violin plots represent the maximum and minimum values. The upper edge of the black box represents the third quartile and the lower edge represents the first quartile. The circle in the box depicts the median.

viewing experience and they believe that changing wireless networks is able to improve the video quality effectively.

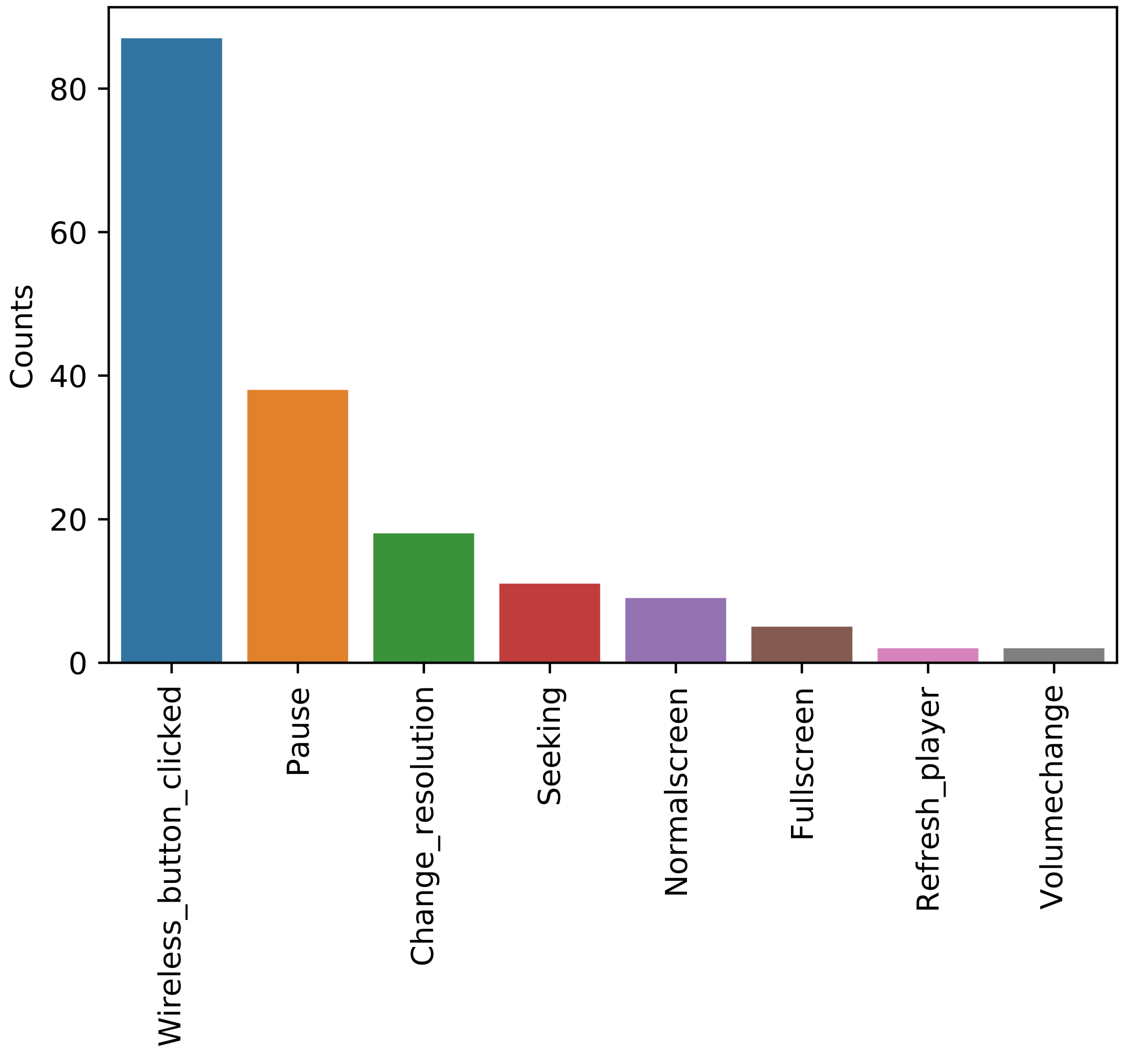


Figure 4.8. The counts of users' first behavior choice in each video clip. The bar plot shows that users were likely to choose to change network first when the video viewing experience was not acceptable.

Video Quality Improvement Can Change Users' Decision on Switching Networks.

We notice that not every wireless button click is followed by the network change. It means that sometimes participants initiated the network changing process by clicking the wireless button, but decided to stop it without choosing another network. Participants changed their mind and did not want to change the network during the process.

We discover that impairment length after wireless button click is much shorter when it is not followed by network change (shown in Table 4.1). Meanwhile, the average time participants spent between wireless button click and network change is 2.0 seconds and this time difference lies between the average impairment length after wireless button click with and without network change. In other words,participants gave up changing networks because the impairment ended and video quality has improved.

Table 4.1. The Relation between Video Impairment Length and Network Change Behavior Sequences

Impairment Length after Wireless Button Click	with Network Change	without Network Change
Mean	2.9 s	1.4 s
Standard	3.8 s	1.3 s

4.3.5 User Video Resolution Changing Behaviors

High video resolutions include more video details and bring people better viewing experiences. However, videos with higher resolutions need more network bandwidth resources in streaming and are more sensitive to network quality variations. Lowering video resolutions sacrifices video details and definitions, but it also reduces the existence of video freezes and block-artifacts and provides smoother video to users.

The Disturbance of Impairments on Video Viewing Experience is More Serious Than That of Low Resolutions.

We have observed 89 video resolution change behaviors. Figure 4.9 shows that 78 of them are lowering current video resolutions while only 11 are switching from a low resolution to a higher one. Meanwhile, 54 lowering video resolution behaviors were occurring within a video impairment and 3 increasing video resolution behaviors were observed in an impairment. In addition, video subjective ratings, shown in figure 4.10, are lower ($M = 1.7, SD = 1.0$) before reducing resolution actions than the ratings ($M = 2.3, SD = 0.6$) before increasing resolution behaviors and the difference is statistically significant, $U = 1.9e4, p < .001$ (Mann–Whitney U test). Lowering resolution within impairments can prevent the continuity of impairments and improve the viewing experience. At the same time, increasing resolution is also an effective method to improve the viewing experience when the network condition is capable to deliver higher resolution videos. Only a few number of increasing resolution behaviors were observed in the experiment depicts that the video impairments bring more dissatisfaction than low resolutions on viewing experience.

4.3.6 User Behaviors and Video Impairments

Impairments caused by network packet losses in streaming videos are explicit factors that motivate users to take actions. We evaluated the video impairment occurrences in pre-behavior windows and discussed their relations with video viewing behaviors.

More Impairments Participants Saw in Pre-behavior Windows, More Quickly They Took Actions during the Current Impairment.

We picked all behaviors that occurred within video impairments and calculated the time difference between behavior occurrence time and impairment's start time. Shorter the time difference is, participants spent less time on making the decision to take actions. Figure 4.11 demonstrates the number of impairments users saw in the pre-behavior window against the time difference between behavior occurrence and

