The Effects of Latency on Player Performance and Experience in a Cloud Gaming System

Robert Dabrowski, Christian Manuel, Robert Smieja

May 7, 2014

An Interactive Qualifying Project Report:
submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the
Degree of Bachelor of Science

Approved by:

Professor Mark Claypool, Advisor
Professor David Finkel, Advisor

Abstract

Due to the increasing popularity of thin client systems for gaming, it is important to understand the effects of different network conditions on users. This paper describes our experiments to determine the effects of latency on player performance and quality of experience (QoE). For our experiments, we collected player scores and subjective ratings from users as they played short game sessions with different amounts of additional latency. We found that QoE ratings and player scores decrease linearly as latency is added. For every 100 ms of added latency, players reduced their QoE ratings by 14% on average. This information may provide insight to game designers and network engineers on how latency affects the users, allowing them to optimize their systems while understanding the effects on their clients. This experiment design should also prove useful to thin client researchers looking to conduct user studies while controlling not only latency, but also other network conditions like packet loss.

Contents

List of Figures

List of Tables

Listings

1 Introduction

As video games become more and more popular, the demand for them increases as well. Distribution plays an important part in the success of video games, with games coming in a multitude of mediums such as disks, cartridges, or simply as data on a computer or handheld device. An important decision for game designers is to choose which medium to distribute their game on. Many modern games can exist on many separate mediums, while being exported to others. When producing a game on more mediums, there is also a higher chance it being illegally copied and downloaded. Cloud gaming can help to solve these problems, as well as make multiplayer game development cheaper and easier.

Development costs aside, some consumers struggle to purchase all of the latest games and the necessary console systems required to play them. Others just can not afford it. PC gaming offers a somewhat cheaper solution by allowing players to have access to a much higher number of games without needing the console hardware. It is also easier for consumers to obtain games, as many of them are just data on the computer and can be downloaded through the Internet. This also solves the problem of trying to keep up with a large number of physical copies of games. Although, the price for new games is often the same whether it is for the PC or console. Some users do not want to pay the full price of a game that they will complete within one or two days time. Cloud gaming is a solution to these problems as well.

The idea of cloud gaming was introduced to the public in 2009, with OnLive announcing their services at the 2009 Game Developers Conference [1]. The idea was to provide a service to gamers that would allow them to access and control games over the Internet, without needing the required hardware in order to play them. This idea manifested itself in the form of OnLive's "thin client" application for PC's or the OnLive home console. For a small monthly fee, the application allowed gamers access to a variety of games, new and old. The users would be able to select a game from a list and then stream it to their home over the Internet. The input from the player would be sent to OnLive's servers, which would run the necessary computations, then send the resulting video back to the player's screen. The application solved the problem of high game prices as well as physical and digital game storage all in one. The advantages to having a thin client is that they can be significantly cheaper for each individual client, have great scalability and lower hardware requirements because the processing can now be done through the network, and multiple clients are able utilize the same server for processing.

Before cloud gaming, networked games could only be played through "fat clients". A "fat client" is defined as the client in a client-server architecture that provides the majority of the functionality and computations independent of the server. A "thin client" on the other hand relies entirely on the server to provide functionally and to process computations. While both a fat client and a thin client require a network connection, a fat client is able to preform many functions independent of a network connection, while a thin client is entirely dependent on the network connection.

OnLive's thin client technology works well with the assumption that users have a decent Internet connection. Even so, unforeseen problems with connectivity can happen, and data flow between the client and the servers can be disrupted. This kind of disruption is known as "latency". Latency occurs when there is either a delay in information to the client or to the server. These disruptions or delays can cause many other problems such as frame rate loss, loss of control precision, and lag. Issues like this are not unique to OnLive's thin client application and can occur in many thin client applications. Additional types of network conditions that can have an effect on thin clients are documented in greater detail in Section 2.1.1.

As the usage of thin clients becomes more prevalent, especially in their usage for video games, which are applications that have extreme sensitivity to latency as well as a high definition media requirement, it is important to understand how network conditions impact a user's experience.

In order to study the effects of latency on players using a cloud gaming system, we researched other experiments involving video games. Some involved thin clients, so we designed an experiment similar to those. We decided to control only latency while participants play a game, with multiple trials at various latencies. The cloud gaming service we used was OnLive and the software for introducing latency was DummyNet.

Before we started our experiment the we went through the process of being reviewed by the Institutional Review Board (IRB) at Worcester Polytechnic Institute because we planned on having human participants. We also went through a similar procedure to use the Psychology Department Participant Pool, which we used to gather additional participants. After that, we began advertising through email, posters and word of mouth for people to sign up to participate on our sign up sheet on Slottr.

For the experiment, we had volunteers come in to the testing environment that we set up in a lab on campus. The participants played short sessions of Crazy Taxi via OnLive while we controlled how much additional latency was introduced. Before the participants played, we had them fill out a survey with some demographic information so that we could know what population we were examining. Next we had participants alternate between playing and filling out short surveys about the session they just played. The questions asked for the participants' opinions on how responsive they thought the controls were or how much they enjoyed that particular session. We also recorded player scores for each session.

By the end our experiment, we gathered 49 participants, which contained a fairly even distribution of "gamers" and "non-gamers". This allowed us to evaluate our target demographic as well as compare against the opposite group. Through our experiment, we found that latency has a negative linear effect on player performance and satisfaction. We found that for every 100 ms of added latency, players reduced their Qualitfy of Experience (QoE) ratings by 14% on average, meaning they were less satisfied and enjoyed the game less. Participants also noticed that the responsiveness of the controls changed the most and the graphics quality changed the least across all trials. Our experiment design can also be duplicated for other game types, different amounts of latencies, or used to control other variables in network quality like packet loss.

Section 2, Background Research, describes the research we did to prepare for this project. Included are references to experiments studying cloud gaming, video games or both. Section 3, Hypotheses, introduces the hypotheses we decided to test for based on the research we had completed. Section 4, Methodology, details every part of our experiment. Section 5, Results and Analyses, reports our results and analysis of the data we collected. Upon completion of our experiment, we reflected on the design and provide a discussion on changes and expansions to our experiment in Section 6, Future Studies. Finally, Section 7, Conclusion, presents our conclusions both on the data we collected and the design of our experiment. Appendices A and B are included to show additional data and graphs not included in Section 5. Appendix C is included to show the documents and forms we worked with during the experimental phase of our project.

2 Background Research

In preparation for our experiment, we read many papers about thin client technologies and studies that had been done either with video games, thin client technologies, or both. Section 2.1 details parameters used to control clients by as well as studies that did so. Section 2.2 is about other studies that conducted studies around video games, as well as how those studies did so. Finally, Section 2.3 is about a metric used to classify video games and thus lend some context to our study.

2.1 Thin Client Technology

As thin-clients become more common, the amount of research done about their behavior surrounding video games will increase. Looking at previous studies, we were able to find useful information about how to successfully control variables, which variables needed to be controlled, and what cloud services were utilized in those studies. Section 2.1.1 discusses the various network variables which can affect cloud games and how previous studies have investigated them and controlled them. Section 2.1.2 discusses various cloud gaming services that were covered used by the studies mentioned in this section.

2.1.1 Control Methods and Parameters

Each study that we looked at identified and attempted to account for or manipulate various aspects of network conditions in order to study how they impacted cloud gaming. These aspects are:

- Delay or Latency - Delay, also known as latency, is the measurement of the amount of time needed to send a packet from the thin-client to the server, or from the server to the thin-client. Delay can affect input from the user and output, such as video and audio, from the server. [2]

- Jitter - Jitter is defined as the amount of change over a given period of time. In this context, jitter refers to a change in the network latency over time. [2]

- Bandwidth - Bandwidth is the measure of the amount of bits able to be transmitted at any given moment, and is typically measured in bits per second (bps). [3]

- Packet Loss - Packet loss is the amount of packets that failed to be delivered during a transmission between the thin client and server, or vice versa. Packet loss is typically measured as a percent against packets successfully delivered. I.E. 5 percent packet corruption means 5 percent of packets have not been delivered. [2]

- Packet Corruption - Packet corruption is the amount of packets that were delivered with incorrect and unusable data during a transmission between the thin client and server, or vice versa. Packet corruption is typically measured as a percent against packets successfully delivered, the same as packet loss. [3]

- Packet Re-ordering - Packet re-ordering is the delivery of packets outside of their intended ordered, and is typically measured in percentages. For example, the last packet might arrive before the first packet. [2]

- Packet Duplication - Packet duplication is the delivery of the same packet multiple times, and is typically measured in a percentage. [2]

- Packet Size - Packet size can be used to restrict the size of each individual communication sent between the thin-client and the server, and is measured in bytes. [3]

From an end user perspective, it has been argued that the source of degradation of network conditions is not important, as the end result of the degradation of the cloud gaming service occurs regardless of the source. In the study by Jarschel, Schlosser, Scheuring and Hofeld, they studied changes in both delay and packet loss to arrive at this conclusion, while controlling jitter[2]. In the study *Thin to Win? Network Performance Analysis of the OnLive Thin Client Game System* by Claypool, Finkel, Grant and Solano, they study the effects of altering bandwidth, packet loss and packet size[3]. We have not encountered any study in our background research that investigates packet re-ordering, packet corruption or packet-duplication specifically.

Different studies have had different approaches to the issue of normalizing network conditions, which can vary from avoiding the usage of an online cloud gaming service by creating a local cloud game server, to changing existing network conditions and accounting for natural latency. The methods used in these studies are as follows:

- DummyNet - DummyNet is a network emulation tool that was originally created for FreeBSD, and has been modified to run on Linux, Mac OS X, Windows, and OpenWRT for routers. It is able to add delay, jitter, packet loss, packet size and bandwidth manipulation. [3]

- NetEM - NetEM is a network emulation tool that was created for Linux, and is included by default in Linux kernels version 2.6 or newer. NetEM is capable of creating packet loss, packet corruption, packet duplication, packet re-ordering, delay, bandwidth manipulation, packet size and jitter. [2]
- Keyboard and Mouse Delay - One study that we encountered a Windows application that used the native Windows API to capture and delay all keyboard and mouse inputs by a fixed amount, which emulated the natural delay in an online cloud game. [4]

2.1.2 Services

Different studies approached cloud gaming differently and attempted to study it using a wide variety of different services. SpawnBox was a device that attempted to take a standard gaming console and enable it for the cloud. The SpawnBox connected to the video output of game consoles and to the network via an Ethernet cable, and returned the input from the network back to the game console through USB. Unfortunately, the SpawnBox is no longer available for sale, due to GameStop acquiring SpawnLabs and later shutting that division of GameStop down [2, 5]. OnLive is a cloud gaming service that offers a variety of games for purchase, or as part of a monthly subscription called the "OnLive Play Pack", that are then able to be played on the OnLive cloud servers from computers, tablets, or the custom OnLive micro-console [3]. A previous study ran a BitTorrent client through Amazon's EC2 cloud service in an attempt to measure the cloud platforms latencies to various end hosts, and to determine the quality of a potential cloud game from the latencies obtained [6].

2.2 User Studies Involving Video Games

While preparing to conduct our experiment, we looked into some previous studies involving human participants. Some experiments were very close to ours in that they studied the effects of latency on users of cloud gaming services, while others provided a lot of insight into how to set up for test involving users and video games. Sections 2.2.1 and 2.2.2 elaborate on some experiments that collected information from users directly, while Section 2.2.3 delves into studies that collected other data like player scores. Finally Section 2.2.4 explores how others have modified games to prepare them for human or artificial intelligence test subjects.

2.2.1 Demographic Data

When performing human user studies, it is often necessary to collect demographic information to give context to data that is collected from various groups of people. It is also important to verify that at least the target audience is well represented. For example, a study involving the enjoyability of a first person shooter game wouldn't mean much if the participants hate video games to begin with.

To ensure that we knew what groups our users belonged to, we created our own set of demographics questions, detailed in Section 4.7.2, with borrowed questions from studies like [2, 7, 8, 9, 10]. The two most common questions ask age and gender, which are commonly asked because the most common demographic for video game players is teenage males [2, 7]. The remaining questions try to determine the skill level of participants in whatever the particular game type of interest. Claypool and Claypool [8, 9] ask the following questions:

- "How many hours per week do you play computer games?"
- "How you would classify yourself as a computer gamer?"
- "How would you evaluate your performance in first person shooters?"

These questions did not yield significant groupings and thus were not reported on extensively. Jarschel et al. asked similar questions, after modifications, in [2, 7] and determined that their users play games at least sometimes and are thus within their target demographics. Studies like [10] did not include their questions and only provided generalized information about their users, indicating that they did not ask the correct questions. Our target demographics will be very similar to these studies and we hope to use our demographic information to confirm that.

2.2.2 Quality of Experience

The studies most similar to ours involved collecting subjective data directly from their users. The easiest subjective questions ask the users how much they enjoyed a particular gaming session. Jarschel et al. in [2, 7] used those questions to calculate a Mean Opinion Score (MOS) to represent the users' Quality of Experience (QoE). In their study they asked users for their rating of their current game experience, then mapped those ratings to a scale of 1 to 5. Their questioning process was very crude because it was up to the participants to weight different aspects as they saw fit in their final score; different users would give different scores depending on whether they felt responsiveness or image quality were more important. Clincy et al. in [10] used the following categories as questions:

- "Loading Time"
- "Responsiveness"
- "Image Quality"
- "Sound Quality"
- "Choppiness/Stuttering"
- "Game Freezes/Disconnects"
- "Overall Experience"
- "Likelihood to play again"

While not all of these questions apply to our study, this set provides a much more accurate QoE rating than previous studies because the various categories can be weighted independently of the users' overall preferences. With this data, [2, 7, 10] did the same thing and combined all scores into one value and reported how that value changed between changing latency and packet loss conditions. Each study left a lot open in terms of conducting tests at more latencies as well as asking more appropriate questions to figure out what the user thought.

2.2.3 Player Scores

To objectively determine the effects of latency and other conditions on players, most studies collect player scores. It is a simple feat to display average player scores per game condition, as Claypool and Claypool did in [8, 9, 11, 12] and Beigbeder et al. did in [13]. Player scores are very valuable because they have the potential to be unbiased and accurate as long as participants put the same amount of effort into each trial. With this data, Claypool and Claypool successfully showed that player performance got worse with lowering frame rates in [8]. Concrete data such as this is very compelling, and will complement our subjective data quite well.

Of course, the player score data will not collect itself. Most of the previously mentioned papers did not explicitly state their methods for collecting player scores, but it could be as simple as someone writing down scores as participants finished rounds. To improve efficiency, Claypool and Claypool in [8, 9] use a program called FRAPS, which is a Windows application designed to take screen shots of games and record times automatically. Using this program, they were able to easily collect all kinds of data with minimal effort or interference. Automation in data collecting is very advantageous.

2.2.4 Game Modified for User Studies

For the purposes of studying the effects of changing game conditions it is very advantageous to have a fully controlled game environment. For example, it would be bad to have a user get into a car and drive around when you want them to experience the first person shooter aspect of the game.

Claypool and Claypool created a very well controlled in-game environment for their users in [8, 9]. Using *Quake 3 Arena*, they built custom maps to test two very specific aspects of first person shooter games: shooting and moving. For the moving aspect, they made a path for the user to run down, and made it so the player would lose a discrete amount of life each time they fell off the path. This allowed Claypool and Claypool to collect score and remaining health data that very precisely told them how the user performed on that map under whatever other conditions they chose. They created shooting scenarios with similar constraints that allowed the user to focus on just one shooting or just moving. Since the only data they were interested in was performance, they created the perfect video game environment for their study.

In [2, 7], Jarschel et al. use existing games like Final Fantasy and set up the user to face only specific parts of the game. They did not actually modify the games they were using for two key reasons; they did not need player scores and they did not want to change the normal enjoyability of the games they used. Sometimes there is no need to actually change the game, but they did control which parts of the game the user would play.

Some really interesting scenarios can be created for different kinds of users. In [11], Claypool and Claypool took a third person perspective tank driving game and set it up so that the game's artificially intelligent "bots" would play all the game sessions for them. They tested the effects of latency on the computer players by providing one with latency and letting the other play on its own, the whole time recording scores and other statistics. This is a very interesting case because all aspects, from participants to game conditions, were under the control of those conducting the experiment.

2.3 Precision vs. Deadline Classification of Games

There are numerous network variables to take into account when measuring player performance in a game.

1. Frame Rate
2. Precision
3. Deadline

Frame rate, or frames per second (FPS), in the context of games, is the rate at which the image produced by the game is refreshed. Precision refers to the amount of accuracy needed between the player and the game to successfully complete their interaction. Precision can relate to how responsive the controls between the player and the game are. Deadlines refer to the time it takes to complete an action. This is the time needed to achieve the final outcome of that action.

In [9], analysis of the performance results of the participants shows that for the ranges tested, the frame rate had a larger impact on player performance than found for streaming video or other interactive multimedia applications. As shown in [11], actions that have a high precision degree and tight deadlines are sensitive to moderate latencies, while actions that have low precision and deadlines are resilient to high internet latencies.

Since this study involves determining the effect of latency on player performance, it is important to study how frame rate, precision degree, and deadline times are affected. Originally, it was hypothesized that the frame rate would not be affected by changes in latency. Many initial reviews of the OnLive system described high frame rates [14]. Upon completion of our study, however, we noted that the frame rate was dramatically affected for the highest latency in a few of the tests. While the frame rate loss was not consistent across all the tests, this notably affected player performance. Frame rate loss due to high latency is often seen in resource intensive games that have high precision and deadline times and make use of networking capabilities. A first person shooter, for example, matches this criteria.

For a given game action, the higher the precision required the greater the impact of latency on performance. As well as that, the tighter the deadline the greater the impact of latency on performance [11]. Since the precision-deadline model can be directly correlated to player performance, our study specifically chose a game with a medium-to-high precision degree and a tight deadline for game action. The rate at which frames are rendered in a computer game directly impacts player performance, inuencing both the game playability and enjoyability [9].

3 Hypotheses

After researching factors that lead into a player's performance, described briefly in Section 2, as well as some intuition from playing video games, we made the following hypotheses all relating to thin client gaming:

1. As latency increases, player performance will decrease.

2. As latency increases, player satisfaction will decrease.

3. There is a range of best-case latencies that a user will not be able to differentiate between.

4. There is a point where any further decreases in latency will cause a user to consider a game unplayable.

While these four statements seem completely independent, a proper test setup can figure out each of them at the same time as well as provide for additional predictions. For example, a fifth hypothesis may be extrapolated from 1 and 2 to say that player performance and satisfaction will increase and decrease at similar rates with different network conditions. An alternate version of our hypothesis could be the following:

- Proportional relations can be made between player performance and satisfaction and added latency in thin client games, except at the extremes where the user either does not notice changes or the game is unplayable.

In order to show these hypothetical characteristics, we designed an experiment, detailed in Section 4, to have users play a game at many latencies and record their scores as well as their ratings of different aspects of the game conditions. We predict that our data will look like Figure 1 once graphed. The x-axis shows the control case and latencies from "normal network" conditions and decrease in quality from there. This graph also shows player score and satisfaction increasing along the y-axis. They are shown together here because our hypothesis is that they both change at similar rates as the network quality decreases. The labels on the y-axis go from "unplayable" to "fun to play" and are mainly meant to show player satisfaction. However, the level of "unplayable" corresponds to players getting scores of 0 or uncontrollably failing levels.