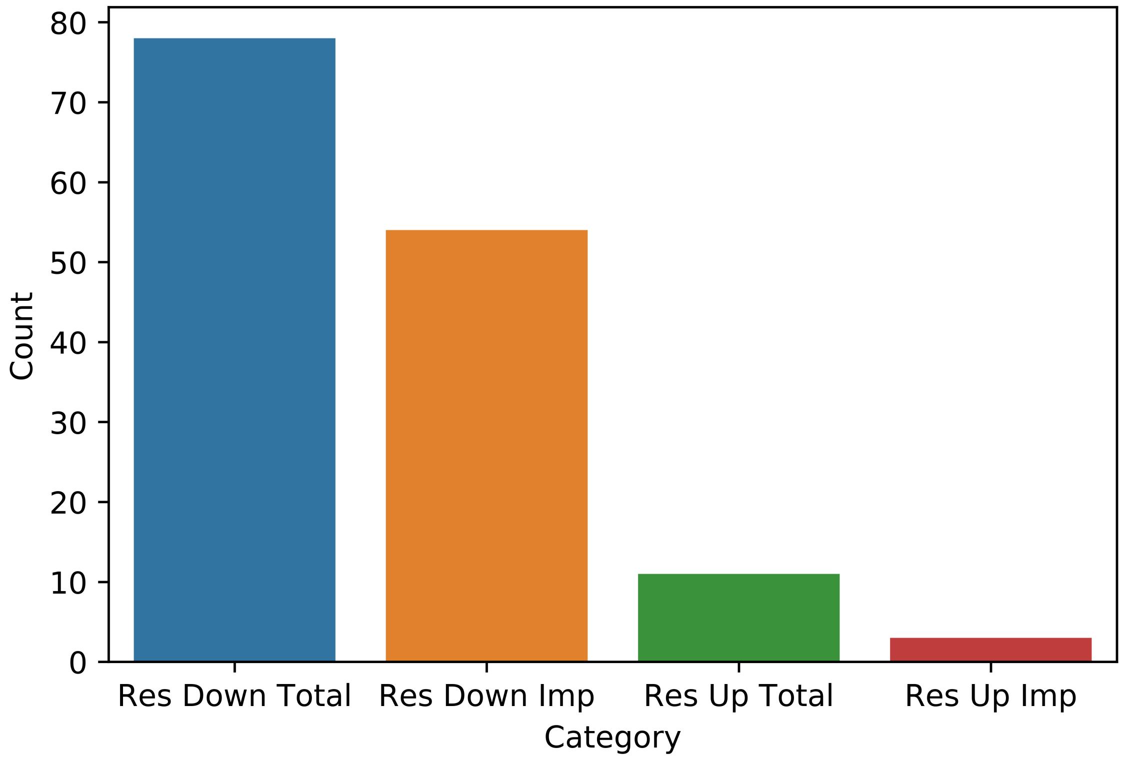


Figure 4.9. The number of raising and lowering video resolution behaviors and the number of these behaviors occurring within an impairment. 'Res Down Total' and 'Res Up Total' mean the total number of moving video resolution down and up behaviors respectively. 'Res Down Imp' and 'Res Up Imp' mean the number of resolution down and up behaviors in an impairment.

impairment start. The time difference decreases when participants have seen more impairments before the current one. It means that previous video qualities and viewing experience effect users' current decision and behaviors.

Modeling User Behavior Decision Time

To derive quantitative evaluation on user behavior decision time, we applied an established cognitive model named ACT-R (Adaptive Control of Thought-Rational (Anderson, Bothell, Lebiere, & Matessa, 1998)). The ACT-R model describes how human memory activation affects human's information retrieval and decision making.

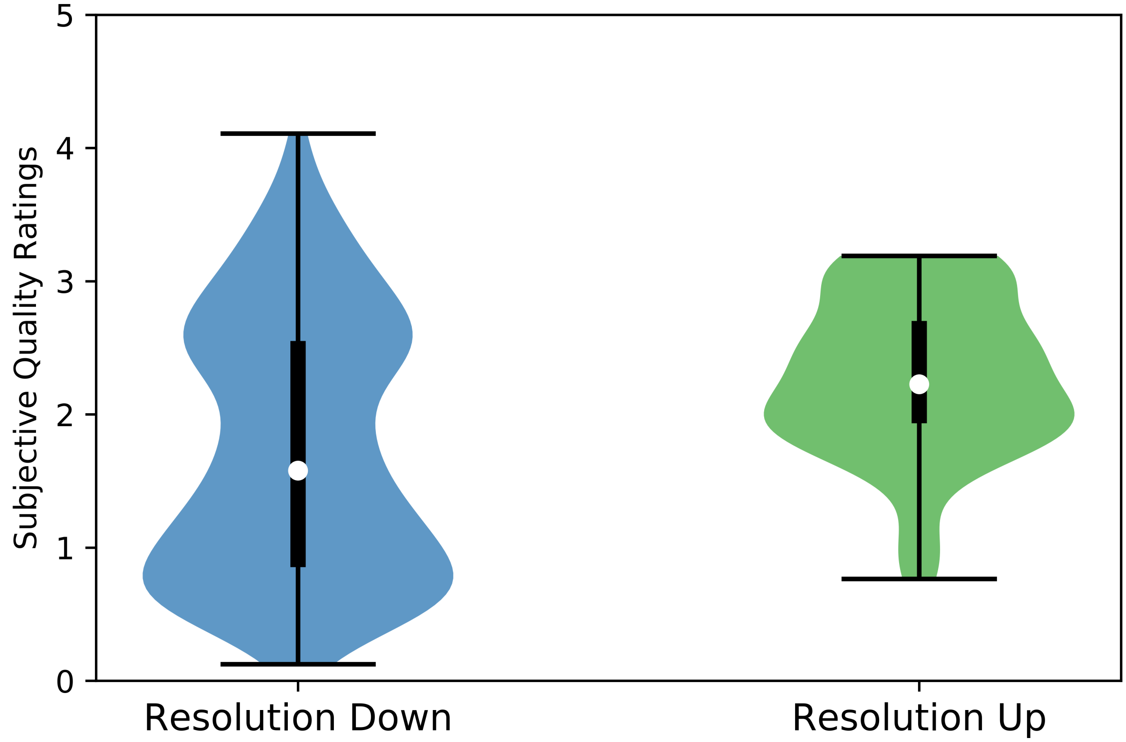


Figure 4.10. Video quality ratings before raising and lowering video resolution behaviors. The violin plot shows that video quality ratings before resolution down behaviors are lower than the ratings before resolution up behaviors. Two bars in the violin plots represent the maximum and minimum values. The upper edge of the black box represents the third quartile and the lower edge represents the first quartile. The circle in the box depicts the median.

In the ACT-R model, stronger activation leads to more quick decision making process. Activation is related to the historical use of this memory chunk and other contextual memory associations (Anderson, 2014; Anderson & Lebiere, 2014). Therefore, the user behavior activation A is

$$A = B + \sum_{i}^{n} W_i S_i \tag{4.1}$$

where A is the activation level for a user behavior, B is the base-level activation, and S_i is the strength of associated activation from element i to the user behavior and W_i

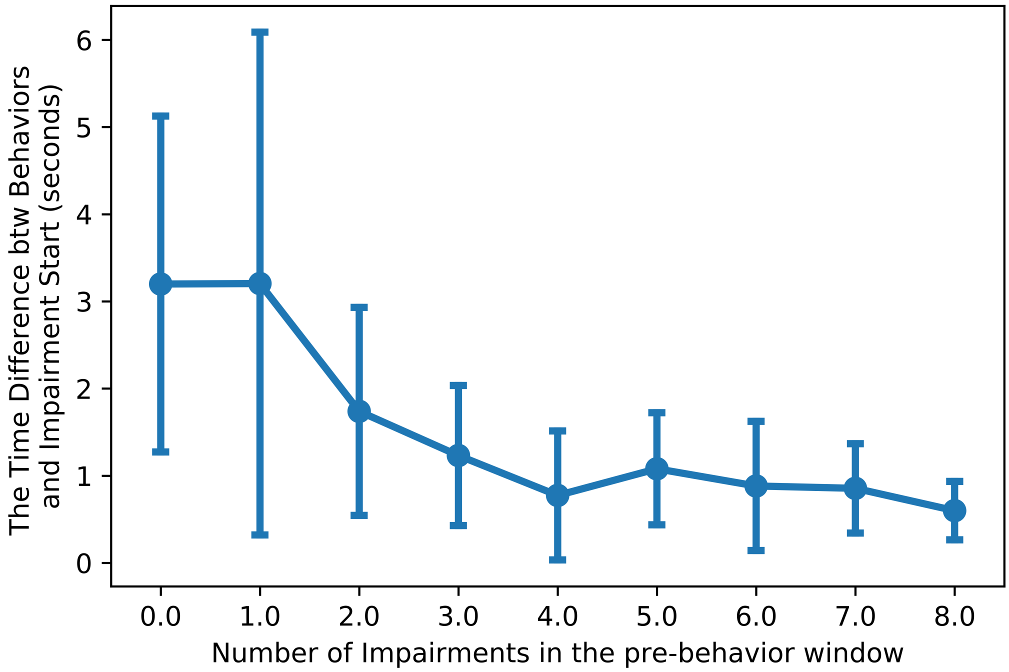


Figure 4.11. The length of time participants spend on taking actions after the start of impairment. The figure shows the time difference between behavior occurrence and the start of impairments when the behavior is within an impairment. More impairments participants saw in the pre-behavior window, shorter the time they spent on deciding to take actions after the current impairment starts. The dot represents average time difference and the stick shows the standard deviation.

is the weight for association S_i. The second term with W_i and S_i is the activation from contextual associations and we interpret it as the video impairments users have seen in the pre-behavior window. The base-level activation B is related to the previous behavior users have performed. B can be calculated throught equation (Anderson & Schooler, 1991)

$$B = ln(\sum_{j}^{m} t_j^{-d}) \tag{4.2}$$

where t_j is the time elapsed since the ith occurrence of the user behavior, m is the total number of previous user behaviors, and d is the memory decay rate. If we assume the occurrences of previous behaviors are evenly distributed, equation (4.2) is approximated to a simpler form:

$$B \approx ln\frac{mL^{-d}}{1-d} \tag{4.3}$$

where L is the time since the first user behavior.