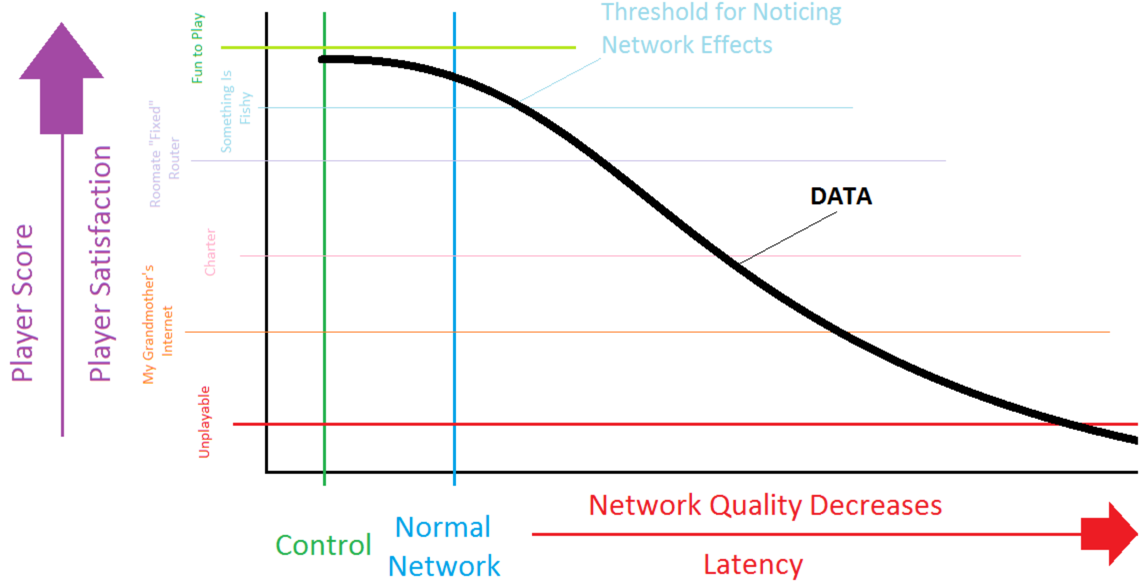

Figure 1: Expected graph of player satisfaction and score vs. network quality with respect to latency

This graph shows the network conditions that players begin to notice latency and the network conditions that render a game unplayable. A particular network quality will be unplayable where the data line intersects the "unplayable" level of the y-axis. The threshold for noticing changes in latency will be harder to determine as it will probably take more data points to actually find when a user notices a change. This is why the level of satisfaction is named "Something is Fishy", because the user might not be 100% sure that something has changed. Ultimately, this graph can be used to predict the effects of added latency on player performance and satisfaction. While the values we plot will be specific to a single game, our experiment can be duplicated with almost any game and produce comparable results.

4 Methodology

In order to test the effects of latency on users while playing games on a thin client system, we designed a full test station, procedure and data collection process using OnLive as our thin-client game provider. Section 4.1 describes how we chose a game to use in our study. The hardware, software and environment used in our study are described in Section 4.2. Sections 4.3 to 4.6 elaborate on how we got approved to and handled having human participants in our study. Section 4.7 shows the questions we asked our participants. Finally, Section 4.9 explains how we conducted our experiment and Section 4.8 describes the latency settings we used during the experiment.

4.1 Game Selection Process

The following criteria, listed in order of importance, was used when determining what game would be used in the study.

1. A new player can sit down and play it immediately
2. OnLive provides it
3. Scores are provided
4. Short challenge levels of varying difficulty are available
5. It is fun
6. It requires higher precision and tighter deadlines than the next choice
7. Gives high response time at little to no latency
8. It can be played with a game controller instead of mouse/keyboard
9. Test subjects are familiar with it
10. Test subjects have played it before

One of the major determining factors when deciding the game to be used was responsiveness to user input. At little to no latencies, some of the games did not respond very well to user input and it would become even worse when higher latencies are applied. Another determining factor was how resource intensive the game was, as well as the quality of the graphics. Participants are questioned about the quality of the graphics during play, and if he or she notices any change in quality. For this reason, it was important to choose a game with good graphic quality. The game that was chosen also needed to be easy to pick for new players or beginners. The desired demographic of participants included those who spent little to no time playing games, and it is important for them to be able to understand the game mechanics to provide meaningful survey data.

To help measure player performance during the study, it was important for the game chosen to provide some of point or score system that gauges how well the player performed. This could be an arbitrary point system or a time limit. To make sure that the experiment lasted to within 35 minutes, each of the 10 game session needed to last from 35-60 seconds.

4.1.1 Other Games Considered

There was a variety of games considered for this study, each meeting most of our criteria.

1. Batman Arkham City
2. Darksiders/Darksiders II
3. Sonic The Hedgehog 2 or 3
4. Unreal Tournament 3
5. Syder Arcade
6. 4 Elements II

Many of these games, however, had a skill curve that was too high. The player's would not be able to fully grasp the skills needed to perform well in the time allocated for the study. Other games, like Sonic The Hedgehog 2, did not give high enough response time at little to no latency to be considered for the study.

4.1.2 Crazy Taxi

Crazy Taxi was chosen because it met most , if not all, of our criteria.

1. Its simplicity makes it easy for new players to quickly pick up
2. It's provided by OnLive
3. It provides a scoring system to measure performance

4. It requires high precision of controls to perform well
5. It gives high response time at little to no latency
6. It can be played with a game controller
7. Test subjects are familiar with it
8. Test subjects have played it before

The main objective of the game is to pick up customers and take them to their chosen destination as quickly as possible. Along the way, players can earn extra money by performing stunts such as near misses with other vehicles or jumping over ramps. The player is directed to a destination by a large green arrow at the top of the screen. When the player arrives at the destination, they must stop within a specified zone. When the destination is reached, that customer's fare is added to the total money the player earned. Customer's fare vary on the distance to their destination. Some destinations are further from customers than others. Each customer has a timer to indicate how long the player has to reach the destination. If the customer's timer runs out before the player reaches the destination, the customer jumps from the taxi and no fare is collected.

When customers are delivered, time is added to the overall game play time. To optimize the time each session took, we changed the game options to increase the traffic difficulty and game the player 35 seconds of start time at the beginning of each round. This would ensure each player did not take too long per session, but was still able to reach a measurable score. The game play was consistent as well, as the player always started in the same position and the customers went to the same destination every time.

4.2 Test Station

Our test station setup is divided into two parts including hardware and software described in Sections 4.2.1 and 4.2.2 respectively.

4.2.1 Hardware

For the computer acting as the gateway in our test setup, we had the following materials:

- MacBook Pro, 17-inch Early 2011, (8,3), with the following specifications:
  - Mac OS X 10.9.2

  - Intel Quad Core i7 2.3 GHz
  - 4 GB DDR3 RAM
  - AMD Radeon HD 6750M
  - Intel HD Graphics 3000
  - Broadcom 57765-B0 Gigabit Ethernet Controller
  - Hitachi HTS725050A9A362 500 GB Hard Drive, 7200 RPM

- OnLive Micro-console
  - OnLive Micro-console controller
- iPad 2
- Nexus 7 (2013)
- Two standard Cat 5E 10/100 Ethernet cables
- UTechSmart USB 2.0 to 10/100 Wired Ethernet Adapter

The WPI network is connected via a Gigabit Ethernet link to the Internet2 network. Internet2 is a high-speed research network that allows WPI to collaborate with other schools, government agencies, and corporations. WPI maintains and operates the Goddard GigaPoP for Internet2, which allows WPI to serve as a connector for other schools and businesses. [15]

The MacBook Pro was the machine that had DummyNet, which was covered in Section 2.1.1, and acted as the "Gateway" between the OnLive micro-console and the Internet and was where the added latency was placed. The OnLive micro-console was used as the OnLive client that connected the OnLive game servers to send input from the OnLive controller, and receive the video output. We provided both the iPad 2 and the Nexus 7 tablet computers to participants to allow them to quickly fill out the Qualtrics survey in between each trial. The Ethernet cables and USB to Wired Ethernet adaptor were used to setup the network connections as shown in Figure 2.

4.2.2 Software

The testing software setup includes:

- OnLive

- DummyNet, which is included as part of the Mac OS X kernel, (October 28, 2002)
- Crazy Taxi for the PC

OnLive was the cloud gaming service provider that we chose for this study, due to having a previous study [3], being done on OnLive. DummyNet is a well documented and widely used network emulation software which has been used for various network related studies in the past. DummyNet allows manipulating varying aspects of network traffic, as described in Section 2.1.1 Using the criteria described in Section 4.1.2, we decided on using Crazy Taxi for our experiment and used the PC version provided through OnLive.

4.2.3 Full Setup

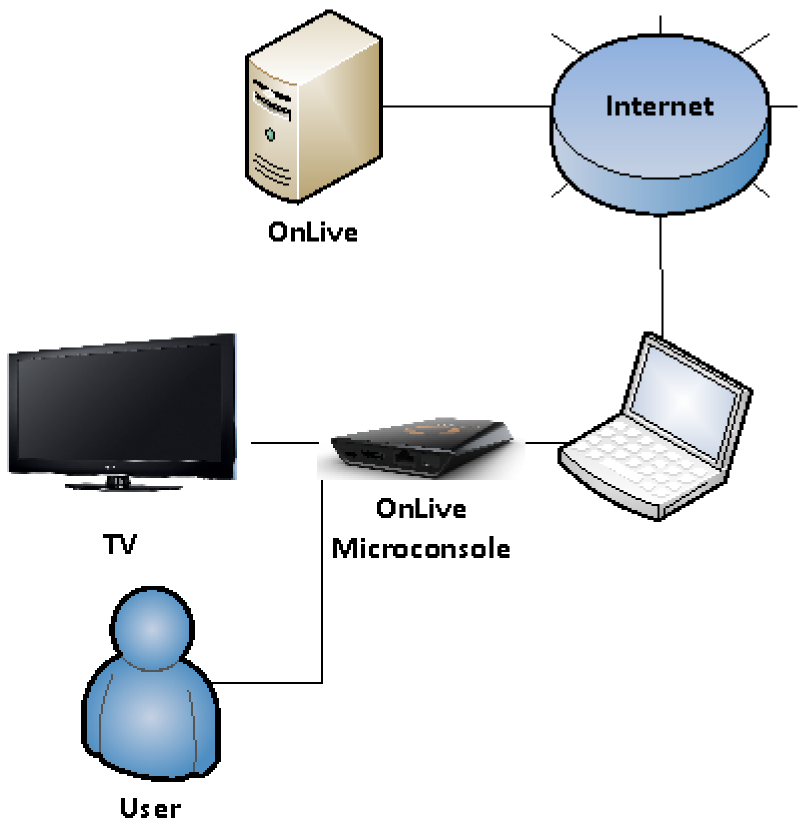

Figure 2: A diagram representing the setup we used.

The setup consisted of one Ethernet cable plugged into an RJ45 wall jack that was located in the room. That Ethernet cable was then connected to the USB Ethernet Adaptor which was attached to the MacBook Pro's integrated Ethernet adaptor. The remaining Ethernet cable was plugged into the OnLive Micro Console and the USB Ethernet Adaptor. The USB Ethernet Adaptor was attached to the MacBook Pro to complete the setup.

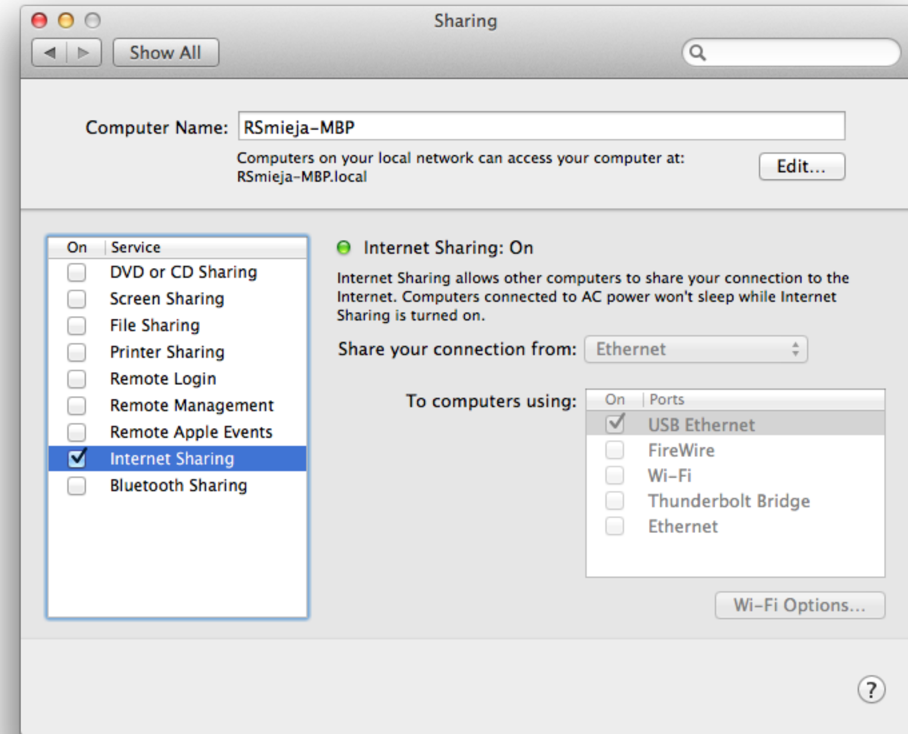

Figure 3: The settings area in Mac OS X that needs to be configured to allow the computer to act as a gateway.

On top of the hardware configuration, there was some minor additional software configuration needed by Mac OS X. Under "System Preferences", in the "Sharing" option, the check mark box needed to be enabled next to "Internet Sharing" as well as the check mark box next to the USB Ethernet Adaptor. This allowed the computer to pass-through the connection from the built-in Ethernet Adaptor to the USB Ethernet Adaptor.

4.2.4 Commands for Testing

When using DummyNet to control the network environment, we utilized the following commands to instruct DummyNet to set the latency according to trial schedule. Our trial schedule can be found in Section 4.8 for reference.

Script	Latency	Command for Trial
Start.sh	-	`sudo ipfw add 100 pipe 1 ip from any to any`

Finish.sh	-	`sudo ipfw pipe 1 delete`
		`sudo ipfw pipe flush`
		`sudo ipfw flush all`

Status.sh	-	`sudo ipfw pipe list`
Trial 1	0 ms	`sudo ipfw pipe 1 config delay 0ms`
Trial 2	50 ms	`sudo ipfw pipe 1 config delay 50`
Trial 3	25 ms	`sudo ipfw pipe 1 config delay 25`
Trial 4	75 ms	`sudo ipfw pipe 1 config delay 75`
Trial 5	150 ms	`sudo ipfw pipe 1 config delay 150`
Trial 6	100 ms	`sudo ipfw pipe 1 config delay 100`
Trial 7	125 ms	`sudo ipfw pipe 1 config delay 125`
Trial 8	75 ms	`sudo ipfw pipe 1 config delay 75`
Trial 9	0 ms	`sudo ipfw pipe 1 config delay 0`
Trial 10	150 ms	`sudo ipfw pipe 1 config delay 150`

Table 1: The scripts used organized by trial and latency.

"sudo" is used in all scripts to elevate the commands to root privilege, which is required since DummyNet affects all network flow going to and from the computer within the kernel "ipfw" is the interface that invokes DummyNet on FreeBSD and Mac OS X. "pipe" is the first command given to DummyNet that specifies we want to work with a DummyNet "pipe". A DummyNet pipe is a type of queue which is used to shape traffic flowing through it with a various set of options, such as delay, loss, and packet corruption. The "start.sh" script creates a DummyNet pipe, labeled as "1" that applies to "100" percent of the traffic. A DummyNet pipe is only applied if a packet that is received matches the specified "from" or "to" IP address, which we specified here as "any".

The trial scripts are all similar in structure, with the first command after "sudo ipfw pipe 1" being "config", which tells DummyNet that we would like a configuration change to pipe 1. "delay 0ms" tells DummyNet that we would like to change the delay for this pipe, to 0 milliseconds. If there is no unit specified, DummyNet defaults to milliseconds.

The "finish.sh" consists of 3 lines. The first one with "delete" tells DummyNet to specifically delete "pipe 1". The second line with "flush" tells DummyNet to clear, or "flush" any rules for DummyNet pipes. The last command "flush all" tells DummyNet to clear any and all rules relating to DummyNet.

The “status.sh” script asks DummyNet to “list” all active pipes, which returns the status of each pipe, including information such as whether it is active, what percent of traffic is flowing through it, and what options it has. We primarily used it to check if a pipe was successful in being active and running.

4.2.5 Pilot Studies

In order to test our setup, we configured the hardware and software in an off-campus location. We used the exact test setup that we planned to replicate in Fuller A17, with the only difference being the location.

Using the previously mentioned setup for our pilot study, we used the game “Dear Ester” while executing our scripts to test whether we could successfully control the latency of the network communications going to the OnLive micro-console. To confirm that latency could be successfully adjusted using this setup, we tested the extremes of 150 milliseconds of latency and higher, until the OnLive service refused to allow us to play the game due to poor network conditions at 200 milliseconds.

We conducted additional trials with our roommate and our chosen game, “Crazy Taxi”, to determine if our methodology would be able to produce constant and repeatable results that we would be able to draw conclusions from. We conducted the experiment in the order the actual trials were intended to happen, and had the pilot study participant fill out the survey after each trial, as stated in our procedure in Section 4.9, Testing Procedure.

4.2.6 Environment

The room that we used to carry out trials for our experiment in is designated a Digital Arts lab, located in the sub-basement of Fuller Labs on the Worcester Polytechnic Institute campus, with a green area set aside for capturing video that could later be edited using “green screen” replacement and a 3D printer to print out 3D models, and a TV with several video game consoles and a cabinet of video games. We used this existing area to configure the OnLive micro console for our trials.

The room was chosen for its relative isolation, as being underground on the sub-basement level provided natural reduced foot traffic and eliminated the possibility of noise pollution from outside of the room with the exception of the hallway. The room also had a minimum of other factors that were easily adjusted for such as turning off the 3D printer, closing the door, and others that are explained in detail in Section 4.9. Figure 4 shows the main part of the room that we used, including the TV.

Figure 4: A picture of the test area in the environment we used for the study, from the entrance.

4.3 Institutional Review Board

At Worcester Polytechnic Institute, all experiments involving human participants must be approved by the Institutional Review Board (IRB) [16]. The IRB ensures that all experiments that involve people are conducted in the proper maner, including informed consent and minimized risk of bodily harm. Due to the benign nature of our experiment, we were able to go through the expedited review process, which took less than a week once the forms were submitted. We did have some major concerns based on the form that we had to fill out including personal data collection and storage, advertising, and consent forms. The advertising that we used is detailed in Section 4.4 and we used Qualtrics, detailed in Section 4.7.1, for survey data storage. The consent form we created is included in Appendix C. Any mention of the experiment itself is kept as vague as possible in the consent form so that the participant is not made aware of exactly what we are testing for. Apart from that, the consent form consists mainly of formalities given by the IRB department, however it does have an extra component to collect email addresses for the raffle, explained in Section 4.4.1 [16].

In order to avoid complications with the approval process, we abandoned a few possible options to include in our experiment. The first was the possibility of recording facial expressions to observe whether participants were enjoying the gaming experience or not. Another restriction was the amount of demographic data we collected, detailed in Section 4.7.2. Avoiding these areas as well as anything that could harm the subjects in any way allowed us to complete IRB process very smoothly.