The memory retrieval time (i.e., user behavior decision making time) is exponentially related to the user behavior activation (Anderson et al., 1998; Anderson, Reder, & Lebiere, 1996)

$$Time = Fe^{-M} \tag{4.4}$$

where F is a scale constant, and $M = A - P$. A is the activation level and P denotes the mismatch penalty referring to the similarity of user behaviors to conditions.

To model user behavior occurrence time, we assume that the variability of time users spent to decide taking actions comes from the associated memories of previous behaviors and video impairments. We apply equation (4.4) to compute the behavior occurrence time and we can calculate the expected value of the time by:

$$E[Time] = E[Fe^{-M}] = Fe^{-E[A-P]} \tag{4.5}$$

where the mismatch penalty P can be considered as a random variable with zero mean. After substituting (Equation (4.1), (4.2), and (4.3) to Equation (4.5)), we obtain

$$E[Time] \approx F\frac{1-d}{E[mL^{-d}]}e^{-E[\sum_i^n W_i S_i]} \tag{4.6}$$

We consider the number of behaviors the user took before, m, and the time since the first user behavior, L, as two independent random variables with different constant expected values. We apply Taylor expansion (Benaroya, Han, & Nagurka, 2005) to approximate the function of random variable, $E[L^{-d}] \approx (E[L])^{-d}$. $\sum_i^n W_i S_i$ is the sum of activation from each video impairment the user have seen in the pre-behavior window and we treat its expected value be proportional to the number of video impairments, n. Therefore, we acquire an equation for the average time users spent on deciding to take actions:

$$E[Time] \approx c_0 \frac{1-d}{c_1^{-d}} e^{-c_2 n} \tag{4.7}$$

where d is the decaying rate, n is the number of video impairments in the pre-behavior window, and c_0, c_1, c_2 are three scale parameters. Figure 4.12 shows the fitting curve of our experiment data to the derived equation (4.7) and the fitted parameters are $c_0 = 2.67$, $c_1 = 4.94$, and $c_2 = 0.24$. The memory decaying rate d is 0.29 and we obtained the root mean square error (RMSE) of 0.53 with the user behavior decision time model.

4.4 User Video Watching Behavior Prediction Models

Users' streaming video watching behavior is an important indicator to show their feelings and altitudes towards their watching experience. Meanwhile, some types of behaviors including changing networks, pausing videos and changing video resolutions modify the network bandwidth consumption and could affect the quality

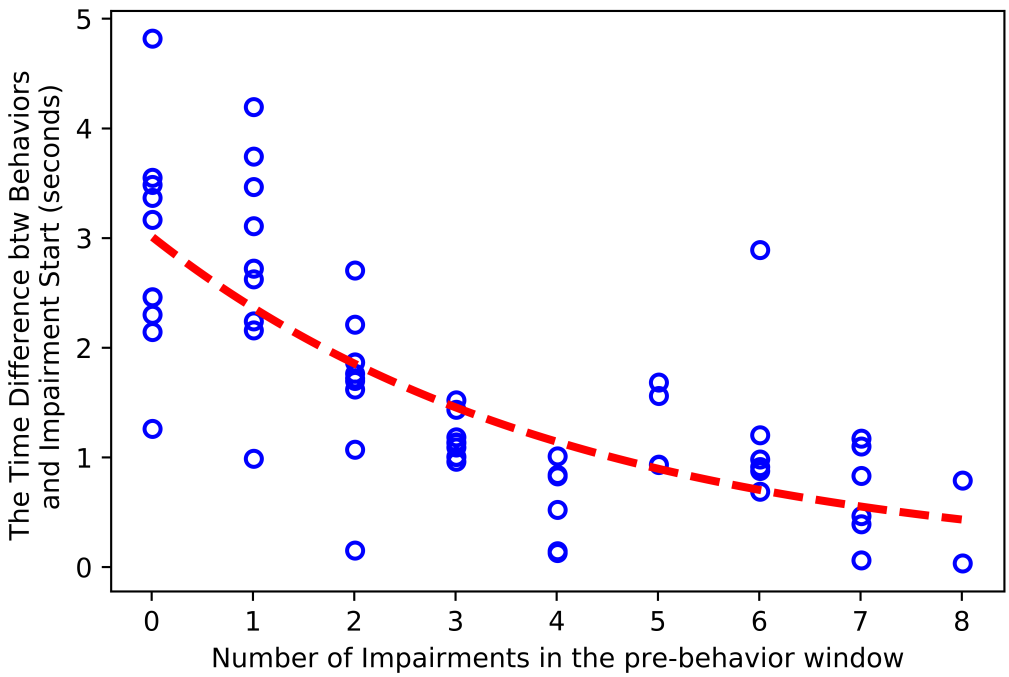


Figure 4.12. Average user behavior decision time per behavior type against number of video impairments in the pre-behavior window. The fitting curve based on the derived equation is shown as the dashed line in the figure.

of experience for other users under the same network. Therefore, users' streaming video watching behavior prediction is valuable for local wireless network bandwidth management. Network controllers can manipulate network bandwidth allocation to prevent or motivate user behaviors for particular purposes. In this section, we compare the performances between different behavior prediction models and discuss the importance for different user behavior features on achieving high model performance.

4.4.1 Multiclass User Behavior Classification

We observed multiple types of user video watching behaviors in our experiments and interpret the behavior prediction as a multiclass classification problem. We chose three classic machine learning algorithms, multinomial logistic regression, random forests and gradient boosted decision trees (GBDT) for our model frameworks.

Multinomial Logistic Regression

Multinomial logistical regression is a classification algorithm that applies logistic regression to predict the probabilities of the different outcomes of a categorical variable based on a set of feature variables (William, 2012). The multinomial logistic regression can be expressed analytically by

$$\Theta = \alpha + X \cdot \beta \tag{4.8}$$

$$P(Y = y_i|X) = \frac{e^{\theta_i}}{\sum_{j=1}^{K} e^{\theta_j}} \tag{4.9}$$

where X is the feature vector and $P(Y = y_i|X)$ shows the probability that the behavior belongs to type i given the feature X. α and β are normal distributed random matrix and random vector respectively. We define probability P as $P = Softmax(\alpha + X \cdot \beta)$ and represent observation Y using a categorical distribution, $Y \sim Categorical(P)$.

Random Forests

Random forests classifier is an ensemble learning method that constructs a number of decision trees at training process and labels the mode of classes of the individual trees (Tin Kam Ho, 1995). Specifically, we created a forest with separate trees and trained the forest with our training dataset. The number of trees in the forest and maximum tree depth parameters are determined according to the model performance in the training and validation. We then input unlabeled behavior sample to the trained forest and set its type by the majority of each individual tree of the forest.

Gradient Boosted Decision Trees (GBDT)

Gradient boosted is also an ensemble learning method that improves the accuracy of a predictive function by minimizing the error term using a serial of weak prediction models (Friedman, 2001). We applied decision trees as the weak prediction models. Each tree in the series is created and fitted the "pseudo residuals" generated by the prediction from previous trees to reduce the error. This leads to the following model

$$F(\mathbf{X}) = \beta_0 + \sum_{i=1}^{n} \beta_i T_i(\mathbf{X}) \tag{4.10}$$

where β_0 is the constant term of the model, $T_1, \ldots, T_n$ are the trees fitted to the pseudo-residuals and, β_i are corresponding coefficients to the trees computed by gradient optimization methods.

4.4.2 User Behavior Feature Extraction

We extracted 23 features for each user behavior sample and categorized them into four groups based on feature properties and relations to other video information. These features capture user behavior characteristics from different aspects and have comprehensive description of each individual user behavior sample.

Video impairment related features

- **Time difference from the impairment starts** (impLenBeforeAction): The time difference between the start of video impairment and the occurrence of this user behavior. It is a numerical variable. If the user behavior occurs outside any video impairment, the value is -1.

- **Time difference from the last impairment ends** (timeDiffPrevImp): The time difference between the end of last video impairment and the occurrence of this user behavior. It is a numerical variable. If the user behavior occurs inside an video impairment, the value is -1.

- **Number of previous impairments** (trueNumPrevImp): The number of video impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Average length of previous impairments** (meanLenPrevImp): The average length of video impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Standard deviation length of previous impairments** (stdLenPrevImp): The standard deviation of length of video impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Type of video impairment when the behavior occurs** (actionImp): The video impairment is labeled 0 for freeze and 1 for block-artifacts. If the behavior occurs outside any video impairment, the label is 2.

- **Is the behavior occurring in an impairment** (isInImp): The behavior is labeled 1 if it occurs inside a video impairment, otherwise 0.