4.4 Advertisement

In order to get a high number of participants for our study, we used a variety of methods to advertise to students. The initial approach to advertising consisted of an email sent to the entire student body briefly detailing the experiment. We created a standard paper sized poster to attract students' attention and placed them around the campus. The poster directed onlookers to a website where they could sign up for a time slot, described in Section 4.5, and incentives for participating in the study. While we did make a few posters, Figure 5 below is the one we printed. We later decided to include the poster in the advertisement email to the student body as well. Some of the participants were friends of the researchers that were contacted via Facebook or in person. Robert Dabrowski also spoke in front of a few CS and IMGD classes to promote the study. The main source for participants came from WPI's Psychology Department Participant Pool, described in Section 4.6.

Figure 5: One poster printed to advertise to participants.

4.4.1 Incentives

Some of the incentives used to encourage participation in our study included free refreshments during the experimentation process (Coca-Cola, water, Orange Fanta) and a 50 dollar Visa Gift Card raffle for any participants who wished to be included. Participants would enter the raffle by including their email address on the consent form. We also heavily promoted the idea of "playing video games for science" to attract participants.

4.5 Scheduling Participants

To allow potential participants the opportunity to pick their own times, we used a website called slottr.com. [17] Figure 6 below is a screen shot of what the people see when they are signing up for our experiment. We made sure to include directions for getting to the lab after the first few people got lost. We must point out that Slottr is not stable and while good for a single day, it was not very helpful for the two weeks that we conducted our experiment.

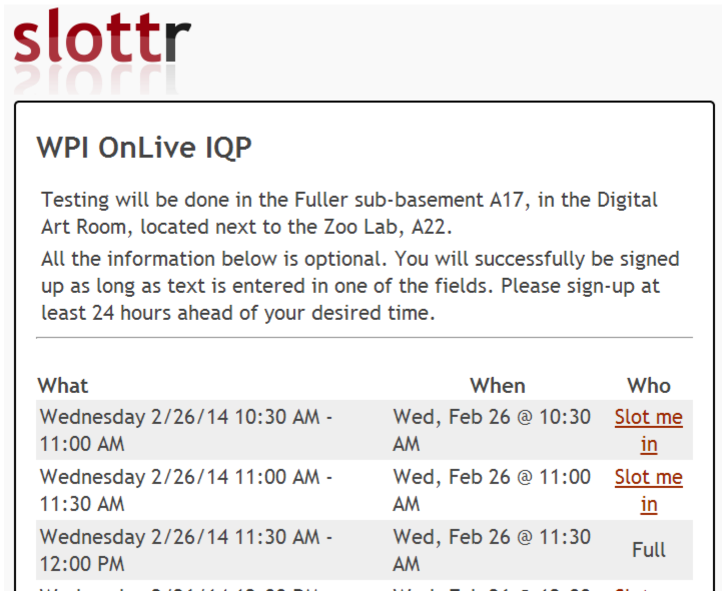

Figure 6: A screenshot of our page on the scheduling website we used.

4.6 Psychology Department Participant Pool

WPI's Psychology department makes use of an on line participation pool, where participants can sign up for studies and receive credits for various reasons. The participants are able to log on, see what studies are available, and select time slots for the studies. The researchers are notified via email of any new participants. Our team made use of this system to pull in participants for the project.

4.6.1 Forms and Approval

In order to make use of the participation system, we needed to provide the psychology department with a debriefing statement about our study and an application that included all researchers involved (and the advisers), a description of the experiment, and the length of the study. In addition to that, our team also needed to provide the IRB approval forms and the consent forms. These forms are included in Appendix C.

4.6.2 Built in Scheduling

The WPI Participation Pool comes with a built in scheduling system that allows researchers to select the times for their studies, and then for participants to select the times they want. This makes it convenient for our team because it was much easier to schedule participants for our study. Sona Systems is a cloud-based participant management software that enables universities to manage research and recruit participants in a cloud-based environment. According to WPI policy, if you are a faculty member, graduate student, postdoctoral fellow, visiting researcher, visiting scholar, MQP student, or IQP student, you are eligible to put in a request to post studies in the Participant Pool.

4.7 Collecting Data via Survey

For our experiment, we planned to collect data primarily through a survey. In general, surveys are great for providing anonymity while still collecting people's opinions in an efficient manner. Through our survey, we were able to ask each participant the same questions in the same way every time, thus providing consistent interactions with participants. Our survey was also digital and automatically stored results for us, detailed in Section 4.7.1 below. We do also collect player score data, but that is separate from the survey and is explained in Section 4.9.3. The entirety of the survey that we used is located in Appendix C.

4.7.1 The use of Qualtrics as a Tool

The service that we used to produce and provide the survey was Qualtrics, which is a product that WPI provides access to for students and faculty [18]. Using this tool, we were able to draft all of our questions and actually take the survey to see what our participants might see. Because Qualtrics is mobile device compatible, we were able to use the tablet mentioned in Section 4.2.1 to give the participants an easy way to answer survey questions.

The survey question creation utilities in Qualtrics are very extensive and allow for nearly any kind of question, however we primarily used multiple selection and multiple choice questions. One very useful question feature that we used is a multiple selection box that holds a complete list of majors at WPI. Data validation is also available, and we used it to make answering the questions for the individual trials mandatory.

Once we started conducting experiments, Qualtrics also recorded all of the survey data for us. With all of the recorded data, Qualtrics can perform its own data analysis or export that data to a file compatible with Microsoft Excel file or MatLab. For example, its analysis tools can create a "Cross Tabulation" analysis that shows how those from different demographics answered questions later during their trials.

4.7.2 Demographic Questions

To give our data some context, we decided to ask a few general and a few very specific demographics questions of our participants. The general questions included gender, age and major. We made sure to allow students to select multiple majors if applicable. We did not ask any more than that because we did not predict that any more general demographics information would be useful.

The specific questions we did ask focused on determining how skilled and familiar the participant is with video games. The first question asks how many hours per week the participant plays, and we gave options in increments of 5 hours from 0 to 25+ hours, with 25+ hours being the average we observe in serious gamers around campus. The next question asks what kinds of video games the participant plays the most. Many genres are included because there are many kinds of games. We expect that those who selected racing games or one of the fast pace gaming genres will perform better and be more attuned to detecting changes in latency. The final video game related question asks about the participant's past experience with video games. Using different combinations of answers to these questions, we hope to notice trends in player scores vs. gaming experience. This data can also help to show only experience gamers, anybody or nobody can notice changes in latency.

The final question asked how the participant found out about and signed up for our experiment so that we could determine how effective our advertising was. The options were email, posters, Psychology Department, word of mouth and other.

4.7.3 Questions for Individual Trials

For each of the ten trials, our survey asks the same four questions. They are each on a scale of one to seven, with one labeled as worst and seven labeled as best. This consistent formatting allows the participant to quickly fill out the survey and continue with their next trial. The first question asks the participant how they think they performed, which we will we used to determine if they think they performed better with lower latencies or not. The next question asks what the participant thought of the graphics and visuals. We plan to use this to see if they notice changes in frame rate or quality with changing latencies. Next we ask what the user thought of the responsiveness of the controls. This is the main question used to determine whether the participant could feel the latency or not because the controls should be the most affected quality of games run using thin client technology. The final question asks how enjoyable the experience was. We hope to use this to show definitively that increasing latency decreases the enjoyability.

4.8 Schedule of Latency During Tests

Our schedule for latency tests was determined by the following criteria:

- A schedule of latencies that we could corroborate to trial numbers in our survey data
- A regular interval to change the latency enough to have a significant impact and small enough to provide accurate data, but large enough to minimize the number of trials needed
- A range of latencies where the maximum value would be on the edge of what the OnLive micro-console could handle
- An ordering that was not predictable by the user
- An ordering that could help distinguish and isolate any improvements made by the player over the course of the experiment
- In deciding the order, we also needed to avoid significant and noticeable changes in latency in order to avoid drawing the user's attention to the varying network conditions.
- A control latency to give all participants a consistent reference point

For all of these reasons, we decided on the latency range of 0 milliseconds to 150 milliseconds, which was determined to be the maximum amount of added latency by our pilot studies in Section 4.2.5. An interval of 25 milliseconds gave us a total of seven total trials with time to do three additional trials with repeated latency values in order to account for changes in each user's improvements. We decided to repeat the latency values of 0 milliseconds, 150 milliseconds, and 75 milliseconds because 0 and 150 represented the two extremes of our experiment, and 75 milliseconds is located in the exact center of those two values. The order of the latencies was determined by the fact that we needed to start at the best case or worst case scenario, and then to avoid large changes in latency. We

Trial Number	Latency (milliseconds)
1	0
2	50
3	25
4	75
5	150
6	100
7	125
8	75
9	0
10	150

Table 2: Our schedule of latencies that we used for our trials.

attempted to avoid a change in latency at any given point over 50 milliseconds, which we were able to accomplish with the exception of the last three trials of repeated latencies, and the change from trial 4 to trial 5. Table 2 below shows the entire schedule.

4.9 Testing Procedure

In order to provide consistency between each experiment, the steps detailed in the following subsections were performed for every test. The final version of our experiment proctoring instructions is in Appendix C. It should be noted that the purple text was not read to the test subject.

4.9.1 Before Subjects Arrive

The steps in the list below are what we had to do to prepare our workstation and room. Steps 1 through 3 were always necessary for this experiment, while the refreshments in Step 4 were optional incentives used just for attracting participants, further detailed in Section 4.4.1. The specific commands used on the network control workstation terminal are detailed in Section 4.2.4.

1. Connect OnLive box through laptop network control workstation

2. Load Crazy taxi on the OnLive box

3. Have consent forms ready

4. Have refreshments available

5. Make sure the 3-D printer is turned off

Step 5 is specific to our lab, Fuller A17, because not every lab will have a 3-D printer. While it could have been any machine making noise in the room, it was critical to turn that machine off so that test subjects would not be distracted. More details about the room are provided in Section 4.2.6.

4.9.2 When the Subject Arrives

When the subject arrives, we greet them with the opening statements listed on the script as well as the consent form explained in Section 4.3. We also give them a minute to have refreshments and get comfortable. Next we have them fill out the demographics part of the survey described in Section 4.7.2 on one of our tablets.

Finally, we tell them that the game will be Crazy Taxi and that the first round they play will be a practice round. This practice round is a special mode of Crazy Taxi in which the player gets exactly 5 minutes to play. We also tell the test subject that the practice round is with "best game conditions", but nothing more. Throughout the entire introduction, we tell the subject very little about the experiment, except that "game conditions may vary". Keeping the subject uninformed helps to reduce the number of biased answers during the experiment.

4.9.3 During the Experiment

Once the subject finishes the practice round, we tell them that they will play ten rounds that last about 60 seconds, then they should fill out a page of the survey after each round. Unknown to them, we record their final scores and associate them with the Qualtrics survey by trial number in our notes. Also without their knowing, we change the added latency between each session, with the commands listed in Section 4.2.4 and the schedule listed in Section 4.8. It is very important that the latency settings be changed while the user is navigating a menu instead of during gameplay, mainly so that the conditions do not change while the user is playing. If the latency conditions are changed during game play, the frame rate drops far too low to be playable with any amount of latency. We believe this is a result of how OnLive reacted to the changes made in network conditions.

4.9.4 Pre-Answered Questions

In anticipation of test subjects asking questions during or before the experiment, we recorded and answered some potential questions. As mentioned in Section 4.9.2, it was important to prevent the participant from knowing the full purpose of the experiment. With that in mind, the answers to the predicted questions are intentionally vague. The following general statement was also included to answer all questions not accounted for: "we will explain that aspect of the experiment once you have finished all ten trials."

4.9.5 After the Experiment

To conclude the experiment, we thank the test subject and read the debriefing statement at the end of the proctoring instructions. This statement is very short, but explains what was different about the sessions of game-play and the general intention of the experiment. Next, they get a chance to ask any questions. Finally, we request that he or she does not disclose anything about the experiment to friends or classmates so that future participants remain unbiased during later experiments.

5 Results and Analyses

In this section we provide our data, multiple ways of examining at it, major trends and some analysis. The data collected from the individual trials of our experiment come in two forms: player score and ratings given by participants in the survey. The player scores are represented in two ways: dollars received as payment for delivering customers and number of customers delivered. The number of customers delivered is data extracted from the player scores based on certain ranges of scores that can only be achieved by delivering a certain number of customers. The survey data is all represented as values from 1 to 7 on a worst to best scale for each of the four questions asked during each trial.

Section 5.1 reports the average trends in answers to those questions. Section 5.2 shows the effects of repeating latencies on player scores and answers to the survey questions. Section 5.3 provides an overview of the survey data collected during the trials and some major trends. Section 5.4 reports on the demographics of our participants. Section 5.6 reports on the effectiveness of our advertising methods.

Appendix A contains additional plots and graphs not seen in this section. Appendix B provides data tables not shown here.

5.1 Average Trends

To examine general trends in our data, we took averages for scores and survey answers across all participants for each trial. Section 5.1.1 shows average trends in survey data and Section 5.1.2 shows average trends in player scores.

5.1.1 Survey Data

All of the following graphs were created from average answers to the survey questions for each trial. They are all formatted in the same way not only because the data was collected in the same format but also because we expect them to have similar trends. The X and Y axes of each graph show latency in milliseconds, on a scale from 0 to 150 milliseconds, and the average rating, on a scale of 1 to 5, respectively. The data for each graph actually has ratings that span the range of 1 to 7 instead of 1 to 5, but none of the averages for any question or trial exceed a value of 5, so the graphs are shortened to better show the region of interest. Each point represents the average for one trial at a particular latency. Some trials repeated a particular latency and these graphs reflect that. Also included in each graph is a best fit linear approximation, accompanied by the equation for that line and coefficient of determination, also known as CoD, R-squared or R^2. For all of the data points used in these plots, see Table 5 in Appendix B.

Figure 7 shows the averages responsiveness ratings of the controls at each latency. The blue line represents a linear fit of the data, and it has the highest CoD of any our survey data, with a value of 0.9737.

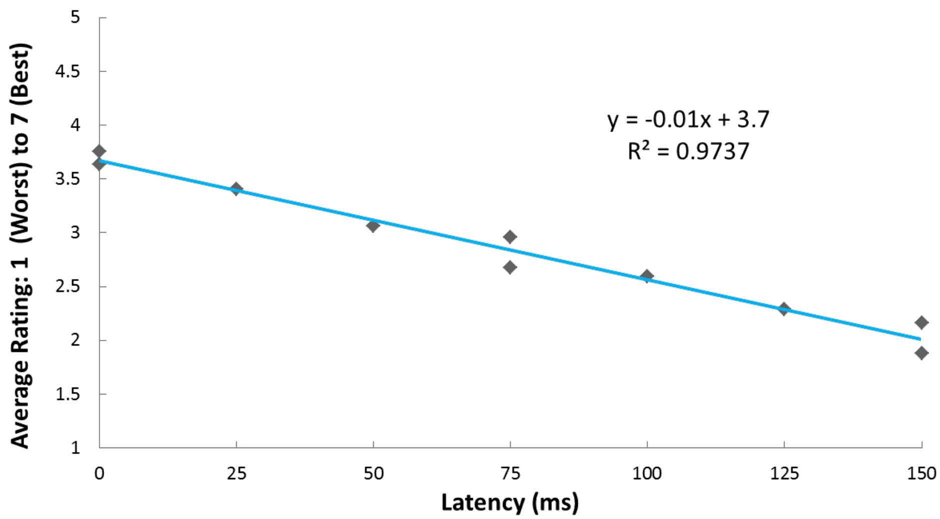

Figure 7: Average responsiveness rating.

The overall enjoyability rating for each trial, shown in Figure 8, also had a high CoD value of 0.8803. Because we expected a high correlation between enjoyability and responsiveness, we also combined those scores to create Figure 31 in Appendix A, which shows the average of those two scores. Note that the linear fits in Figures 7, 8 and 31 each have the same slope of -0.01 rating points per millisecond of added latency. At that slope, a full rating point or 14% of the maximum rating is lost for every 100 milliseconds of added latency o

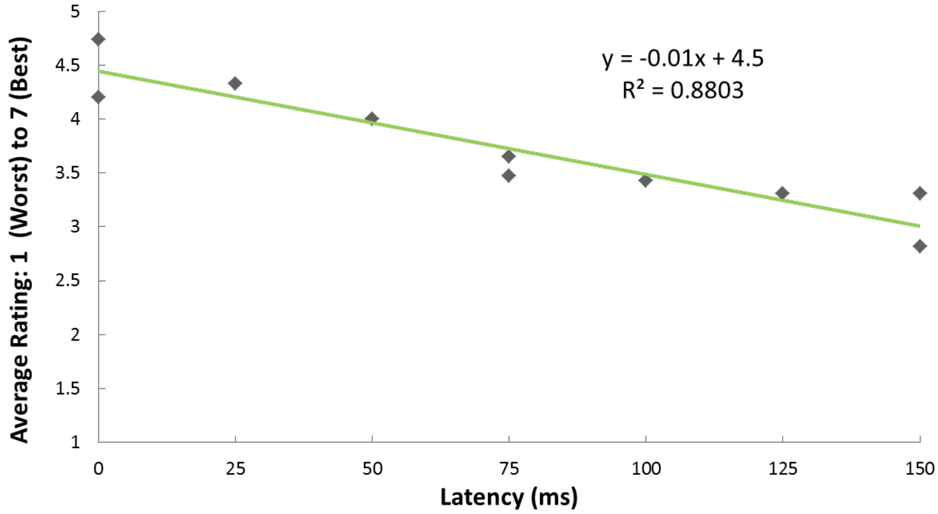

Figure 8: Average enjoyability rating.

The trend lines in Figures 7, 8 and 31 are very strong predictors for what ratings participants will give when between 0 and 150 milliseconds of latency are added. We predict that this trend line will approach the minimum value of 1 as latency increases. While the line could also continue to go up as latency decreases, it does not make sense to predict responsiveness ratings with negative additional milliseconds of latency because it is not possible.

By comparison to the responsiveness and enjoyability, the average ratings for image quality rating and self evaluation of performance were not linearly approximated as well. Figure 9 shows a yellow line very poorly fitting the performance rating data. The CoD is 0.4338, but by looking at the points around the line it is clear that there is not a strong relationship in the data. We tried some other methods of fitting equations to this data without success. Despite the poor CoD, the slope of the line was -0.006 rating points per millisecond of added latency, which is 60% of the slope of the responsiveness and enjoyability ratings.

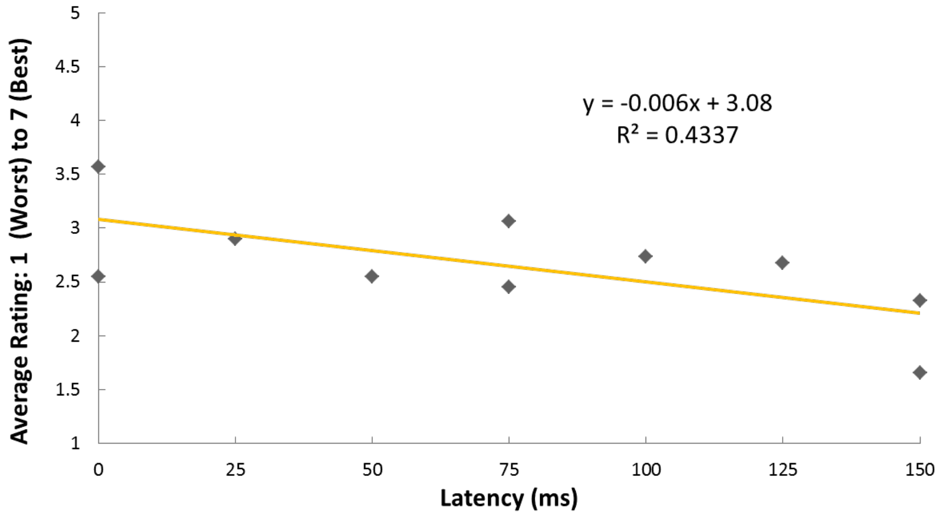

Figure 9: Average self evaluation of performance.