- **Number of previous freeze impairments** (trueNumPrevFreezeImp): The number of freeze impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Average length of previous freeze impairments** (meanLenPrevFreezeImp): The average length of freeze impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Standard deviation length of previous freeze impairments** (stdLenPrevFreezeImp): The standard deviation of length of freeze impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Number of previous block-artifacts impairments** (trueNumPrevBlockImp): The number of block-artifacts impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Average length of previous block-artifacts impairments** (meanLenPrevBlockImp): The average length of block-artifacts impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

- **Standard deviation length of previous block-artifacts impairments** (stdLenPrevBlockImp): The standard deviation of length of block-artifacts impairments observed in the pre-behavior window of the target behavior. It is a numerical variable.

User behavior related features

- **Order of behavior occurring in the impairment** (action_order_inImp): The order of the target behavior in all behaviors observed within the same video impairment. It is an ordinal variable. If the target behavior occurs outside any video impairment, the order value is -1.

- **Order of behavior occurring in the video clip** (action_order_inVideo): The order of the target behavior in all behaviors observed in the same video clip. It is an ordinal variable.

- **Number of previous behaviors** (numPrevBehaviors): The number of behaviors observed in the pre-behavior window of the target behavior for the same user. It is a numerical variable.

- **Type of the previous behavior** (lastPrevBehavior): The type of behavior before this targeted behavior for the same user. The behavior type is a categorical variable and labeled with different positive integers. If the targeted behavior is the first one, its previous behavior is labeled as -1.

- **Time difference from the previous behavior** (timeDiffToPrevBehavior): The time difference between the previous behavior and the target behavior. It is a numerical variable.

Video subjective quality related features

- **Average of video subjective ratings**: The average of video ratings in the pre-behavior window of the target behavior. It is a numerical variable.

- **Standard deviation of video subjective ratings**: The standard deviation of video ratings in the pre-behavior window of the target behavior. It is a numerical variable.

Other features

- **Type of video content** (content): The type of video content. The label is 0, 1, 2, and 3 for news, education, movie and sports videos respectively.

- **Video resolution level** (action_res): The level of video resolution when the target behavior occurs. It has three values which are 360, 480 and 720.

- **Network packet loss rate** (action_net): The network packet loss rate when the target behavior occurs. It has three values which are 2.5, 5 and 10.

4.4.3 Balancing User Behavior Samples

Our observation shows that the number of switching connected networks related behaviors (i.e., Wireless_button_clicked and Changing_network) surpasses the half of total user behaviors in the experiment. The imbalanced dataset across different label classes results in serious degradation of model performance and misleads model evaluation metric interpretations (Japkowicz & Stephen, 2002; Batista, Prati, & Monard, 2004). In order to handle the imbalanced dataset issue, we simply applied random oversampling approach on behavior types with small amount of samples. In this method, we chose the behavior type class with the largest sample size and picked its total sample number as the target amount. For each of the rest behavior type class, we randomly selected and replicated the samples and added them to the original dataset until the total sample size of this type class reaches the target amount. The total number of samples in every behavior type class was equal after the dataset balancing.

4.4.4 "No Action" Behaviors During Video Watching

Users behaviors during video watching show their reactions towards video impairments and quality changes. Users were not satisfied with the video quality and wanted to improve the watching experience through different actions. Meanwhile, if users did not take any actions, it means they felt that the video watching experience is acceptable and there is no need to take any movement to change the current status. Therefore, we can categorize "No action" as a new type of behavior reflecting users' attitude to the current video quality and watching experience.

We created "No action" behavior samples and added them to the user behavior dataset as one additional behavior type. We made the number of "No action" behavior samples the same as the number of other behaviors in order to keep the balance of dataset. We selected the same features as other user behaviors and formed the corresponding feature vector for every "No action" behavior sample.

4.4.5 User Behavior Feature Selection and Model Performance

We applied five-fold cross validation method and compared averaged accuracy, precision and recall values of different behavior prediction models. K-fold cross validation is widely used to evaluate machine learning models with limited data samples. It splits data into k groups/folds approximately with equal size. One fold is used as validation set and the model is trained on the rest k-1 folds. We chose five-fold cross validation and got the average performance of the five runs for each model.

Precision and recall are key metrics to evaluate the performance of binary classification models. In our multiclass behavior classification, we used averaging approach to get the model's precision and recall. In detail, we first got the precision and recall for each class label as a binary classification, and calculated the average value across all classes to acquire the final precision and recall values for the model.

We ran recursive feature elimination to rank features and applied the cross validation to select the best number of features according to average classification accuracy for every model. Recursive feature elimination (RFE) is a feature selection algorithm (Kuhn, Johnson, et al., 2013) and it first includes entire set of features and selects features by recursively considering smaller subset of features and eliminates the features with lower important ranks. At each step, the model is trained using dataset with selected features and the average classification accuracy of cross validation is recorded. We chose the highest average accuracy to present the performance for each model. Table 4.2 shows the accuracy, precision, recall, and the corresponding number of selected features for every model.

Table 4.2. The Accuracy, Precision, Recall and Number of Selected Features for Different User Video Behavior Prediction Models

Model	# of Features	Accuracy (%)	Precision (%)	Recall (%)
Baseline	-	10	-	-
MLR	21	44.1	42.9	44.1
RF	16	93.1	92.7	92.7
GBDT	**21**	**94.7**	**92.8**	**94.7**

The baseline model does not use any behavior feature information and simply as-sign the behavior type to each unlabeled sample randomly with equal probability. We have ten candidate behavior types and the baseline model accuracy is 10%. Multinomial logistic regression (MLR) has the lowest performance among all chosen models except the baseline model. It has less than 45% prediction accuracy. Two tree based approaches, random forests (RF) and gradient boosted decision trees (GBDT), gain similar prediction accuracy and have a huge jump from MLR to go over 93%. GBDT model achieves the best accuracy of 94.7%.

4.4.6 User Behavior Feature Importance

Gradient boosted decision trees (GBDT) model reaches the highest prediction accuracy with 21 features. We analyzed the importance of features in this model and discussed the subset of features that are crucial to identify different behavior types. Impurity-based feature importance and permutation importance are two applied metrics to rank the feature importance for tree-based models (Breiman, 2001). The two importance metrics are computed on training and test set statistics respectively and provide comprehensive evaluations on the model and dataset. Figure 4.13 shows the impurity-based feature importance (MDI) and feature permutation importance ranking of our GBDT based user behavior classification model and figure 4.14 illustrates the user behavior prediction accuracy against the number of selected features. We can see that the cross validation score increases quickly and goes above 0.9 when the first five most important features are included and scores stay with small increment if more features are added. It tells us that the first five features on the importance ranking list are decisive for the GBDT model to give highly accurate user behavior prediction.

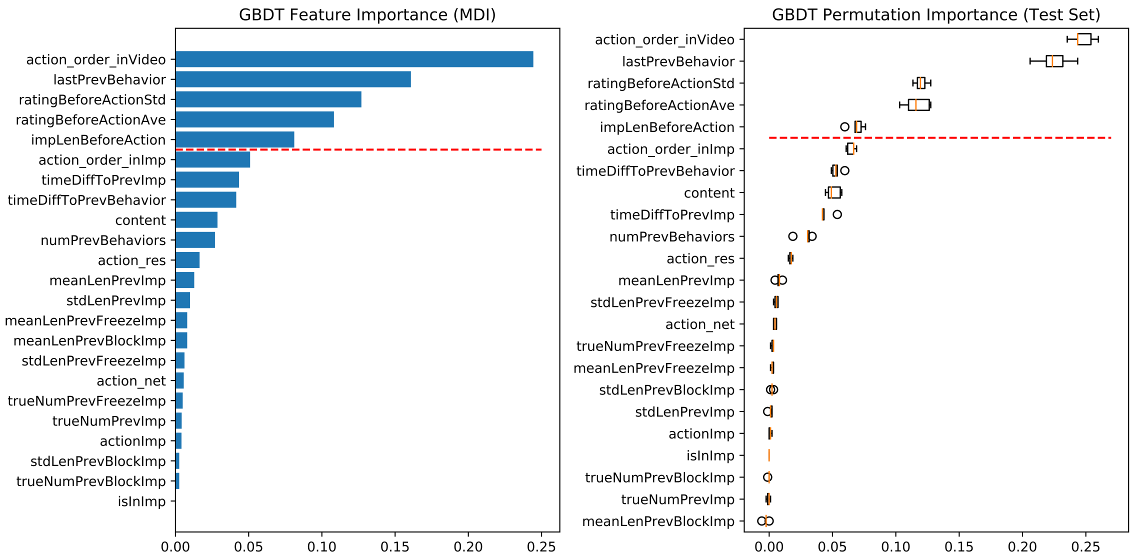


Figure 4.13. Impurity-based feature importance (MDI) (Left) and feature permutation importance (Right) of the gradient boosted decision trees (GBDT) model for user video watching behavior prediction. Features above the red dashed line are decisive for good behavior prediction.