Figure 10 shows the average image quality rating with a higher CoD of 0.7696. The linear fit of this line has a slope of -0.0025 rating points per millisecond of added latency, which is nearly flat compared to the other plots indicating that participants did not notice a change in graphics quality. This matches the observation we had during pilot studies that the graphics quality does not change noticeably with latency for Crazy Taxi.

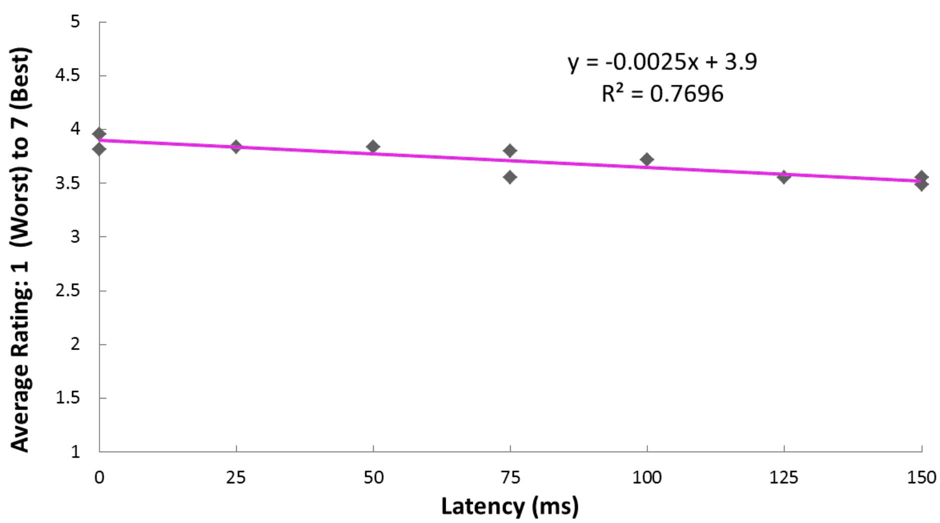

Figure 10: Average quality of graphics rating.

To put these average trends in the context of the Quality of Experience (QoE) index used in [2, 7, 10], we also calculated a QoE score. Similar to those studies, we too used the answers from each question without any weights for each trial. However, instead of adding the scores together and scaling the result to an arbitrary range, we took the average of the answers to the survey questions per trial in order to match the format of the individual trial graphs shown previously.

The result of our QoE calculation is shown in Figure 10. We were surprised by this plot because the CoD has a high value of 0.9298 despite having the poor self evaluation of performance rating included in the calculation. Even the slope of -0.007 rating points per added millisecond of latency for the linear approximation is very close to the slopes observed in the responsiveness and enjoyability ratings in Figures 7, 8 and 31.

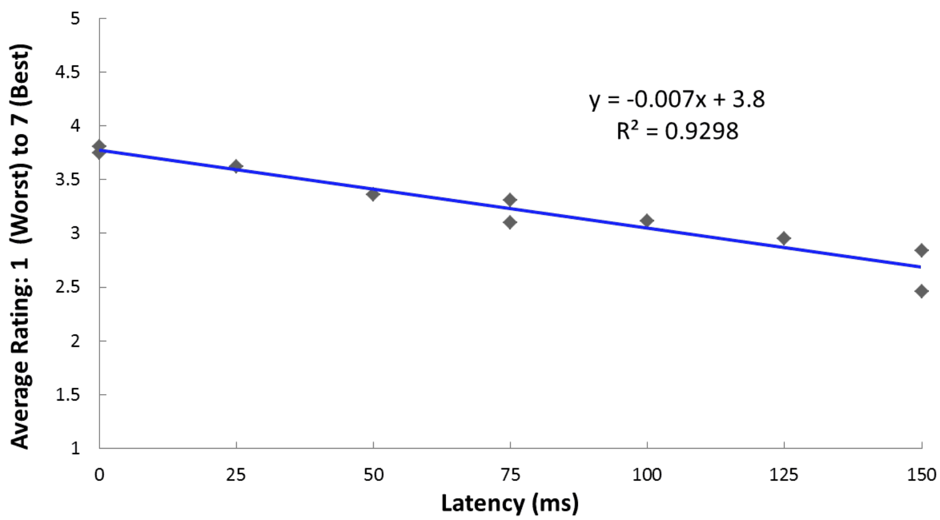

Figure 11: Average combined Quality of Experience rating.

Table 3 summarizes the graphs shown in this section. The rows represent each of the questions and the columns show the slopes, CoD values and y-intercepts of each trend line examined.

Rating	**Slope**	**Y-intercept**	**CoD**
Responsiveness	-0.01	3.7	0.9737
Enjoyability	-0.01	4.5	0.8803
Performance	-0.006	3.1	0.4337
Graphics Quality	-0.0025	3.9	0.7696
QoE	-0.007	3.8	0.9298

Table 3: Summary table of the average trends for each survey question as well as QoE.

5.1.2 Player Scores

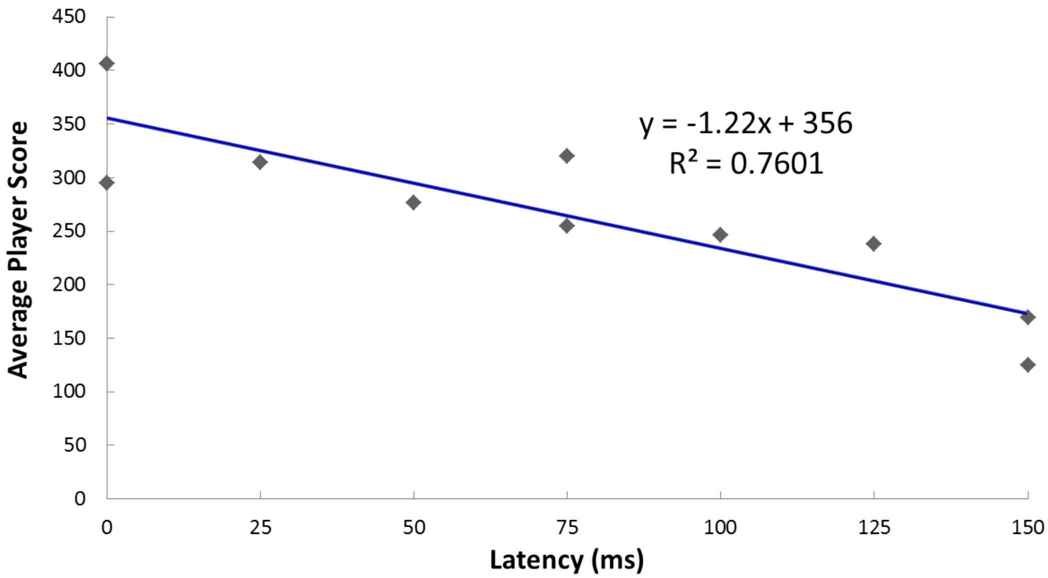

Figure 12: The average player score across each latency tested.

Figure 12 shows the correlation between average player score and latency. There is a downward trend in player scores as latency increases. This downward trend isn't completely linear, and fluctuates between lower and higher scores. The players seem to have gotten skilled enough to get adequate scores despite high latencies. However, high latency was still enough of an effect on the player's performance to prevent them from achieving higher scores.

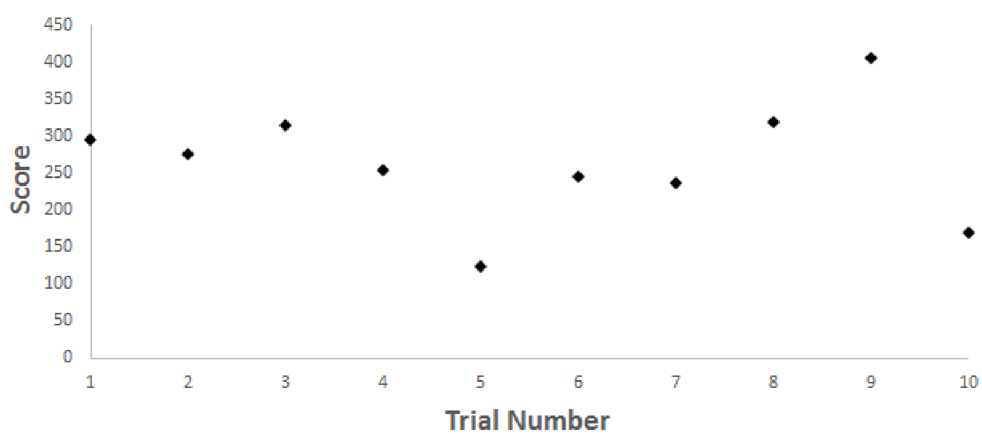

Figure 13: The average player score across each trial.

The game play and game levels during each of the trials were similar enough to each other for participants to become familiar with the game objectives and strategize. The participants got better at the game as the trials continued. There is evidence of this in Figure 13. This graph shows the average scores of players according to trial number. At trial 3, the average player scores were higher than the ones previous to it. This was at 25ms of latency. These results would suggest that latency at these low levels did not effect participant performance enough to cause them to get lower scores. The participants performed better than at 50ms of latency, but also at no latency. Trials 6 and 7 were at 100ms and 125ms, respectively, and the averages for both are fairly similar. From these results, it would seem that a latency level of +/- 50ms is needed for significant effect on participant performance.

Figure 14: The average number of customers participants were able to deliver at each latency.

Participant scores were mainly determined by how many customers they managed to deliver. This was where a majority of the score was achieved. To get a broader group of player performance, we kept track of how many customers participants were able to deliver. Figure 14 shows the total average number of customers delivered at each latency tested. On average, participants were able to deliver, at most, 1 customer during their trial. The more skilled participants managed to deliver 2-3 customers, which was more difficult than delivering 1. There is an increasing trend in the number of participants who delivered no customers. Figure 15 shows the decreasing trend in the average number of customers delivered per latency.

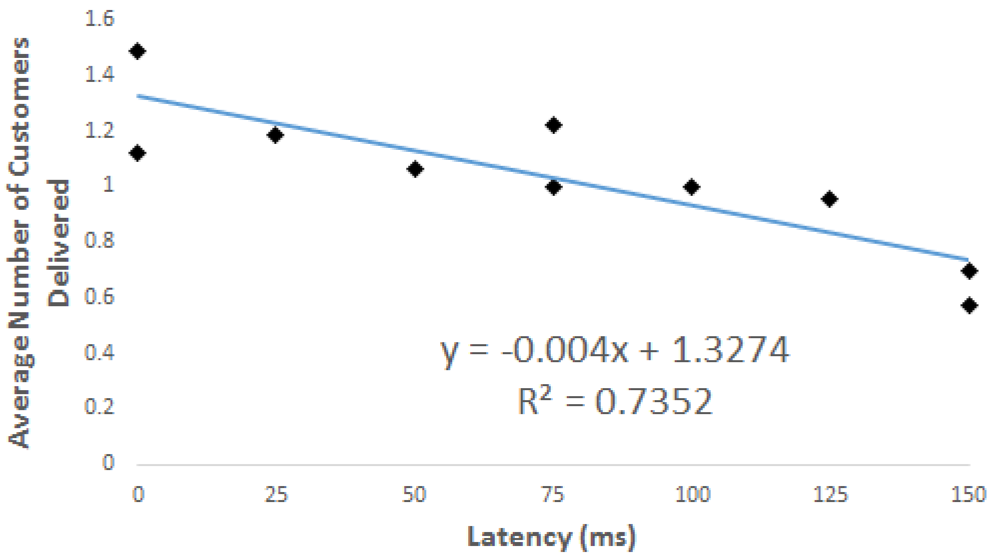

Figure 15: Average number of customers delivered.

5.2 Repeated Latencies

Our schedule of latencies, as described in Section 4.8, repeated three particular latency trials once each and this section shows how those pairs of trials compared to each other. The three repeated latencies were 0, 75 and 150 milliseconds representing no added latency, a medium amount of added latency and the maximum added latency, respectively. These particular latency levels were chosen to help distribute the difficulty of game play that comes with higher latency levels. The trials were conducted according to this schedule in order to help us gauge how well players adjusted to the various latency levels. Section 5.2.1 shows how participants changed their answers to the survey questions and Section 5.2.2 shows how their scores changed.

5.2.1 Survey Data

In order to see how our participants changed their answers when latencies were repeated, we calculated the change in score from the first exposure to that latency to the second. Figure 16 shows the average change in answer given for each question at each of the repeated latencies. Each question is included here as well as the QoE score, which is calculated as the average of the answers to the four survey questions. This graphs shows that only QoE and the participants' self evaluation of their performance improved with the second exposure to a given latency. While the answers to overall responsiveness and image quality were positive or negative depending on the latency, the magnitude of the average change is less than three tenths of a point on a seven point scale.

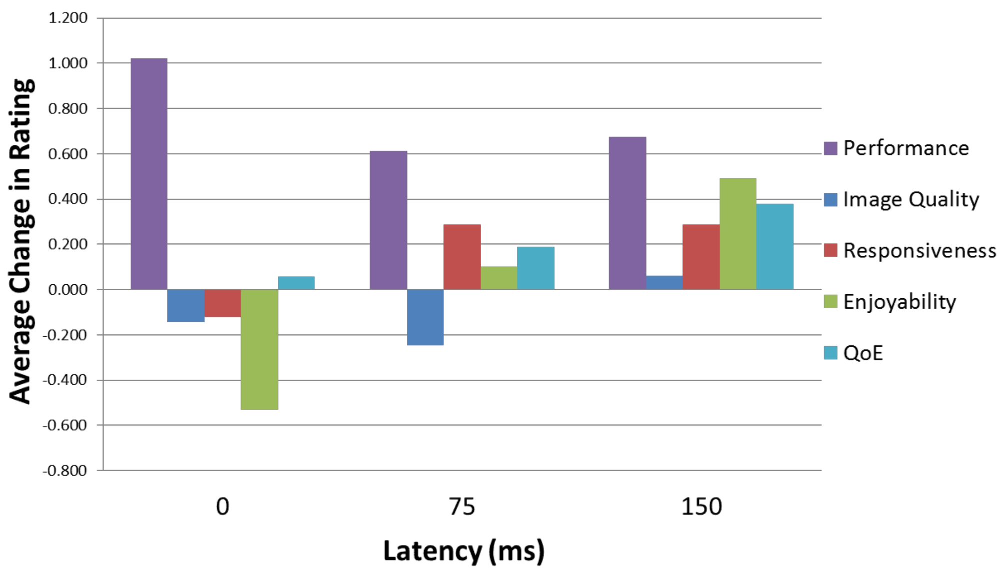

Figure 16: The average change in rating for each of the survey questions at each of the repeated latencies.

Figure 17 shows how participants changed their answers for each question in the trials at 150 milliseconds of added latency. From this graph it is clear that many participants did not change their scores for all of the questions, but many also changed by one or two points. Figures 32 and 33 in Appendix A show how participants changed their answers at 0 and 75 milliseconds. The change in QoE score is not shown in Figure 17 because for most participants the change was not a whole number and the maximum magnitude of change was 2.25 points.

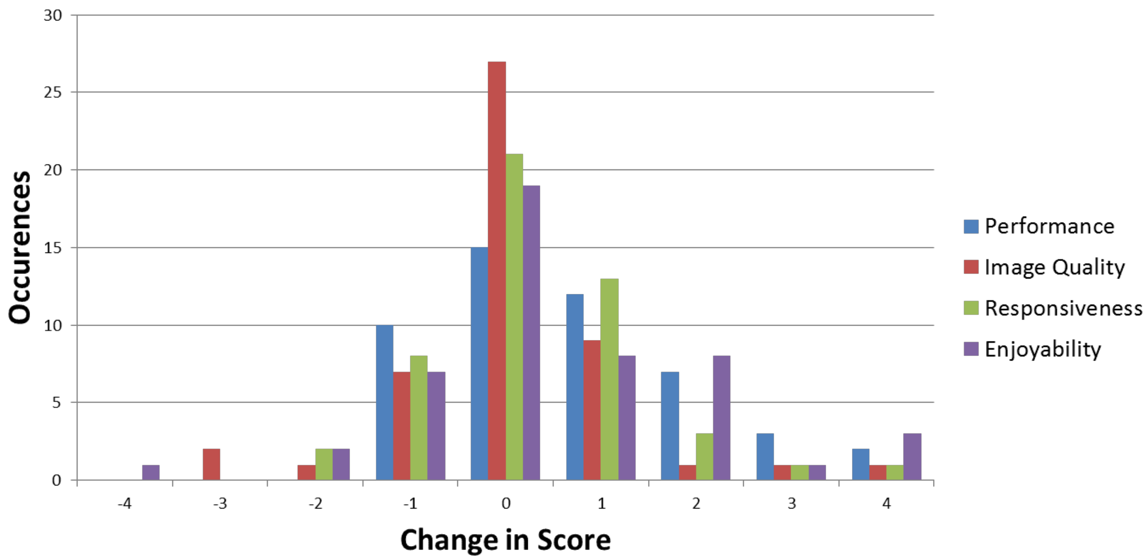

Figure 17: Change in ratings at 150 milliseconds.

Our data shows that it is difficult to identify similar or identical network conditions. If players could not notice that the latency had been repeated, then they might not be able to notice more harsh changes in latency. This means that our experiment would have to be repeated many more times at many more latency settings to prove or disprove all of our hypotheses from Section 3 about tolerances for noticing latency.

5.2.2 Player Scores

Figure 18 shows the distributed averages among the players that improved their score during the repeated latency tests.

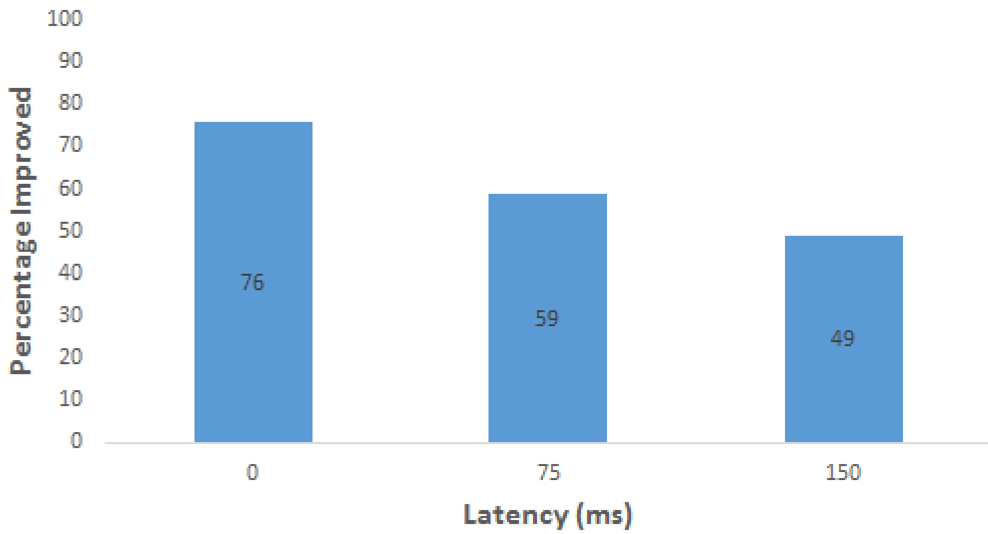

Figure 18: The average improvement percentage among the repeated latency tests

Across the tests with repeated lower latencies (0ms-75ms), there was an increase in the performance of a majority of the participants. There was also an increase in performance at higher latencies, though not as significant. This seems to suggest that the higher latency levels did not affect player performance enough to hinder their score a second time. Figure 19 shows the average score improvement among all of the players for each of the repeated latency tests.