Figure 4.13 depicts that the two feature importance metrics recommend the same set of features with top-five importance. These features include

- **Order of behavior occurring in the video clip** (action_order_inVideo): The order of the target behavior in all behaviors observed in the same video clip. We collected the behaviors each participant performed in every video clip separately and labeled them according to their chronological orders. This feature indicates the participants' preferences in different types of behaviors.

- **Type of last behavior** (lastPrevBehavior): The type of behavior before this targeted behavior. This feature captures correlations between consecutive behaviors.

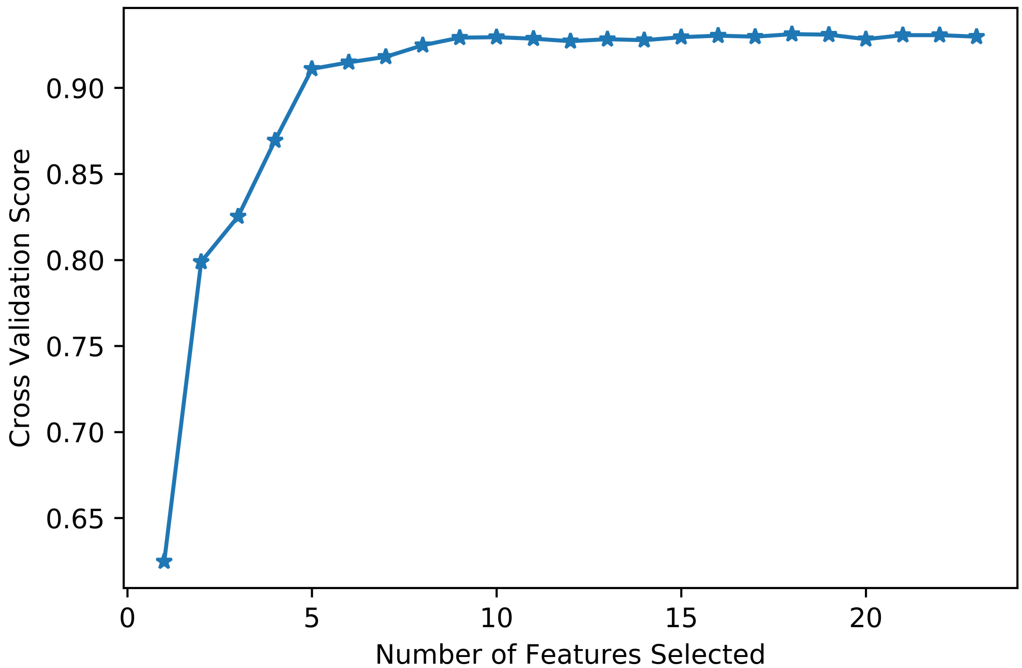


Figure 4.14. The cross validation scores of GBDT model with different number of feature selected.

- **Standard deviation of video subjective ratings**: The standard deviation of video ratings in the pre-behavior window of the target behavior. This feature tells us the level of video quality variations before participants' behaviors.

- **Average of video subjective ratings**: The average of video ratings in the pre-behavior window of the target behavior. This feature shows the general video quality levels before participants taking actions.

- **Time difference from the impairment starts** (impLenBeforeAction): The time difference between the start of video impairment and the occurrence of the target behavior. This feature demonstrates how long participants waited to take actions after they saw a video impairments.

4.4.7 Human Cognition behind Users' Choice of Behaviors

The five most important features characterize how users make decisions to take different actions during streaming video watching. In this section, we analyze the human cognition behind the process of user decision and behavior.

Soar is a widely used cognitive architecture to approximate human cognitive processes and explain human behaviors (Rosenbloom, Laird, Newell, & McCarl, 1991; Laird, 2012). Figure 4.15 illustrates the Soar-based user video watching behavior model. The model is hierarchical and built by levels. The first level is the environment and external incentives. In our model, it includes, for instance, video impairment occurrences, different video resolutions and smooth playback. The environment information is received by users' visual and auditory perceptions and influences their short-term working memory. Short-term working memory is the core level and have decisive influence on human decision and behavior. Our feature importance analysis shows that users pay more attention on previous video quality and its variations, what type of behaviors they have performed and what feedback they have received after that. Three parts of long-term memory are lying above the short-term memory. The semantic memory stores general facts. For instance, the available functions of the interface (i.e., video player), the concepts of network packet loss and video resolution are all belongs to semantic memory. The second part is the episodic memory and it relates to the former personal experience to the specific incident. In streaming video watching, it includes users previous experience on their behaviors when they were watching online video at home or in other occasions. The procedural memory forms a person's character and includes knowledge about what to do and when to do it. In our experiment, taking some actions or doing nothing and just watching the video are parts of procedural memory.

The short-term working memory and long-term memories are associated. Users' previous video watching experience in other occasions, understandings of some network and video related facts, and personal characteristics can affect and modify the short-term working memory. Meanwhile, the short-term memory can form and update their long-term memories. Along with the goals of having better video watching experience, users make decisions and take actions based on a mixture of external environment incentives, contents in working memory and three parts of long-term memory.

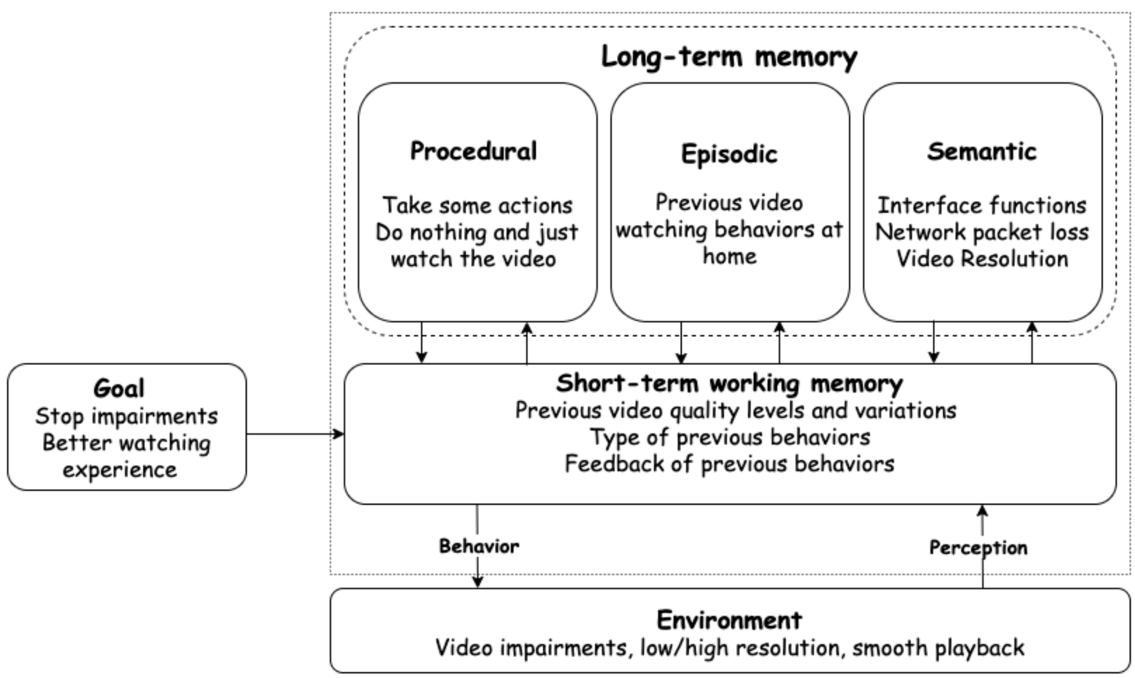


Figure 4.15. Human cognition behind users' choices of video watching behavior.

4.5 Summary

We first organized a user experiment to observe participants' video viewing behaviors under different packet loss wireless networks. We analyzed the video quality ratings before user behaviors and compared their relations with different types of behaviors and found that low video qualities caused user behaviors and users were trying to take action, including change networks, change video resolutions, refresh the player and pause the video, to change the current video qualities. Additionally, users had preferences on choosing particular types of actions. Our observation shows that they usually choose to change networks first when they have bad video viewing experience and video quality improvement can also change users' decisions on changing networks. Moreover, we noticed that more impairments users have seen before, more quickly they took actions during the current one. In other words, the bad video viewing experience memories make users spend less time to take actions if the bad experience continuous. We established an ACT-R based mathematical model to quantify how human memories affect viewers' decisions on taking actions. Our findings provide evidence on predict users viewing behaviors according to video qualities and ideas to optimize network management and deliver acceptable video viewing experiences to users under resource-constrained network environments. Based on the experiment data, we created different users' video viewing behavior prediction models and compared their performances. Our model achieves 94.7% prediction accuracy. Furthermore, we analyzed the importance of different features for precise behavior prediction and built a human cognition model to explain the process of human decision making and behaviors. User behaviors indicate their opinions towards video watching experience and their decisions to change the current condition. In the same time, behaviors can change the network resource demand for single users and affect the quality of services for other users in the network. User behavior prediction provides clues of users' feelings about perceived service qualities and experiences as well as potential influences on the entire network resource consumption. Our models bring network controllers tools to design, analyze and practice network resource management protocols via human engineering.

CHAPTER 5

DISCUSSIONS

5.1 User Subjective Rating on Streaming Videos under Packet Loss Wireless Networks

People are increasingly preferable to view streaming videos with their mobile devices on wireless networks. High packet losses in wireless networks are major issues to introduce video impairments. In this study, we analyzed the influences of different individual and sequences of impairments on subjective video viewing experience over networks with controlled packet loss rates that emulate streaming video viewing on mobile devices over heavily utilized wireless networks. Next, we will highlight our most important results.

Participants preferred high-resolution videos with occasional impairments over smooth low-resolution videos. The average ratings for 720p videos with impairment ratio less than 0.2 were higher than all 360p videos without any impairments, and introducing some video impairments does not always lower video viewing experience. This observation disagrees with the current HTTP adaptive streaming protocol that prioritizes video smoothness over resolution. It shows that using this method for online streaming under networks with limited bandwidth is not always the best choice for better subjective viewing experience. Under networks with known available bandwidth, choosing videos with good trade-off between resolution and impairment occurrence can maximize network resource utilization and provide users with better viewing experience. This idea is beneficial when users are streaming videos with their mobile devices under wireless networks where bandwidth is precious.

Meanwhile, participants felt block-artifacts impairment after a video freeze is acceptable and video freezes happening after block-artifacts impairments keep worsening the viewing experience. Most of subsequent freezes were observed in high-bitrate videos on low-bandwidth networks, and high-bitrate videos on low-bandwidth networks had more and longer freeze and block-artifacts impairments. Also, participants agreed that longer video impairments have larger influences on video viewing experience. These findings indicate that providing network bandwidth larger than streaming video's bitrate is an effective approach to reduce video impairments and improve video viewing experience.

However, giving adequate network bandwidth to every individual mobile device user is impractical in heavily utilized wireless networks. We notice that participants had different evaluations on short impairments. Some short impairments have a very small amount of rating drops, which shows that participants ignored these impairments or believed that they did not worsen the video viewing experience. These impairments either have a relatively long or a very short impairment-free period in front of themselves. Specifically, the impairments with high minimum ratings and small rating drops have an average 9.5-second long impairment-free period, and participants did not drop the rating too much and it remained at a high value at the end of impairments. Meanwhile, the impairments with low minimum rating and small rating drops have an average 0.2-second long impairment-free period, and participants felt that these short impairments did not worsen the viewing experience even further. In addition, participants felt that video viewing experience improves during 81% of block-artifacts impairments which are occurring after a freeze and it tells us that block-artifacts after a freeze is acceptable. These observations provide evidence that changing impairment's occurring order and time can improve viewing experience even the total impairment number is unchanged. It is encouraging and brings alternatives to optimize network management in resource-constrained situations.

5.2 User Viewing Behaviors on Videos Streaming under Packet Loss Networks

We studied how users behave when they are watching streaming videos under controlled packet loss networks that emulate conditions of extensively utilized wireless networks. We have several significant findings and results.

Most importantly, our study reveals how users' video viewing behaviors are related to video quality and impairment occurrences. Based on our findings, we novelly proposed several user video viewing behavior prediction models and the gradient boosted decision trees (GBDT) model reaches 94.7% prediction accuracy. Additionally, we found that the GBDT model achieves more than 90% accuracy with the five most important features and we discussed the how human cognition works behind users' choices of different types of behaviors.

Also, participants take actions because video quality is low and choose actions they believe can improve video quality and viewing experience. Video packet losses during data transmission results in impairments and degrades video quality. Low video qualities destroy the smoothness and clearness of video delivery to viewers and hurts their viewing experience. Our study proves that low video quality is responsible for the occurrences of user behaviors. The video quality ratings before user behaviors ($M = 1.7, SD = 1.1$) are much lower than the ratings during the rest of videos ($M = 2.7, SD = 0.9$). When viewers received video contents with bad qualities, they expressed more willingness to change the current condition and chose behaviors including switching to another network, changing video resolution, refreshing video player and pausing the video which can provide direct influence on network condition and video quality. Therefore, bad video qualities motivate participants to take actions and the types of behavior they chose are the ones can change the video quality and improve viewing experience.

In addition, participants' decisions on whether or when to take an action is mainly based on what they have seen recently. We found that more impairments participants saw during the pre-behavior window, more quickly they took actions during the current impairment. Human's short-term memory plays an important role in the decision making process. Bad experience in the past stays in participants' memories and cause them to make quick decision to change the current conditions when the video quality and viewing experience becomes unacceptable again. We also built ACT-R based cognitive model to quantify the relations between users' decision time and the number of video impairments within the pre-behavior window. In addition, our study also shows that the time participants spent on making the decision was not affected by the length and types of video impairment in pre-behavior window. This observation tells us that participants paid less attention on impairments details and more focused on the influences on video viewing experience due to their occurrences.

Furthermore, participants have preferences on particular types of behaviors. We found that switching to another available network was frequently chosen as the first option when participants felt that the video quality is bad. At the same time, video quality improvement can change their decisions on switching networks. Our observation proves that participants evaluated the video viewing experience and made decision to take actions in a dynamic manner. The video quality changes during this process are able to alter participants' following decision and behaviors. We also noticed that most changing video resolution behaviors were decreasing the resolution and about 70% (54 out of 78) of them were happening within an impairment. Lowering video resolution reduces the consumption of network bandwidth and can prevent video impairments from occurring and improve video viewing experiences under resource-constraint networks. Meanwhile, participants rarely increased video resolution to have much clearer video contents. We draw conclusion that video impairments bring more and severe disturbance than low resolution does to viewers' video viewing experiences.

CHAPTER 6

LIMITATIONS

We conducted both of the two experiments in a well controlled laboratory. Participants were invited to an office room and asked to use a computer for the experiment. This may impact their assessment on the video viewing experience and viewing behaviors compared to their daily lives. We took several steps towards achieving validity for both studies. In the video rating experiment, we provided three video clips as the training session at the beginning of the experiment and let participants get familiar with the interface and the types of video impairments you will meet in formal experiment video clips. In addition, we examined and conducted sanity checks on the rating data provided by every participant. In the video viewing behavior experiment, we applied deception in the experiment to conceal the real experiment purpose. This method prevented participants from behaving intentionally during experiment and introduced potential bias into the dataset. Furthermore, each participant was asked to complete an exist survey that included evaluation of their own performance and results shown that all participants believed they did well in the experiment.

CHAPTER 7

CONCLUSIONS

We have shown that low video quality drives people to take actions to change current conditions and try to improve the watching experience and users' previous watching experience plays a key role in their decisions to perform different behaviors. We proposed high accurate user behavior prediction models that provide clues of users' feelings about perceived service experiences and potential influences on the entire network resource consumption. We are the first to study the effect of impairment occurring orders on users' video watching experience and demonstrate that it is feasible to improve the experience without additional resource usage by changing video impairments types and occurrence orders. Our study provides encouraging ideas to improve user video watching experience on mobile devices with limited network resources and opens the door to design and practice human engineering based protocols to optimize wireless network management in bandwidth resource constrained environments.

BIBLIOGRAPHY

Afanasyev, M., Chen, T., Voelker, G. M., & Snoeren, A. C. (2008). Analysis of a mixed-use urban wifi network: When metropolitan becomes neapolitan. In *Proceedings of the 8th acm sigcomm conference on internet measurement* (pp. 85–98). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1452520.1452531 doi: 10.1145/1452520.1452531

Amsel, A., & Roussel, J. (1952). Motivational properties of frustration: I. effect on a running response of the addition of frustration to the motivational complex. *Journal of experimental Psychology*, *43*(5), 363. doi: http://dx.doi.org/10.1037/h0059393

Anderson, J. R. (2014). *Rules of the mind*. Psychology Press.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341–380.