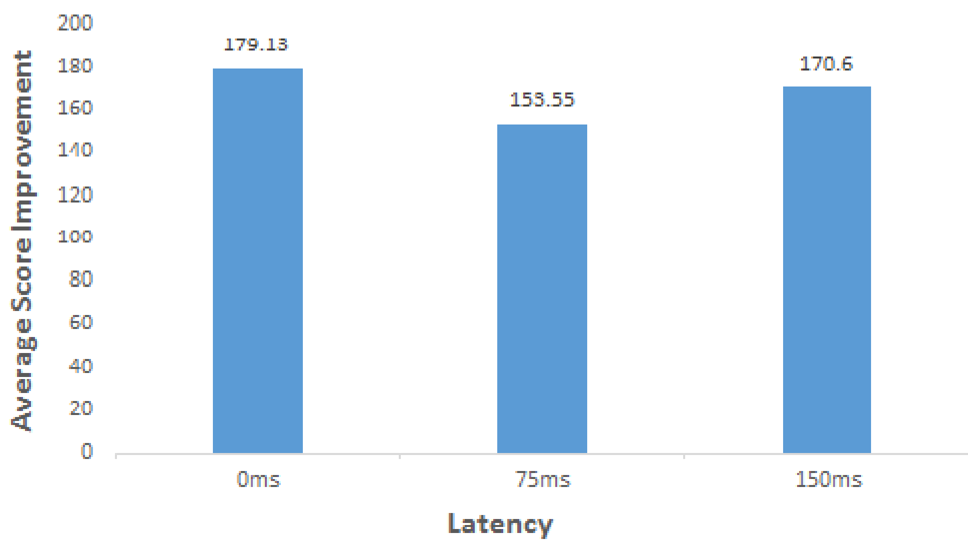

Figure 19: The average player score improvement across repeated latency tests.

Average player scores seem to increase the most at 0ms latency, while they increased the least at 75ms latency. The average distribution of scores led us to believe players would improve the least at 150ms latency instead of 75ms latency. At no latencies, most participants were able to get a grasp of the game mechanics and achieve a high score. The percentage of players that managed to score higher at 150ms were the players with much more gaming experience than the other participants. The more experienced players were able to adapt better to the higher latencies. Those with less experience were able to score adequately at 75ms, though not as high as the more experienced participants, thus lowering the average. This shows that the higher latency levels affected more experienced participant's performances less than those with less experience. This also shows that the participants were able to adapt to changing game conditions at lower latencies better than they were at higher latencies.

5.3 General Observations from the Survey Data

The graphs below are formatted to show histograms of answers given for each trial sorted by latency. The horizontal axis shows answers given, the vertical axis shows the number of times those answers were given and the depth axis shows each trial. Once we created these graphs, we made a few observations about the data as a whole.

The first observation we made from these graphs was that many users did not like to use the rating options on the better end of the scale when rating individual trials. Figure 20 shows that very few users used 5 and 6 as ratings and even fewer used 7. This could have happened because they did not think the controls were ever good or they used their own metrics to decide their rating.

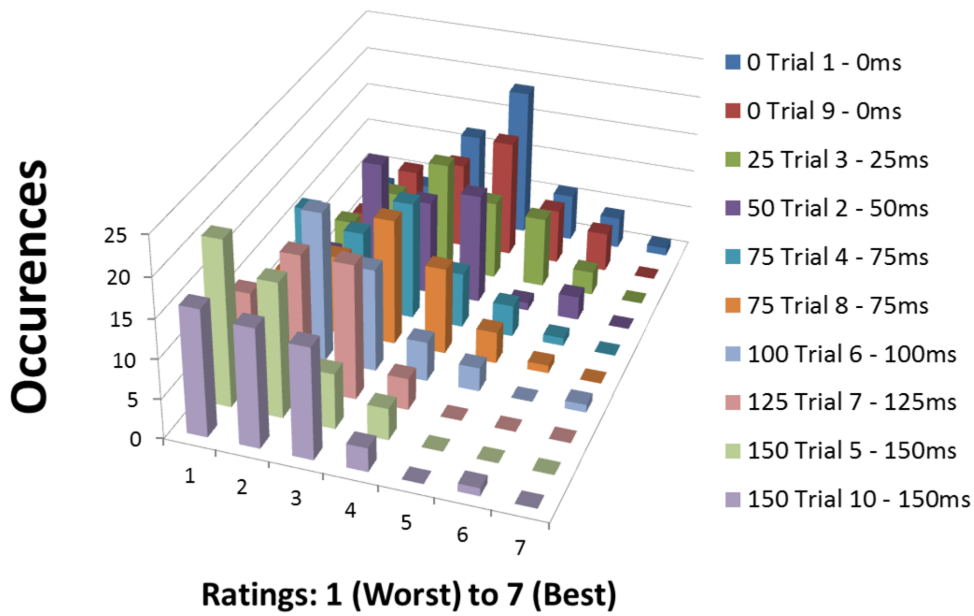

Figure 20: Distribution of responsiveness ratings by trial ordered by latency.

To compensate for the fact that many users did not use the full range of values offered, we normalized the ratings used by each user for each question, then scaled those values to be on the intended scale of 1 to 7. Figure 21 shows the result of that normalization on the responsiveness rating data. Here we observed that many users often did use the extreme values of their scales, and in general the usage of those extremes did correspond to extremes in added latency. This can be seen by looking at the tallest columns on the close left-hand side and far right-hand side, which are the extremes of both rating and latency. To encourage the full use of the rating scale, it could be beneficial to tell the user that their first trial should get the maximum score then not use that trial's data during analysis to eliminate bias.

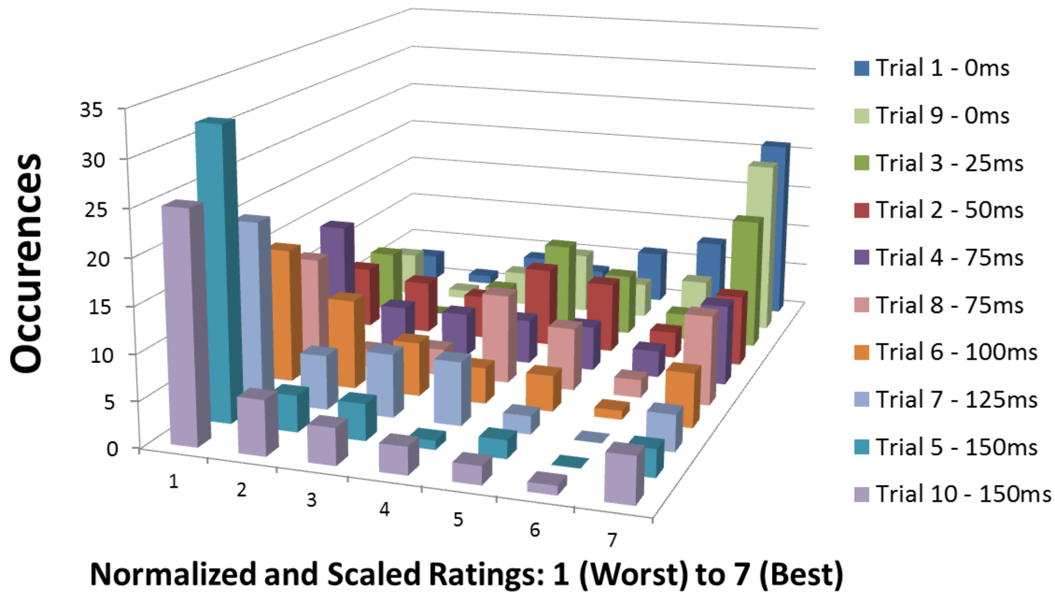

Figure 21: Distribution of normalized responsiveness ratings by trial ordered by latency.

While doing the previous analysis, we discovered a surprising distribution of ratings shown in Figure 22. Only the image quality rating has such a uniform distribution of ratings across all trials. Note that each trial closely resembles a normal distribution of ratings in this plot.

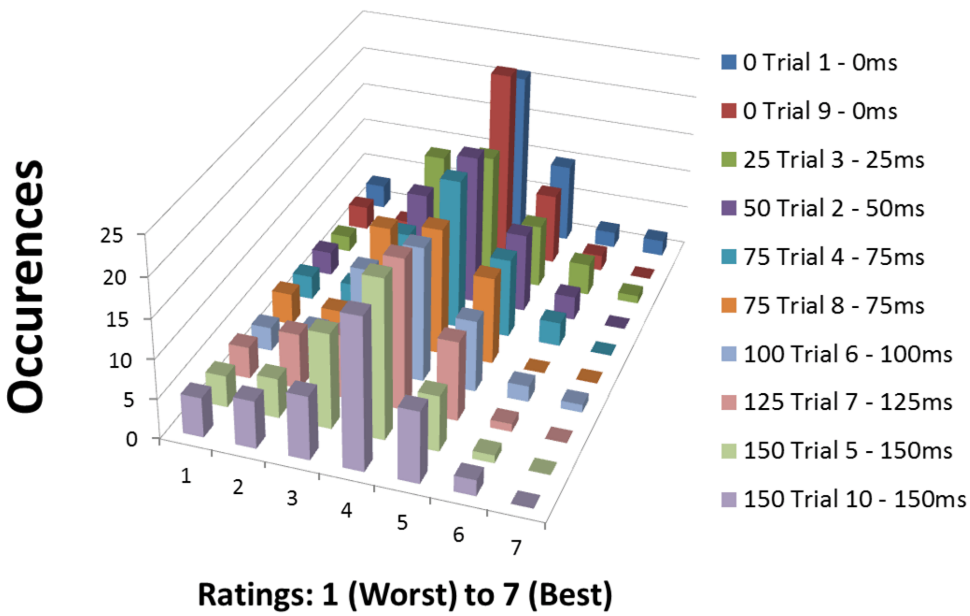

Figure 22: Distribution of image quality ratings by trial ordered by latency.

The figures above were also created for the rest of the trial survey questions and can be seen in Appendix A formatted in the same way. The additional plots exhibit similar features to those described in this section.

5.4 Demographics Represented

The following set of graphs represent our participants grouped according to various demographic information they provided in the first section of our survey.

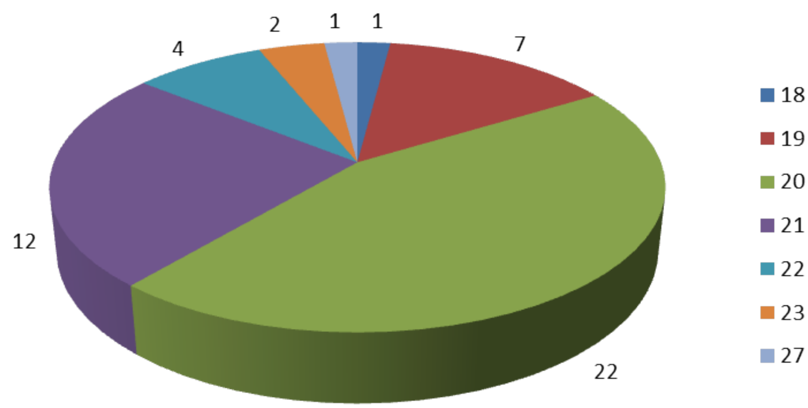

Figure 23: Distribution of participants by age.

Figure 23 shows the age range of participants, with the legend showing the various ages. The age range of participants is spread out between the ages of 18 to 27, with 84% of participants falling between the ages of 19 to 23.

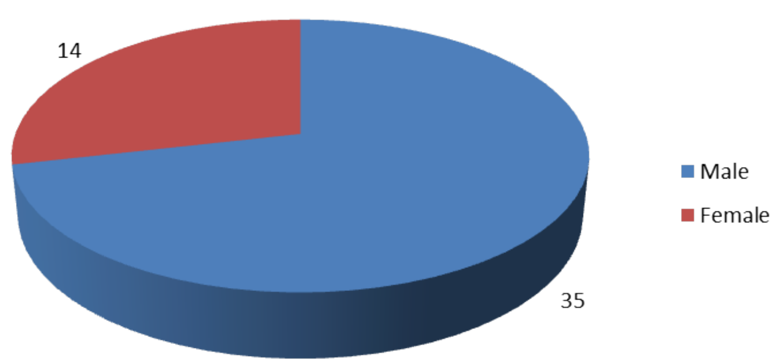

Figure 24: Distribution of participants by gender.

Figure 24 shows that the gender distribution for participants is weighted towards males, with 71% of our participants being male and 29% female.

Table 4 breaks down the distribution of participants according to the major they reported, and displays any major that occurred more than five times. It is important to note that when looking at this information, there is a potential for overlap, since a user may

Major	Number of Participants
Computer Science	14
Electrical & Computer Engineering	6
Interactive Media & Game Development	6
Robotics Engineering	8

Table 4: The number of participants grouped according to their reported major.

have multiple majors. The distribution of participants according to majors is a fairly even split, with the majority of being Computer Science, followed by Robotics Engineering, and Electrical & Computer Engineering and Interactive Media & Game Development being tied for third.

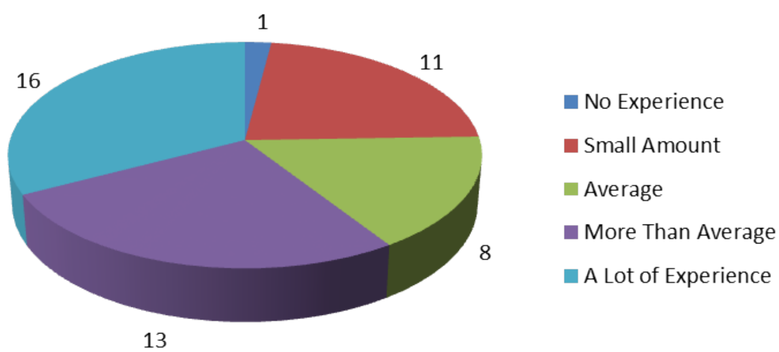

Figure 25: The previous video game experience of our participants.

Figure 25 shows the distribution of participants according to their previous experience with video games is fairly even in each classification except for the "No Experience" category. More than half of participants reported themselves as having "More Than Average" or more experience with video games.

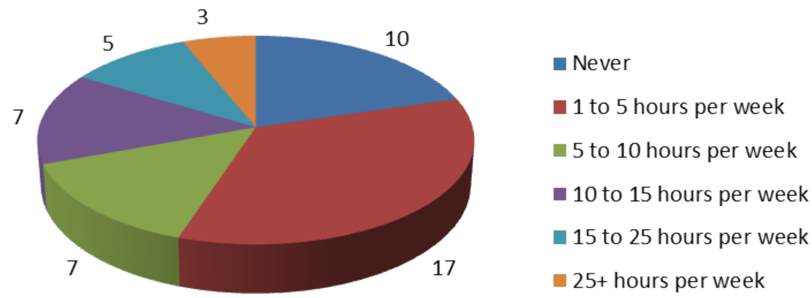

Figure 26: Weekly hours of video games played by our participants.

Figure 26 shows the distribution of participants according to average time they spend playing video games per week, which is a very different result compared to Figure 25. While many participants have "More Than Average" experience with video games, ten participants, or 20 percent of participants, reported that they never play video games on average. More than half, 55 percent of participants, report that they played less than 5 hours a week. This can be explained as a change in playing habits after accumulating a large amount of experience.

One of the goals of our study was to obtain a demographic of participants that would be considered as "gamers", or in other words rated themselves as having "a lot of experience" with video games or played more than 5 hours of video games per week. Looking at the data presented in this section, we achieved that goal with 25 participants fitting the profile of gamers and 24 non-gamers.

5.5 Gamers and Non-Gamers

In this section we explored the possibility that "gamers" would be able to notice changes in latency better than "non-gamers". The two survey questions used to make this distinction asked participants how many hours of video games they play per week and how much experience they have with games. We classify gamers as participants who play more than 5 hours of video games per week or claim to have "a lot of experience" with video games. Those who play fewer than 5 hours or do not claim to have much experience with games are considered non-gamers. By that classification, we had 25 non-gamers and 24 gamers, which splits our data in half.

The formatting used for the graphs presented in this section is the same as the graphs in Section 5.1, except that each here has two sets of points with a linear fit for each. The red points and lines represent the portion of the population who we classified as gamers and the green points and lines represent non-gamers.

To see if being a gamer or non gamer makes a difference, we started by comparing the average scores for each group. Figure 27 shows that gamers scores about 80 points on average more than non-gamers. Tables 7 and 8 in Appendix B also show that gamers got higher average scores than non-gamers in every trial.

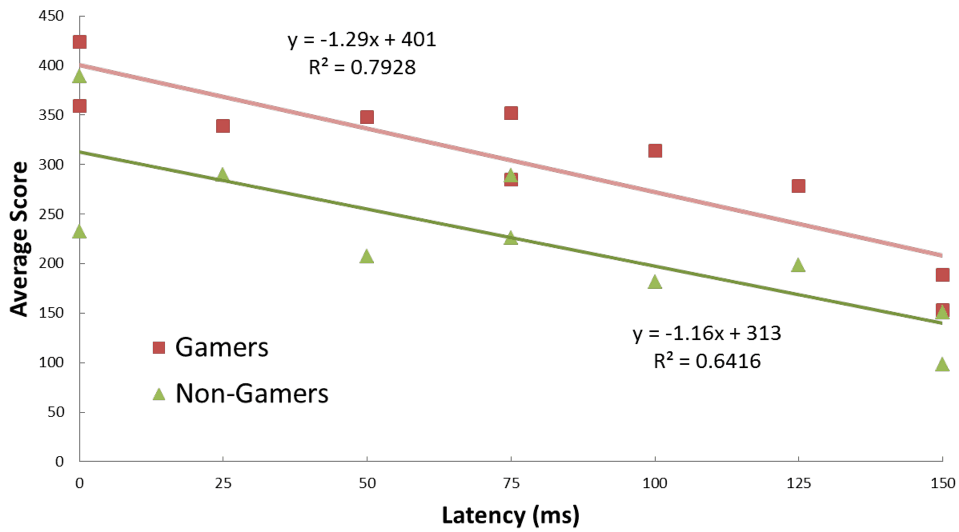

Figure 27: Average scores of gamers and non-gamers.

Because Figure 27 shows that gamers are better on average at Crazy Taxi at all latencies, we decided to see if they were also better at noticing changes in latency. Figure 28 shows how each group rated the responsiveness of the controls for each trial. Also shown in this graph is the linear fit of each set of data. Here, the difference in slopes of those linear fits can indicate a difference in perception of the latency. Because the slope of the linear fit for gamers is steeper than for non-gamers, we claim that non-gamers were better at noticing changes in latency. Figures 28 and 39 to 42 all show steeper slopes for the linear fit of gamers than for non-gamers. It is also expected that those with experience in gaming would notice a characteristic such as latency while play a networked game.