Anderson, J. R., & Lebiere, C. J. (2014). *The atomic components of thought*. Psychology Press.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive psychology*, *30*(3), 221–256.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396-408. Retrieved from https://doi.org/10.1111/j.1467-9280.1991.tb00174.x doi: 10.1111/j.1467-9280.1991.tb00174.x

Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), (Vol. 2, p. 89 - 195). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079742108604223 doi: https://doi.org/10.1016/S0079-7421(08)60422-3

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 35–44). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1753326.1753333` doi: 10.1145/1753326.1753333

Balachandran, A., Voelker, G. M., Bahl, P., & Rangan, P. V. (2002). Characterizing user behavior and network performance in a public wireless lan. In *Proceedings of the 2002 acm sigmetrics international conference on measurement and modeling of computer systems* (pp. 195–205). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/511334.511359` doi: 10.1145/511334.511359

Baltrunas, D., Elmokashfi, A., & Kvalbein, A. (2014). Measuring the reliability of mobile broadband networks. In *Proceedings of the 2014 conference on internet measurement conference* (pp. 45–58). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2663716.2663725` doi: 10.1145/2663716.2663725

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004, June). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.*, *6*(1), 20–29. Retrieved from `https://doi.org/10.1145/1007730.1007735` doi: 10.1145/1007730.1007735

Benaroya, H., Han, S. M., & Nagurka, M. (2005). *Probability models in engineering and science* (Vol. 193). CRC press.

Boulos, F., Parrein, B., Le Callet, P., & Hands, D. S. (2009, Jan). Perceptual effects of packet loss on h.264/avc encoded videos. In *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-09* (p. -). Scottsdale, Arizona, United States. Retrieved from `https://hal.archives-ouvertes.fr/hal-00354947`

Boyce, J. M., & Gaglianello, R. D. (1998). Packet loss effects on mpeg video sent over the public internet. In *Proceedings of the sixth acm international conference on multimedia* (pp. 181–190). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/290747.290770` doi: 10.1145/290747.290770

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

BT.500, I.-T. (2012). Methodology for the subjective assessment of the quality of television pictures. *ITU-T RECOMMENDATION, BT.500*. Retrieved from http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-12-200909-S!!PDF-E.pdf

Campion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goal-setting and changing process. *Organizational Behavior and Human Performance*, *30*(2), 265 - 287. Retrieved from http://www.sciencedirect.com/science/article/pii/0030507382902215 doi: https://doi.org/10.1016/0030-5073(82)90221-5

Che, X., Ip, B., & Lin, L. (2015, Apr). A survey of current youtube video characteristics. *IEEE MultiMedia*, *22*(2), 56-63. doi: 10.1109/MMUL.2015.34

Chen, K.-T., Huang, P., & Lei, C.-L. (2006, November). How sensitive are online gamers to network quality? *Commun. ACM*, *49*(11), 34–38. Retrieved from http://doi.acm.org/10.1145/1167838.1167859 doi: 10.1145/1167838.1167859

Chen, L., Zhou, Y., & Chiu, D. M. (2013, July). Video browsing - a study of user behavior in online vod services. In *2013 22nd international conference on computer communication and networks (icccn)* (p. 1-7). doi: 10.1109/ICCCN.2013.6614209

Chen, X., Jin, R., Suh, K., Wang, B., & Wei, W. (2012). Network performance of smart mobile handhelds in a university campus wifi network. In *Proceedings of the 2012 internet measurement conference* (pp. 315–328). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2398776.2398809 doi: 10.1145/2398776.2398809

Chen, Z., & Wu, D. (2012, March). Prediction of transmission distortion for wireless video communication: Analysis. *IEEE Transactions on Image Processing*, *21*(3), 1123-1137. doi: 10.1109/TIP.2011.2168411

Cheng, X., Dale, C., & Liu, J. (2008, June). Statistics and social network of youtube videos. In *2008 16th interntional workshop on quality of service* (p. 229-238). doi: 10.1109/IWQOS.2008.32

Copyright law of the united states and related laws contained in title 17 of the united states code. (2016). Retrieved from https://www.copyright.gov/title17/title17.pdf

De Simone, F., Naccari, M., Tagliasacchi, M., Dufaux, F., Tubaro, S., & Ebrahimi, T. (2009, July). Subjective assessment of h.264/avc video sequences transmitted over a noisy channel. In *2009 international workshop on quality of multimedia experience* (p. 204-209). doi: 10.1109/QOMEX.2009.5246952

Dobrian, F., Awan, A., Joseph, D., Ganjam, A., Zhan, J., Sekar, V., ... Zhang, H. (2013, March). Understanding the impact of video quality on user engagement. In (Vol. 56, pp. 91–99). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2428556.2428577` doi: 10.1145/2428556.2428577

Duanmu, Z., Ma, K., & Wang, Z. (2017). Quality-of-experience of adaptive video streaming: Exploring the space of adaptations. In *Proceedings of the 25th acm international conference on multimedia* (pp. 1752–1760). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/3123266.3123418` doi: 10.1145/3123266.3123418

Falaki, H., Lymberopoulos, D., Mahajan, R., Kandula, S., & Estrin, D. (2010). A first look at traffic on smartphones. In *Proceedings of the 10th acm sigcomm conference on internet measurement* (pp. 281–287). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1879141.1879176` doi: 10.1145/1879141.1879176

Feild, H. A., Allan, J., & Jones, R. (2010). Predicting searcher frustration. In *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval* (pp. 34–41). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/1835449.1835458` doi: 10.1145/1835449.1835458

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. Retrieved from `http://www.jstor.org/stable/2699986`

Frnda, J., Voznak, M., & Sevcik, L. (2016, Jun 01). Impact of packet loss and delay variation on the quality of real-time video streaming. *Telecommunication Systems*, *62*(2), 265–275. Retrieved from `https://doi.org/10.1007/s11235-015-0037-2` doi: 10.1007/s11235-015-0037-2

Fukuda, K., Asai, H., & Nagami, K. (2015). Tracking the evolution and diversity in network usage of smartphones. In *Proceedings of the 2015 internet measurement* *conference* (pp. 253–266). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2815675.2815697 doi: 10.1145/2815675.2815697

Garcia, M. ., De Simone, F., Tavakoli, S., Staelens, N., Egger, S., Brunnström, K., & Raake, A. (2014, Sep.). Quality of experience and http adaptive streaming: A review of subjective studies. In *2014 sixth international workshop on quality of multimedia experience (qomex)* (p. 141-146). doi: 10.1109/QoMEX.2014.6982310

He, Z., & Xiong, H. (2006, Sep.). Transmission distortion analysis for real-time video encoding and streaming over wireless networks. *IEEE Transactions on Circuits and Systems for Video Technology*, *16*(9), 1051-1062. doi: 10.1109/TCSVT.2006.881198

Index, C. V. N. (2019). global mobile data traffic forecast update, 2017–2022. *Cisco White paper*.

ITU-T RECOMMENDATION, H. (2017). *264 advanced video coding for generic audiovisual services.* Retrieved from https://www.itu.int/rec/T-REC-H.264-201704-S/en

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, *6*(5), 429–449.

Joskowicz, J., & Sotelo, R. (2014). A model for video quality assessment considering packet loss for broadcast digital television coded in h.264. *International Journal of Digital Multimedia Broadcasting*, *2014*. Retrieved from https://doi.org/10.1155/2014/242531

Joumblatt, D., Teixeira, R., Chandrashekar, J., & Taft, N. (2011). Performance of networked applications: The challenges in capturing the user's perception. In *Proceedings of the first acm sigcomm workshop on measurements up the stack* (pp. 37–42). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2018602.2018612 doi: 10.1145/2018602.2018612

Korhonen, J. (2018, June). Study of the subjective visibility of packet loss artifacts in decoded video sequences. *IEEE Transactions on Broadcasting*, *64*(2), 354-366. doi: 10.1109/TBC.2018.2832465

Krishnan, S. S., & Sitaraman, R. K. (2013, Dec). Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs. *IEEE/ACM Transactions on Networking*, *21*(6), 2001-2014. doi: 10.1109/TNET.2013.2281542

Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling* (Vol. 26). Springer.

Laird, J. E. (2012). *The soar cognitive architecture.* MIT press.