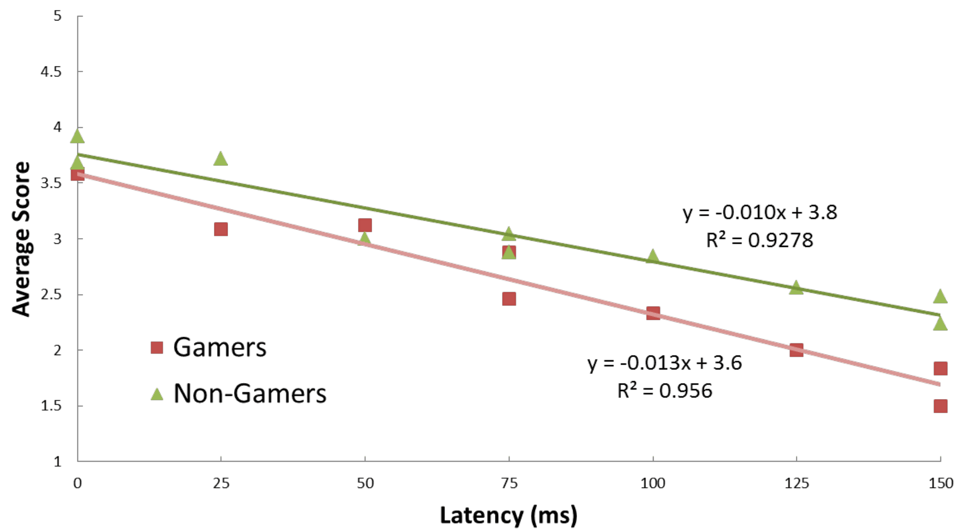

Figure 28: Average responsiveness ratings for gamers and non-gamers.

See Appendix A for the graphs showing how each group rated enjoyability, QoE, graphics quality and performance. Table 9 in Appendix B provides a summary table of the trend line values for each demographic.

5.6 Results of Advertising

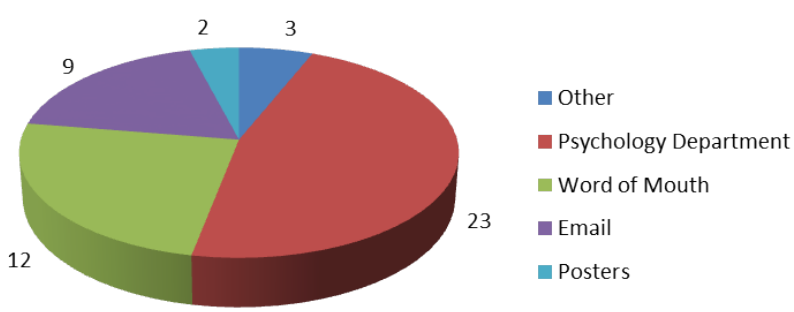

Figure 29: Distribution of participants by how they heard about the study.

Figure 29 shows that the distribution of participants according to how they heard about the study is heavily weighted towards the Psychology Department, with almost half of the participants coming from the Psychology Department Participant Pool. The other top two sources that participants reported were Word of Mouth and Email respectively. A significant portion of the Word of Mouth participants were likely people contacted by us directly or pulled in from the Zoo Lab and outside the room by Robert Dabrowski in an attempt to obtain more data. The Email source was likely due to our emails sent out to majors@cs.wpi.edu, imgd-majors@cs.wpi.edu, imgd-grads@cs.wpi.edu and rbe-majors@wpi.edu.

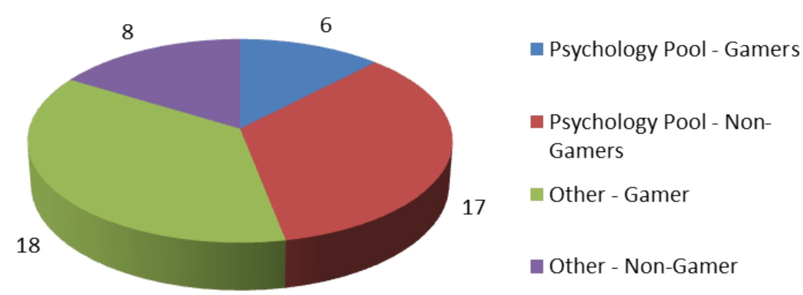

Figure 30: Distribution of gamers who were in or out of the Psychology pool.

Figure 30 shows participants that were either in our out of the participants Psychology pool and whether or not they fit the criteria of being a "gamer". While roughly half of participants are from the Psychology pool and half are gamers, among the Psychology pool participants there are only 26 percent fit the criteria for a "gamer", while 69 percent of the remaining participants fitting the criteria for being a gamer. Because the Psychology pool did not produce a high percentage of gamers, it might not be the best source of participants for a study involving video games such as ours.

6 Future Studies

As we performed our experiment, we ran into some issues and discovered a few things that we should have done differently. Section 6.1 is our reflection on our experimental procedure overall. Section 6.2 is our review of the game we used and how it could have been better. Section 6.3 is about where this project could go next. Section 6.4 is about some other practical technologies for conducting experiments involving cloud gaming.

6.1 Revisions to the Experimental Procedure

While conducting our experiments, we observed a few things that could have been improved in our experimental procedure. The first was that our experiment took 30 minutes per user to finish only ten trials. The same number of trials could have been completed faster with a more efficient introduction and game-play procedure, as well as a game more properly suited to repeated trials. More often than not, our participants were getting bored or frustrated multiple sessions before finishing, and this could have had an adverse effect that could have been prevented. Additionally, we could have had more participants if our sessions were shorter.

Another thing that would have improved our efficiency is automated player score recording. Studies like [8, 9] used programs like FRAPS to take screen shots while the user is playing. We could have picked up the pace of our experiments with automation such as this and also reduced human error.

6.2 Game Selection

In section 3.1 we listed out the criteria that was used when deciding which game to use for the study. Going off these, we eventually chose the game Crazy Taxi. However, after completion of the study, there are a few attributes of the game that, if they had been different, could have improved the quality of the data collected.

6.2.1 The Score

Player's score was determined by the number of customers delivered, the distance traveled from the customer to their destination, the time it took to reach the destination, and any tricks the player may have done along the way. The largest factor of these was the number of customers delivered. Given 35 seconds starting time, the players were able to deliver, on average, 0-3 customers before the round was over. The other factors made enough of a variation to distinguish them amongst players who delivered the same number of customers, but not enough to cause outliers in the three groups. This made for generalized data and it was not fine grained. Choosing a game with a more precise scoring system would be able to more accurately gauge player performance at different latencies.

6.2.2 Game Difficulty

During our initial pilot studies, we looked for games with a low skill curve, that is, the how difficult it is for a completely new player to pick up the game mechanics and perform well in the game with little information from the researchers. Our initial reaction with this game, based on first participant's reactions, was that it would be good for people with no experience with the game. However, upon implementing the study, it was discovered that a lot of the participants had a problem grasping the game mechanics and the controls. Some players actively tried to avoid obstacles like other cars or pedestrians standing on the sidewalk. In most cases, this was simply because the player was not aware that there was no penalty for hitting other cars (other than losing time) and that the pedestrians would always jump out of the way of the taxi.

This knowledge seemed obvious, though it clearly was not to these participants. This, combined with the low time given to get points in each trial, made the game much more difficult for some participants. This is possibly due to the fact that we received participants that were far outside our projected demographic. Some of the participants spent little to no time playing video games, which would cause them difficulty with this game. Choosing a simpler game, may have led to higher player scores which could have provided more meaningful data.

6.2.3 Finer Controls

The player drives the taxi using the joystick on the OnLive game pad; pressing left will turn the taxi left, and pressing right will turn it right. The problem is with the precision of the steering. Crazy Taxi was originally created as an arcade game in 1999 and the ported version on OnLive is the same in game play and mechanics. This version of the game had binary steering, meaning that as long as the game pad stick is moved the taxi will always turn the same amount, no matter how much or how little pressure is applied to the stick. This makes it more difficult to finely control the taxi and it has less precise movements.

Another hindrance for some players was manual shifting between drive and reverse. In sticking with the original arcade machine's design of a forward and reverse "stick shift", the player would press "A" to shift to reverse and "B" to shift to drive. Some racing games will have the player automatically move in reverse when fully stopped or have the reverse function mapped to a different button. It took some time for a few of the participants to become familiar with this mechanic and quite possibly hindered their performance.

6.2.4 Recording Time

From the data collected, the numbers drawn from the scores vs latencies and survey ratings led to broad, generalized results. It would have been more beneficial to our study to have refined, more precise data to draw conclusions from. During the time of each trial, the players were about to deliver 0-3 customers to their destinations. To get better performance ratings, we could have recorded the total accumulative time it took the player to complete each trial. Each trial starts off with 35 seconds, but the total time to complete each round depended on how many customers the player picked up. The more total time it took a player to complete a trial, the more customers the player was able to pick up. This number can attributed to the player's performance and provide more precise data.

6.2.5 Alternative Game Modes

Console versions of Crazy Taxi feature alternative game modes known as Crazy Box. This is a set of minigames that feature challenges like picking up and dropping off a number of customers within a time limit, bowling using the taxi as a ball, and popping giant balloons in a field. While unexplored in pilot testing, the last mode provides a suitable replacement for trial testing. The player starts with 60 seconds to drive around a wide open area, running into balloons and popping them. The time and number of balloons are consistent, removing complicated variables.

6.2.6 Modern Game Graphics

One of the questions on the survey that each participant filled out after each trial inquired about the participants opinion on the quality of the graphics. Latency can affect the frame rate, which in turn would affect the quality of the graphics, though this was only seen in small number of cases. That being said, a majority of the participants never rated the quality of the graphics above a 5 (with 7 being the best). It is quite possible that when asked about the quality of the graphics, those participants were basing their answers in relation to modern day video game graphics. These answers may have been biased. Crazy Taxi was released in 1999, at time where 3D polygon models were improving, but nowhere near the standards of today's models. The easiest way to avoid this bias would have been to tell the participants to choose their answers relative to the baseline trial. Another way that may have fixed this bias would be to pick a game with higher quality graphics.

6.3 Continued Research

This experiment was designed to use the OnLive service without modification to the service or games. The only modification is the external control of network conditions. With this framework, any game provided by OnLive could be used to duplicate this test, as long as it fits the criteria we described in Section 4.1. If the test is repeated with different latency conditions, the survey will still apply, and only minor changes need to be applied to conduct the test with changing packet loss or other changes in network conditions. The next step in continuing this research would be to test with other game types at more latencies or varying packet loss settings and compare those results to those presented in this paper.

In order to fully prove or disprove the hypotheses presented in Section re3, it will be necessary to repeat this experiment with smaller changes in latency at both extremes of the latency scale that we used. If the correlations are still linear, then our hypotheses will be undoubtedly proven wrong, but the results of our experiment are not conclusive in this respect.

6.4 Other Cloud Gaming Options

While our study has been conducted on OnLive and proven results for the service, our study was designed with the cloud gaming service acting as a "black box" so that it can be repeated with these different technologies to know if the results will differ. For these future studies in the area of cloud gaming, latency, and player satisfaction, we include the following section that gives a brief overview of various other cloud gaming infrastructures. We considered these technologies either while preparing for or conducting our study.

6.4.1 GamingAnywhere

GamingAnywhere is an open-source system created by Chun-Ying Huang, Cheng-Hsin Hsu, Yu-Chun Chang, and Kuan-Ta Chen who wanted to test the viability of creating a personal gaming “server”, which could send video to a client and receive controller input [19]. It functions over a Local Arena Network (LAN) and over the Internet, and is compatible with any video game, given the correct configuration. We recommend that an interested reader see another Interdisciplinary Qualifying Project, “Network Latency and Cloud Games: A Study Using GamingAnywhere” by Anouna and Estep and French, which was done at the same time as ours using the GamingAnywhere platform [20].

6.4.2 GameNow

GameNow[21], formerly UGameNow, is a cloud gaming service that functions on many devices offering timed free trials of many different video games that range from “hardcore” to “casual” and some Massively Multiplayer Online (MMO) games. The PC client is in the form of a browser plug-in that installs and allows playing free trials in the browser. There also exists Google TV and Smart TV clients, as well as a Verizon mobile device client through an exclusivity deal.

6.4.3 Steam In-Home Streaming

Steam In-Home Streaming [22], at the time of this writing, is currently still in the beta stage, but is similar to GamingAnywhere or Remote Play by Sony/Gaikai mentioned below. Steam In-Home Streaming uses a host computer and a client computer on the same Local Area Network (LAN) and streams video and input over the network. It streams any normal Steam games and non-Steam shortcuts that have been added.

6.4.4 Sony/Gaikai

Gaikai is a subsidiary of Sony that previously provided 30 minute web demos of video games through the web browser, and offered partnerships to stream games to devices. After being purchased by Sony, Gaikai’s technology is now on the PlayStatation 4 and being used to create the Remote Play and PlayStation Now services.

Remote Play is a service that allows streaming from a PlayStation 4 to a compatible PlayStation Portable or a PlayStation Vita with a supported game over a private home network or the Internet [23]. It would be possible to configure the router to create an added delay, or add a delay between the PlayStation 4 to the router using a similar type of DummyNet setup that we used for our research.

PlayStation Now, once released, will be a service that allows older titles that were available on the PlayStation, PlayStation 2, or PlayStation 3 available for purchase or rental on a PlayStation 4, PlayStation Vita, and Sony BRAVIA Televisions by streaming them over the Internet [24]. If an experiment was to be conducted in this fashion, the setup would be identical to the one used in our experiment, switching the micro-console for a full console and switching the OnLive service for the Sony/Gaikai service.

6.4.5 Kalydo

Kalydo is a unique cloud provider in the sense that it does not do “video” streaming, but rather “file” streaming [25, 26]. This allows any user who has downloaded the Kalydo player to instantly start playing a game with the main files that are required immediately to the computer, such as the first level, and the rest are streamed on-demand. This difference allows them to work better on low bandwidth connections, and makes them less likely to suffer issues directly related to latency. Even though this isn't a typical thin-client, it still is a thin-client since the file-storage is not performed on the client-side.

7 Conclusion

Because of the nature of thin clients, latency is always present. Because the effects of latency on players is not fully understood, we developed an experiment to study the effects of latency on player experience as well as performance. To prepare for the experiment, we researched other studies centered around video games. While researching how factors like screen resolution impact player performance, we came up with the following hypotheses as described in Section 3:

1. As latency increases, player performance will decrease.
2. As latency increases, player satisfaction will decrease.
3. There is a range of best-case latencies that a user will not be able to differentiate between.
4. There is a point where any further decreases in latency will cause a user to consider a game unplayable.

Next we designed an experiment where we would control how much additional latency would be added to a thin client system while a participant plays a game. Our full experiment design is documented in Section 4, including the approval processes and participant solicitation that we did. During the experiment, we collected demographic information, scores from game sessions and ratings about different aspects of the experience from the participants.

Over the course of the study, we collected information supporting the first two of our hypotheses, while we were unable to find any information supporting or disproving our third and fourth hypotheses. As described in Section 5.1.1, and displayed by Figures 7, 8, 10 in Section 5.1.1, we were able to determine from the survey data collected that ratings given by participants decreased linearly as latency increased:

- Average Responsiveness
- Average Enjoyability
- Quality of Experience

The information presented in Figure 14 from Section 5.1.2 lends support to our first hypothesis, as it shows a negative trend between latency and player performance. Participants average scores decreased as the latency increased. The graph shows a non-linear relationship where participant's score did not always decrease at the same rate. Latencies within +/- 25ms of each other did not have consistent drops and stayed relatively similar, save for the cases at 0ms, 75ms, and 150ms. Figure 13 in Section 5.1.2 shows the average scores for participants after each trial. Here, participant scores did not have a negative trend. This supports the notion that players were able to grasp a greater mastery of the game controls and mechanics enough to have little to no effect on their performance. Even so, the survey results show that participants felt a decrease in their performance at latency increased, while their actual scores did not.

The information presented in Section 5.1.1, Figure 8 similarly lends support to our second hypothesis, as there is a clear negative linear relationship presented. The third and fourth hypotheses expect the line to have a "falloff", or a point where the line would appear asymptotic, on both ends of the trend line in Figure 7 from Section 5.1.1. As shown in the previous figures, our data range of 0 ms to 150 ms does not display any such asymptotic behavior, and only suggests that the negative linear trend will continue.

While we can positively say that increased latency has a negative effect of participant performance and experience, it can also be concluded that there is a psychological effect on the participants. Early on, many participants were able to notice the latency and it made them feel as though they performed worse than previously, even if that was not always the case. For example, in the case of trial 3 at 25ms, participants' scores were higher than trial 1 at 0ms. However, many users rated their performance worse at Trial 3 than they did at Trial 1.

As part of our analysis, we looked at the effectiveness of our advertising methods. We looked at the demographics of our users, and determined that the amount of participants that fit the criteria of a "gamer". From the Psychology pool there were significantly fewer "gamers" as compared to the rest of the population as described in Section 5.6. While our projected demographic had a significant portion of "gamers", we achieved an even split between "gamers" and "non-gamers". The participants from the Psychology pool primarily consisted of "non-gamers" and participants from other sources were primarily "gamers".

As thin clients increase in popularity, these findings will be useful when considering the deployment or usage of video games using thin clients. Our results will prove useful to game designers and other researchers of cloud streaming technology. When game designers make the important decision of what medium to distribute their games, and consider cloud gaming, our results will provide insight into the unique problem of latency that can occur when using thin clients and how a player's experience will be impacted. Our analysis will prove useful to developers in determining when a game, with a similar deadline and precision mapping to Crazy Taxi and provided through a thin client, would give a player an experience that does not meet expectations at a given latency. This information is also useful to players attempting to understand the impact of latency on their performance and experience while playing a game through a thin client. Due to the successful design of our study, the procedure can be re-used to study different services or technologies, different network conditions and almost any game. Our research will help others who wish to explore similar topics in the area of cloud gaming.

References

[1] C. Roper, "GDC 09: OnLive Introduces The Future of Gaming," March 2009. http://www.ign.com/articles/2009/03/24/gdc-09-onlive-introduces-the-future-of-gaming.

[2] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hobfeld, "Gaming in the Clouds: QoE and the Users' Perspective," in *Mathematical and Computer Modelling*, December 2011.

[3] M. Claypool, D. Finkel, A. Grant, and M. Solano, "Thin to Win? Network Performance Analysis of the OnLive Thin Client Game System," in *Proceedings of the 11th ACM Network and System Support for Games (NetGames)*, November 2012.

[4] Y.-T. Lee, K.-T. Chen, H.-I. Su, and C.-L. Lei, "Are All Games Equally Cloud-Gaming-Friendly? An Electromyographic Approach," in *Proceedings of IEEE/ACM NetGames*, Oct 2012.

[5] S. Nunneley, "GameStop Closes Spawn Labs Because Consumers Aren't Ready for Cloud Gaming," March 2014. http://www.vg247.com/2014/03/27/gamestop-closes-spawn-labs-because-consumers-arent-ready-for-cloud-gaming/.

[6] S. Choy, B. Wong, G. Simon, and C. Rosenberg, "The Brewing Storm in Cloud Gaming: A Measurement Study on Cloud to End-User Latency," in *NetGames*, pp. 1–6, IEEE.

[7] M. Jarschel, D. Schlosser, S. Scheuring, and T. Hobfeld, "An Evaluation of QoE in Cloud Gaming Based on Subjective Tests," in *Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2011.

[8] K. Claypool and M. Claypool, "On Frame Rate and Player Performance in First Person Shooter Games," in *Springer Multimedia Systems Journal (MMSJ)*, vol. Volume 13, Number 1, pp. 3–17, 2007.

[9] M. Claypool, K. Claypool, and F. Damaa, "The Effects of Resolution on Users Playing First Person Shooter Games," in *Proceedings of ACM/SPIE Multimedia Computing and Networking (MMCN)*, (San Jose, California, USA), January 2006.

[10] V. Clincy and B. Wilgor, "Subjective Evaluation of Latency and Packet Loss in a Cloud-based Game," (Las Vegas, NV), April 2013.

[11] M. Claypool and K. Claypool, "Latency Can Kill: Precision and Deadline in Online Games," in *Proceedings of the First ACM Multimedia Systems Conference*, (Scottsdale, AZ, USA), February 2010.

[12] M. Claypool and K. Claypool, "Latency and Player Actions in Online Games," *Commun. ACM*, vol. 49, pp. 40–45, Nov. 2006.

[13] T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, and M. Claypool, "The Effects of Loss and Latency on User Performance in Unreal Tournament 2003," 2003.

[14] DigitalFoundry, "Digital Foundry vs. OnLive Page 4," April 2014. Eurogamer.net.

[15] WPI, "Netops," April 2014. https://www.wpi.edu/Academics/CCC/Netops/Network/WPI-Network.jpg.

[16] WPI, "Institutional Review Board," December 2013. http://www.wpi.edu/offices/irb.html.

[17] Canadians, "Online Time Slot Management," December 2013. Slottr.com.

[18] Qualtrics, "WPI Qualtrics Control Panel," February 2014. https://wpi.qualtrics.com/.

[19] C.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, "GamingAnywhere: An Open Cloud Gaming System," in *Proceedings of ACM Multimedia Systems 2013*, Feb 2013.

[20] J. Anouna, Z. Estep, and M. French, "Network Latency and Cloud Games: A Study Using GamingAnywhere," 2014. IQP MLC - TC14.

[21] GameNow, "Gamenow," March 2014. http://about.ugamenow.com/about.html.

[22] Valve, "In-Home Streaming - General Troubleshooting - Knowledge Base - Steam Support," March 2014. https://support.steampowered.com/kb_article.php?ref=3629-RIAV-1617.

[23] Sony, "Remote Play for PS3 and PSP," March 2014.

[24] Sony, "PlayStation Now - Streaming Game Service on Consoles," March 2014. http://us.playstation.com/playstationnow/.

[25] Kalydo, "Technology — Kalydo," March 2014. http://kalydo.com/technology/.

[26] "Kalydo Demolishes Entry Barriers to True Cloud Gaming," March 2014. http://gamasutra.com/view/pressreleases/195826/.

A Figures

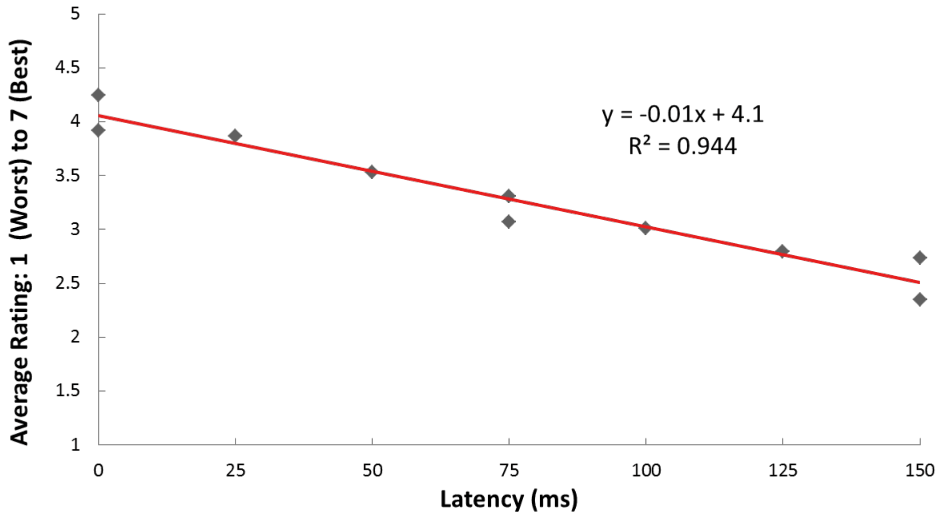

Figure 31: Average combined responsiveness and enjoyability rating.

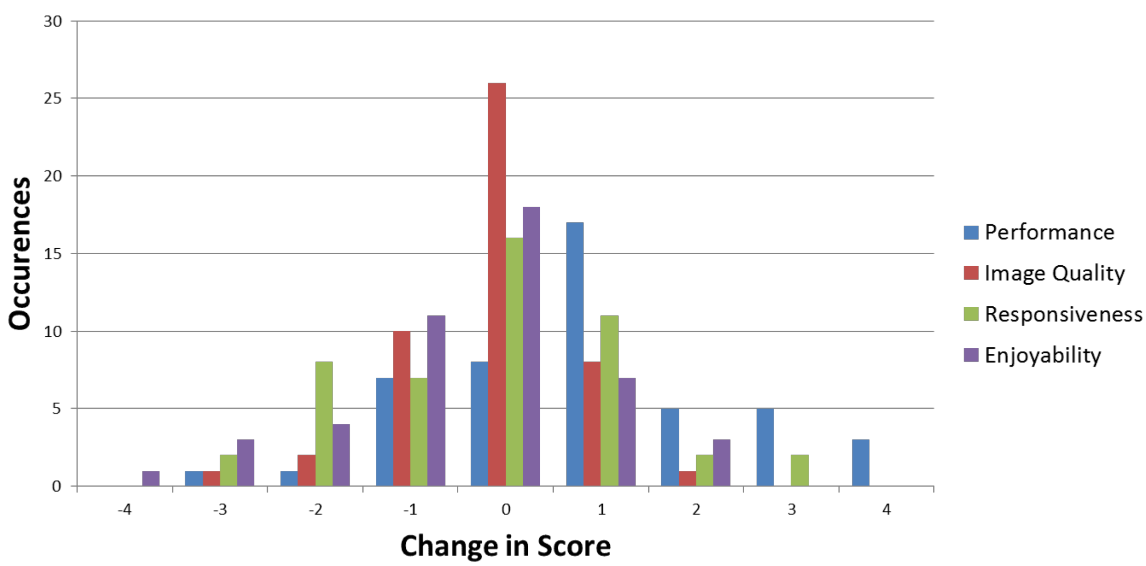

Figure 32: Change in ratings at 0 milliseconds.

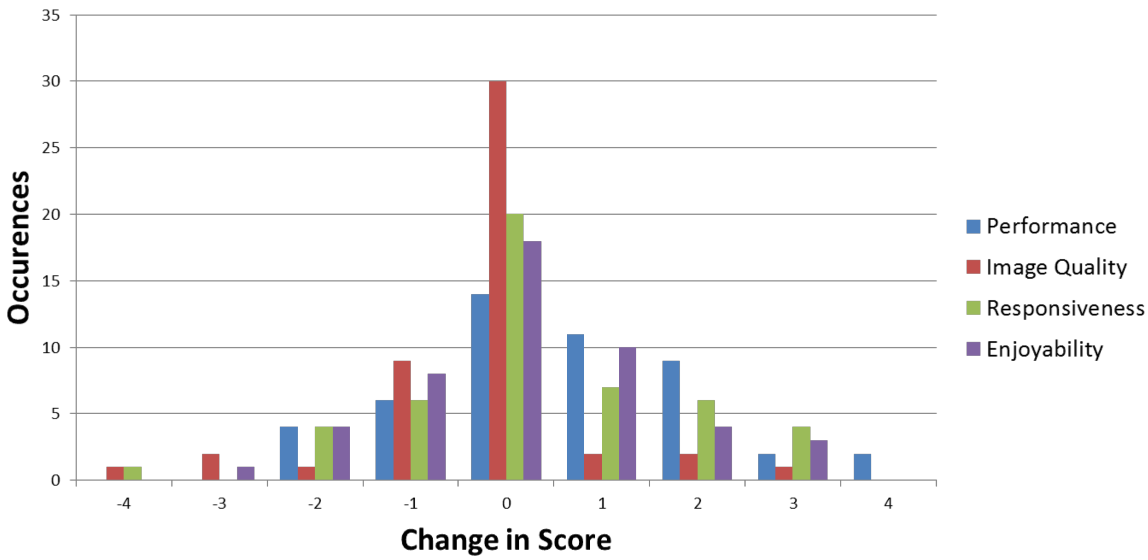

Figure 33: Change in ratings at 75 milliseconds.

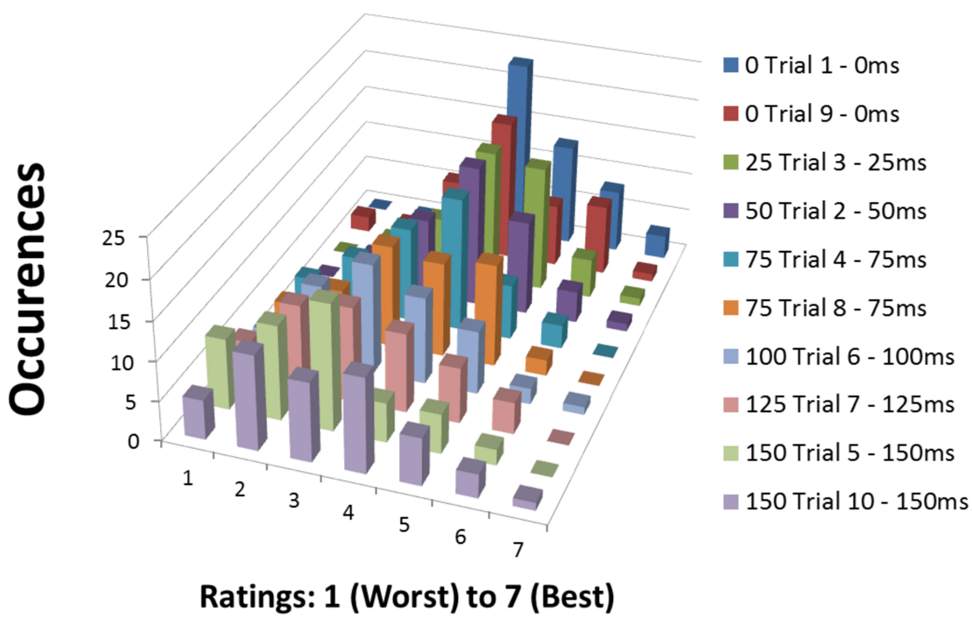

Figure 34: Distribution of enjoyability ratings by trial ordered by latency.

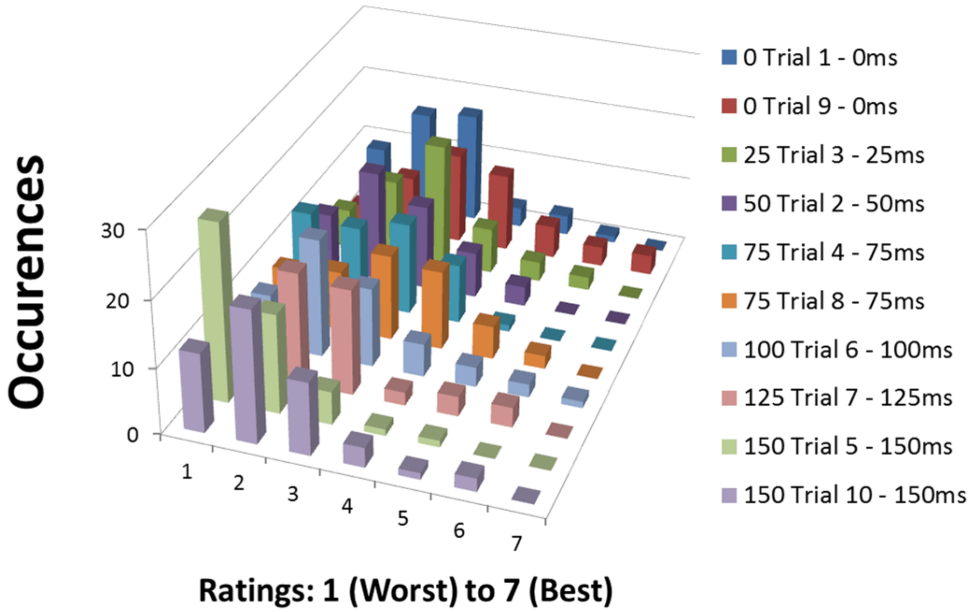

Figure 35: Distribution of self performance ratings by trial ordered by latency.

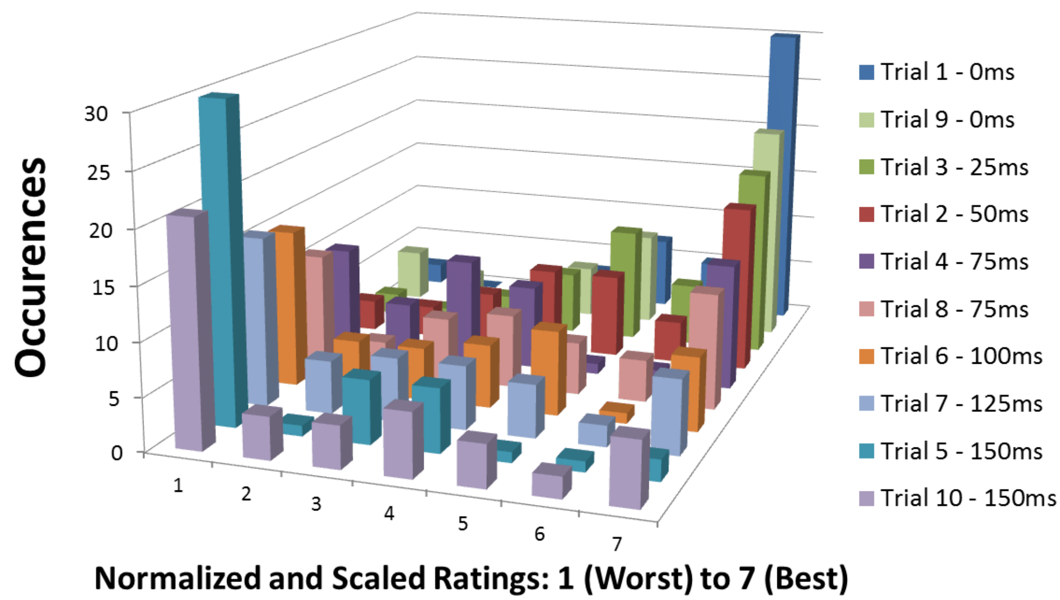

Figure 36: Distribution of normalized enjoyability ratings by trial ordered by latency.

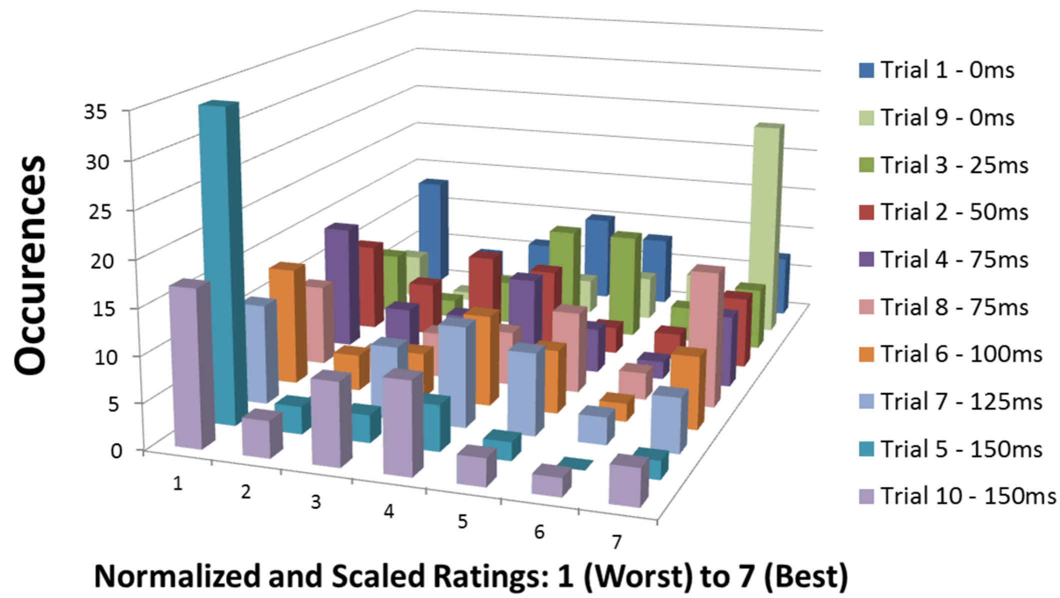

Figure 37: Distribution of normalized self performance ratings by trial ordered by latency.

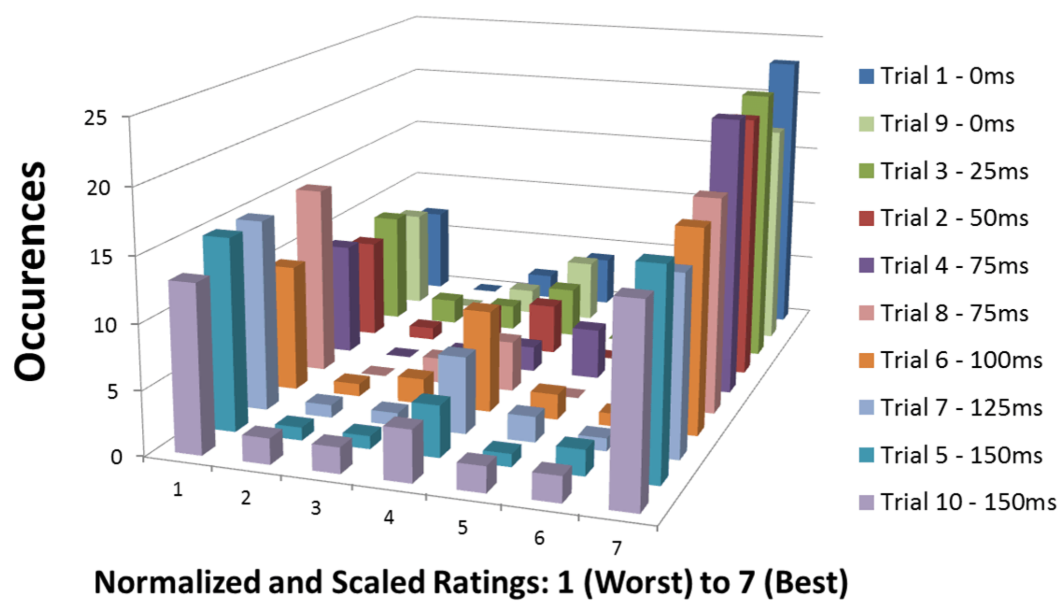

Figure 38: Distribution of normalized image quality ratings by trial ordered by latency.