Liang, Y. J., Apostolopoulos, J. G., & Girod, B. (2008, July). Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(7), 861-874. doi: 10.1109/TCSVT.2008.923139

Liu, Y., Dey, S., Ulupinar, F., Luby, M., & Mao, Y. (2015, Dec). Deriving and validating user experience model for dash video streaming. *IEEE Transactions on Broadcasting*, *61*(4), 651-665. doi: 10.1109/TBC.2015.2460611

Mathis, M., Semke, J., Mahdavi, J., & Ott, T. (1997, July). The macroscopic behavior of the tcp congestion avoidance algorithm. *SIGCOMM Comput. Commun. Rev.*, *27*(3), 67–82. Retrieved from http://doi.acm.org/10.1145/263932.264023 doi: 10.1145/263932.264023

Mok, R. K., Chan, E. W., Luo, X., & Chang, R. K. (2011). Inferring the qoe of http video streaming from user-viewing activities. In *Proceedings of the first acm sigcomm workshop on measurements up the stack* (pp. 31–36). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2018602.2018611 doi: 10.1145/2018602.2018611

Moorthy, A. K., Choi, L. K., Bovik, A. C., & de Veciana, G. (2012, Oct). Video quality assessment on mobile devices: Subjective, behavioral and objective studies. *IEEE Journal of Selected Topics in Signal Processing*, *6*(6), 652-671. doi: 10.1109/JSTSP.2012.2212417

Moorthy, A. K., Seshadrinathan, K., Soundararajan, R., & Bovik, A. C. (2010, April). Wireless video quality assessment: A study of subjective scores and objective algorithms. *IEEE Transactions on Circuits and Systems for Video Technology*, *20*(4), 587-599. doi: 10.1109/TCSVT.2010.2041829

Nasralla, M. M., Hewage, C. T. E. R., & Martini, M. G. (2014, July). Subjective and objective evaluation and packet loss modeling for 3d video transmission over lte networks. In *2014 international conference on telecommunications and multimedia (temu)* (p. 254-259). doi: 10.1109/TEMU.2014.6917770

Ooyala. (2018). Ooyala global video index q2 2018. *Ooyala white paper*.

P.910, I.-T. (2008). Subjective video quality assessment methods for multimedia applications. *ITU-T RECOMMENDATION, P.910*. Retrieved from `https://www.itu.int/rec/T-REC-P.910-200804-I`

Paudyal, P., Battisti, F., & Carli, M. (2016, Dec 01). Impact of video content and transmission impairments on quality of experience. *Multimedia Tools and Applications*, *75*(23), 16461–16485. Retrieved from `https://doi.org/10.1007/s11042-015-3214-0` doi: 10.1007/s11042-015-3214-0

Pinson, M. H., & Wolf, S. (2003). Comparing subjective video quality testing methodologies. In *Visual communications and image processing 2003* (pp. 573–582). Retrieved from `https://doi.org/10.1117/12.509908`

Reibman, A. R., & Poole, D. (2007, Sep.). Characterizing packet-loss impairments in compressed video. In *2007 ieee international conference on image processing* (Vol. 5, p. V - 77-V - 80). doi: 10.1109/ICIP.2007.4379769

Rosenbloom, P. S., Laird, J. E., Newell, A., & McCarl, R. (1991). A preliminary analysis of the soar architecture as a basis for general intelligence. *Artificial Intelligence*, *47*(1), 289 - 325. Retrieved from `http://www.sciencedirect.com/science/article/pii/000437029190057Q` doi: https://doi.org/10.1016/0004-3702(91)90057-Q

Rosenthal, R., & Rosnow, R. L. (2009). *Artifacts in behavioral research: Robert rosenthal and ralph l. rosnow's classic books*. New York, NY, USA: Oxford University Press.

Rui, H.-x., Li, C.-r., & Qiu, S.-k. (2006, Jan 01). Evaluation of packet loss impairment on streaming video. *Journal of Zhejiang University-SCIENCE A*, *7*(1), 131–136. Retrieved from `https://doi.org/10.1631/jzus.2006.AS0131` doi: 10.1631/jzus.2006.AS0131

Shafiq, M. Z., Erman, J., Ji, L., Liu, A. X., Pang, J., & Wang, J. (2014). Understanding the impact of network dynamics on mobile video user engagement. In *The 2014 acm international conference on measurement and modeling of computer systems* (pp. 367–379). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2591971.2591975` doi: 10.1145/2591971.2591975

Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., Venkataraman, S., & Wang, J. (2013). A first look at cellular network performance during crowded events. In *Proceedings of* *the acm sigmetrics/international conference on measurement and modeling of computer systems* (pp. 17–28). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2465529.2465754 doi: 10.1145/2465529.2465754

Smith, S. W., et al. (1997). The scientist and engineer's guide to digital signal processing.

Staelens, N., De Meulenaere, J., Claeys, M., Van Wallendael, G., Van den Broeck, W., De Cock, J., ... De Turck, F. (2014, Dec). Subjective quality assessment of longer duration video sequences delivered over http adaptive streaming to tablet devices. *IEEE Transactions on Broadcasting*, *60*(4), 707-714. doi: 10.1109/TBC.2014.2359255

Stauss, B., Schmidt, M., & Schoeler, A. (2005-07). Customer frustration in loyalty programs. *International Journal of Service Industry Management*, *16*(3), 229,252.

Streijl, R. C., Winkler, S., & Hands, D. S. (2016, Mar 01). Mean opinion score (mos) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*, *22*(2), 213–227. Retrieved from https://doi.org/10.1007/s00530-014-0446-1 doi: 10.1007/s00530-014-0446-1

Tin Kam Ho. (1995, Aug). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, p. 278-282 vol.1). doi: 10.1109/ICDAR.1995.598994

Trestian, R., Vien, Q., Nguyen, H. X., & Gemikonakli, O. (2015, June). On the impact of video content type on the mobile video quality assessment and energy consumption. In *2015 ieee international symposium on broadband multimedia systems and broadcasting* (p. 1-6). doi: 10.1109/BMSB.2015.7177206

Venkataraman, M., & Chatterjee, M. (2012, April). Quantifying video-qoe degradations of internet links. *IEEE/ACM Transactions on Networking*, *20*(2), 396-407. doi: 10.1109/TNET.2011.2167684

Venkataraman, M., Chatterjee, M., & Chattopadhyay, S. (2009, Nov). Evaluating quality of experience for streaming video in real time. In *Globecom 2009 - 2009 ieee global telecommunications conference* (p. 1-6). doi: 10.1109/GLOCOM.2009.5425806

Video.js: The player framework. (2018). Retrieved from http://videojs.com

Videolan media player. (2018). Retrieved from `http://www.videolan.org`

Wei, X., Valler, N. C., Madhyastha, H. V., Neamtiu, I., & Faloutsos, M. (2015, April). A behavior-aware profiling of handheld devices. In *2015 ieee conference on computer communications (infocom)* (p. 846-854). doi: 10.1109/INFOCOM.2015.7218455

William, G. (2012). *Econometric analysis.* Boston, MA: Pearson Education.

Winkler, S., & Campos, R. (2003). Video quality evaluation for internet streaming applications. In *Human vision and electronic imaging viii* (Vol. 5007, pp. 104–116). Retrieved from `https://doi.org/10.1117/12.477766`

Wireshark network protocol analyzer. (2018). Retrieved from `https://www.wireshark.org`

APPENDIX

APPENDIX A

VIDEO QUALITY ASSESSMENT STUDY EXIT SURVEY

1. Please provide your full name

2. Please provide your email address

3. I have paid attention to the videos and tried my best throughout the experiment.

    - Completely Disagree
    - Somewhat Disagree
    - Neither Agree or Disagree
    - Somewhat Agree
    - Completely Agree

4. I am confident with my rating performance in the experiment.

    - Completely Disagree
    - Somewhat Disagree
    - Neither Agree or Disagree
    - Somewhat Agree
    - Completely Agree

5. How many hours do you spend on watching online videos per week?

6. How many hours have you slept the night before?

7. Do you need to wear glasses or contact when watching videos? If yes, did you wear them in the experiment?

8. Have you seen any of the videos used in the experiment before? If so, please identify which ones.

    - No
    - Movie Trailer
    - Basketball Highlights
    - News Interview
    - Doodle Cartoon

9. How often do you make careless mistakes when you have to work on boring or difficult projects?

    - Never
    - Rarely
    - Sometimes
    - Often
    - Very Often

10. How often do you have difficulty keeping your attention when you are doing boring or repetitive work?

- Never
- Rarely
- Sometimes
- Often
- Very Often

APPENDIX B

USER VIDEO VIEWING BEHAVIOR STUDY EXIT SURVEY

1. Please provide your full name

2. Please provide your email address

3. I have paid attention to the videos and tried my best throughout the experiment.

    - Completely Disagree
    - Somewhat Disagree
    - Neither Agree or Disagree
    - Somewhat Agree
    - Completely Agree

4. What is your entire watching experience during the experiment?

    - Bad
    - Poor
    - Fair
    - Good
    - Excellent

5. The video quality and my watching experience get improved when I have interacted with the experiment interface.
    - Completely Disagree
    - Somewhat Disagree
    - Neither Agree or Disagree
    - Somewhat Agree
    - Completely Agree
6. How many hours do you spend on watching online videos per week?
7. How many hours have you slept the night before?
8. Do you need to wear glasses or contact when watching videos? If yes, did you wear them in the experiment?
9. Have you seen any of the videos used in the experiment before? If so, please identify which ones.
    - No
    - Movie Trailer
    - Basketball Highlights
    - News Interview
    - Doodle Cartoon
10. Have you realized the real purpose of this experiment before debriefing? If yes, please describe.

- No
- Yes