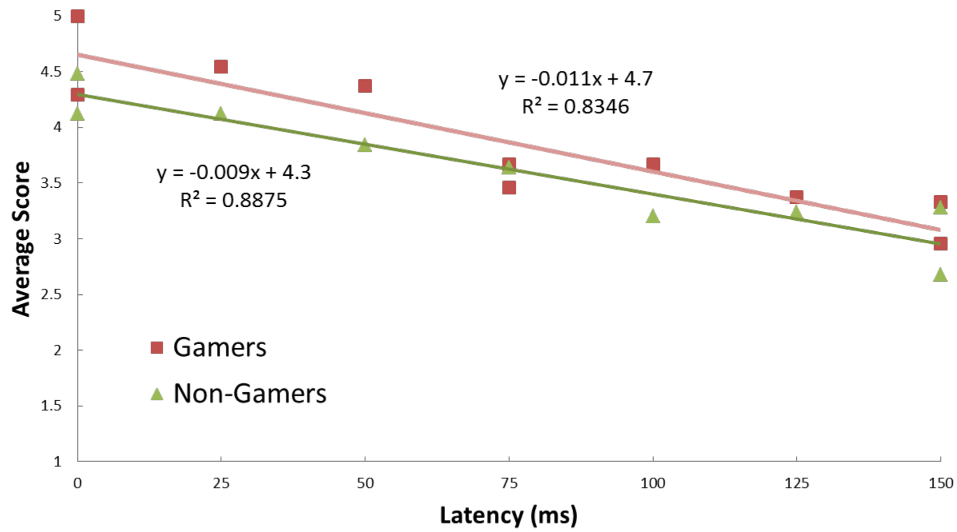

Figure 39: Average enjoyability ratings for gamers and non-gamers.

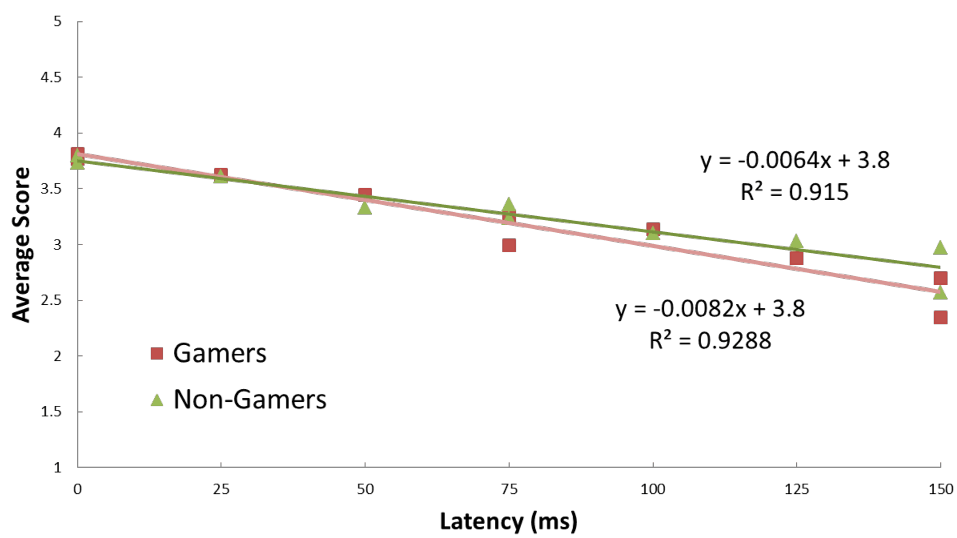

Figure 40: Average QoE ratings for gamers and non-gamers.

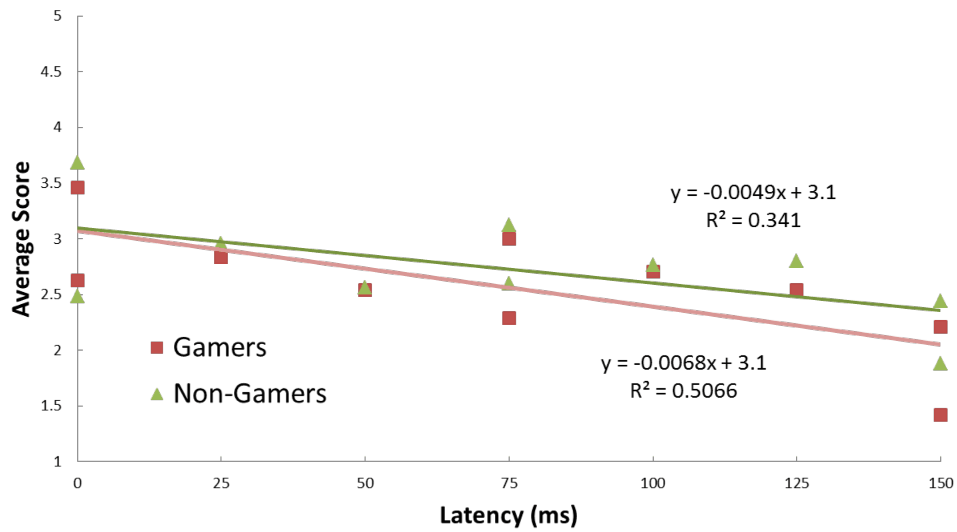

Figure 41: Average performance ratings for gamers and non-gamers.

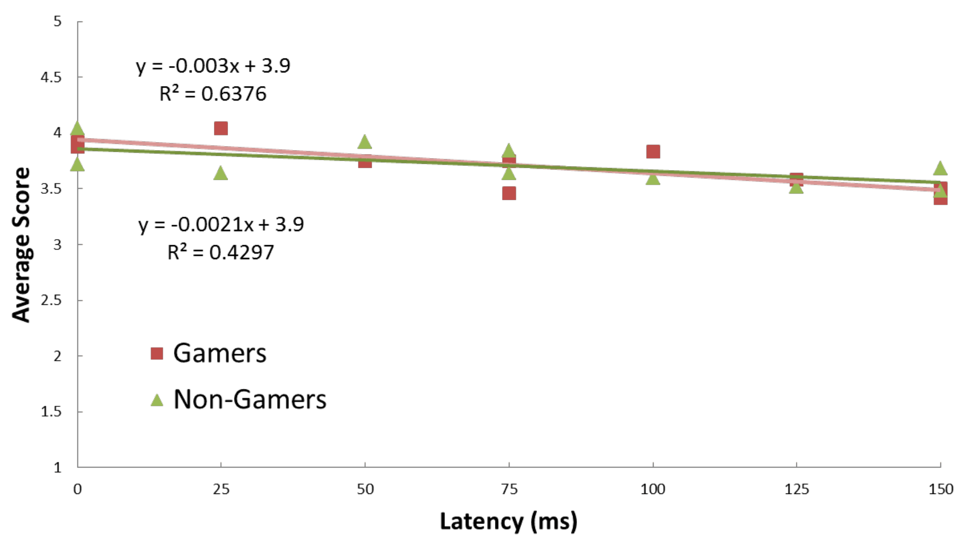

Figure 42: Average image quality ratings for gamers and non-gamers.

B Tables

Latency (ms)	0	0	25	50	75	75	100	125	150	150
Trial Number	**1**	**9**	**3**	**2**	**4**	**8**	**6**	**7**	**5**	**10**
Responsiveness	3.76	3.63	3.41	3.06	2.67	2.96	2.59	2.29	1.88	2.16
Image Quality	3.96	3.82	3.84	3.84	3.80	3.55	3.71	3.55	3.49	3.55
Enjoyability	4.73	4.20	4.33	4.08	3.54	3.65	3.43	3.31	2.82	3.31
Performance	2.55	3.57	2.90	3.96	2.45	3.06	2.73	2.67	1.65	2.33
R & E	4.24	3.92	3.87	3.57	3.11	3.31	3.01	2.80	2.35	2.73
QoE	3.75	3.81	3.62	3.74	3.12	3.31	3.12	2.95	2.46	2.84

Table 5: Averages for the answers to each question by latency for each trial. "E & R" is the average of Responsiveness and Enjoyability. "QoE" is calculated by taking the average of the average answers to the original four questions.

Latency (ms)	0	0	25	50	75	75	100	125	150	150
Trial Number	**1**	**9**	**3**	**2**	**4**	**8**	**6**	**7**	**5**	**10**
Average Score	295	406	314	276	255	320	246	238	125	169
Avg. Customers	1.12	1.49	1.18	1.06	1.00	1.22	1.00	0.96	0.57	0.69

Table 6: The average player scores and number of customers delivered per trial sorted by latency.

Latency (ms)	0	0	25	50	75	75	100	125	150	150
Trial Number	**1**	**9**	**3**	**2**	**4**	**8**	**6**	**7**	**5**	**10**
QoE	3.73	3.80	3.61	3.33	3.24	3.36	3.10	3.03	2.57	2.97
Performance	2.48	3.68	2.96	2.56	2.60	3.12	2.76	2.80	1.88	2.44
Image Quality	4.04	3.72	3.64	3.92	3.84	3.64	3.60	3.52	3.48	3.68
Responsiveness	3.92	3.68	3.72	3.00	2.88	3.04	2.84	2.56	2.24	2.48
Enjoyability	4.48	4.12	4.12	3.84	3.64	3.64	3.20	3.24	2.68	3.28
Score	233	389	290	207	226	289	181	198	98	151

Table 7: The average scores and ratings of participants considered to be non-gamers.

Latency (ms)	0	0	25	50	75	75	100	125	150	150
Trial Number	**1**	**9**	**3**	**2**	**4**	**8**	**6**	**7**	**5**	**10**
QoE	3.77	3.81	3.63	3.45	2.99	3.25	3.14	2.88	2.34	2.70
Performance	2.63	3.46	2.83	2.54	2.29	3.00	2.71	2.54	1.42	2.21
Image Quality	3.88	3.92	4.04	3.75	3.75	3.46	3.83	3.58	3.50	3.42
Responsiveness	3.58	3.58	3.08	3.13	2.46	2.88	2.33	2.00	1.50	1.83
Enjoyability	5.00	4.29	4.54	4.38	3.46	3.67	3.67	3.38	2.96	3.33
Score	359	424	339	348	285	352	314	279	153	189

Table 8: The average scores and ratings of participants considered to be gamers.

	Gamers			Non-Gamers		
Rating	**Slope**	**Y-intercept**	**CoD**	**Slope**	**Y-intercept**	**CoD**
Responsiveness	-0.013	3.6	0.956	-0.010	3.8	0.9278
Enjoyability	-0.011	4.7	0.8346	-0.009	4.3	0.8875
Performance	-0.0068	3.1	0.5066	-0.0049	3.1	
Image Quality	-0.003	3.9	0.6376	-0.0021	3.9	0.4297
QoE	-0.0082	3.8	0.9288	-0.0064	3.8	0.915
Score	-1.29	401	0.7928	-1.16	313	0.6416

Table 9: Summary table of trend lines for gamers and non-gamers.

C Documents

The following documents are included in this section of the Appendix.

- IRB Expedited Review Form
- Consent Form
- Experimental Procedure
- Survey
- Psychology Department Participant Pool Application

WPI IRB use only
IRB #
Date:

WORCESTER POLYTECHNIC INSTITUTE
Institutional Review Board
Application for Approval to Use Human Subjects in Research

This application is for: *(Please check one)* ☒ Expedited Review ☐ Full Review

WPI IRB use only

Principal Investigator (PI) or Project Faculty Advisor: (*NOT a student or fellow; must be a WPI employee)*

Name: Mark Claypool Tel No: 1-508-831-5409 E-Mail Address: claypool@cs.wpi.edu ☐

Department: Computer Science and IMGD

Co-Investigator(s): *(Co-PI(s)/non students)*

Name: David Finkel Tel No: 1-508-831-5416 E-Mail Address: dfinkel@wpi.edu ☐

Name: Tel No: E-Mail Address: ☐

Student Investigator(s):

Name: Robert Dabrowski Tel No: 603-965-6388 E-Mail Address: rpdabrowski@wpi.ed u ☐

Name: SEE ATTACHED Tel No: E-Mail Address: onlive-iqp@wpi.edu ☐

Check if: ☒ **Undergraduate project** *(MQP, IQP, Suff., other)*
☐ **Graduate project** *(M.S. Ph.D., other)*

Has an IRB ever suspended or terminated a study of any investigator listed above?
No ☒ Yes ☐ *(Attach a summary of the event and resolution.)*

Vulnerable Populations: The proposed research will involve the following (Check all that apply):
pregnant women ☐ human fetuses ☐ neonates ☐ minors/children ☐ prisoners ☐
students ☒ individuals with mental disabilities ☐ individuals with physical disabilities ☐

Collaborating Institutions: (*Please list all collaborating Institutions.)*
n/a

Locations of Research: (*If at WPI, please indicate where on campus. If off campus, please give details of locations.)*
Fuller Labs A17, Digital Arts Studio

Project Title: The Effects of Lag on Thin Client Gaming

Funding: *(If the research is funded, please enclose one copy of the research proposal or most recent draft with your application.)*

Funding Agency: n/a WPI Fund: n/a

Human Subjects Research: (***All*** *study personnel having direct contact with subjects* ***must*** *take and pass a training course on human subjects research. There are links to web-based training courses that can be accessed under the Training link on the IRB web site* http://www.wpi.edu/offices/irb/training.html. ***The IRB requires a copy of the completion certificate from the course or proof of an equivalent program.****)*

Anticipated Dates of Research:

Start Date: 2/10/2013 Completion Date: March 7, 2013

X

WORCESTER POLYTECHNIC INSTITUTE

Institutional Review Board

Application for Approval to Use Human Subjects in Research

WPI IRB use only
IRB #
Date:

Instructions: Answer all questions. If you are asked to provide an explanation, please do so with adequate details. If needed, attach itemized replies. Any incomplete application will be returned.

1.) Purpose of Study: *(Please provide a concise statement of the background, nature and reasons for the proposed study. Insert below using non-technical language that can be understood by non-scientist members of the IRB.)*

"Thin Clients" are a video game technology that is internet based. We will be studying the effects of lag on player satisfaction while playing games hosted through a thin client. This will help us show that there is a correlation between lag and player satisfaction.

2.) Study Protocol: *(Please attach sufficient information for effective review by non-scientist members of the IRB. Define all abbreviations and use simple words. Unless justification is provided this part of the application must not exceed 5 pages. Attaching sections of a grant application is not an acceptable substitute.)*

A.) For **biomedical, engineering and related research**, please provide an outline of the actual experiments to be performed. Where applicable, provide a detailed description of the experimental devices or procedures to be used, detailed information on the exact dosages of drugs or chemicals to be used, total quantity of blood samples to be used, and descriptions of special diets.

B.) For applications in the **social sciences, management and other non-biomedical disciplines** please provide a detailed description of your proposed study. Where applicable, include copies of any questionnaires or standardized tests you plan to incorporate into your study. If your study involves interviews please submit an outline indicating the types of questions you will include.

C.) If the study involves **investigational drugs or investigational medical devices**, and the PI is obtaining an Investigational New Drug (IND) number or Investigational Device Exemption (IDE) number from the FDA, please provide details.

D.) Please note if any **hazardous materials** are being used in this study.

E.) Please note if any **special diets** are being used in this study.

3.) Subject Information:

A.) Please provide the exact number of subjects you plan to enroll in this study and describe your subject population. *(eg. WPI students, WPI staff, UMASS Medical patient, other)*

Males: 40 Females: 10 Description: WPI students

B.) Will subjects who do not understand English be enrolled?
No ☒ Yes ☐ *(Please insert below the language(s) that will be translated on the consent form.)*

C.) Are there any circumstances under which your study population may feel coerced into participating in this study?
No ☒ Yes ☐ *(Please insert below a description of how you will assure your subjects do not feel coerced.)*

D.) Are the subjects at risk of harm if their participation in the study becomes known?
No ☒ Yes ☐ *(Please insert below a description of possible effects on your subjects.)*

E.) Are there reasons for excluding possible subjects from this research?
No ☐ Yes ☒ *(If yes, please explain.)*
If a subject is unable to use a mouse and keyboard and see a computer screen, they may not participate.

F.) How will subjects be recruited for participation? *(Check all that apply.)*

☐ Referral: *(By whom)* __________
☒ Other: *(Identify)* Word of mouth
☐ Database: *(Describe how database populated)*

F.) Have the subjects in the database agreed to be contacted for research projects? No ☐ Yes ☐ N/A ☒

☒ Direct subject advertising, including: *(Please provide a copy of the proposed ad. All direct subject advertising must be approved by the WPI IRB prior to use.)*

☐ Newspaper ☒ Bulletin board
☐ Radio ☒ Flyers
☐ Television ☐ Letters
☐ Internet ☒ E-mail

G.) Are the subjects being paid for participating? *(Consider all types of reimbursement, ex. stipend, parking, travel.)*
No ☐ Yes ☒ *(Check all that apply.)* ☐ Cash ☐ Check ☐ Gift certificate ☒ Other: opportunity to win a raffle for an Amazon Giftcard, refreshments during the experiments
Amount of compensation __________

4.) Informed Consent:

A.) Who will discuss the study with and obtain consent of prospective subjects? *(Check all that apply.)*
☐ Principal Investigator ☐ Co-Investigator(s) ☒ Student Investigator(s)

B.) Are you aware that subjects must read and sign an Informed Consent Form prior to conducting any study-related procedures and agree that all subjects will be consented prior to initiating study related procedures? No ☐ Yes ☒

C.) Are you aware that you must consent subjects using only the IRB-approved Informed Consent Form? No ☐ Yes ☒

D.) Will subjects be consented in a private room, not in a public space? No ☐ Yes ☒

E.) Do you agree to spend as much time as needed to thoroughly explain and respond to any subject's questions about the study, and allow them as much time as needed to consider their decision prior to enrolling them as subjects? No ☐ Yes ☒

F.) Do you agree that the person obtaining consent will explain the risks of the study, the subject's right to decide not to participate, and the subject's right to withdraw from the study at any time? No ☐ Yes ☒

G.) Do you agree to either 1.) retain signed copies of all informed consent agreements in a secure No ☐ Yes ☒

location for at least three years or 2.) supply copies of all signed informed consent agreements in .pdf format for retention by the IRB in electronic form?

(If you answer No to any of the questions above, please provide an explanation.)

5.) Potential Risks: *(A risk is a potential harm that a reasonable person would consider important in deciding whether to participate in research. Risks can be categorized as physical, psychological, sociological, economic and legal, and include pain, stress, invasion of privacy, embarrassment or exposure of sensitive or confidential data. All potential risks and discomforts must be minimized to the greatest extent possible by using e.g. appropriate monitoring, safety devices and withdrawal of a subject if there is evidence of a specific adverse event.)*

A.) What are the risks / discomforts associated with each intervention or procedure in the study?

There will be no risks

B.) What procedures will be in place to prevent / minimize potential risks or discomfort?
Players will not sit at the game for the full duration of the study, they will take breaks to fill out short surveys every few minutes.

6.) Potential Benefits:

A.) What potential benefits other than payment may subjects receive from participating in the study?

They get to play and enjoy video games they may not have played before.

B.) What potential benefits can society expect from the study?
Video game developers and providers of thin clients will be able to look at our results and make technical decsions based on our results, which will show player satisfactions at varying network conditions.

7.) Data Collection, Storage, and Confidentiality:

A.) How will data be collected?

Qualtrics, hosted on "wpi.qualtrics.com" will be used for all data collection in the form of surveys.

B.) Will a subject's voice, face or identifiable body features *(eg. tattoo, scar)* be recorded by audio or videotaping?
No ☒ Yes ☐ *(Explain the recording procedures you plan to follow.)*

C.) Will personal identifying information be recorded? No ☐ Yes ☒ *(If yes, explain how the identifying information will be protected. How will personal identifying information be coded and how will the code key be kept confidential?)*
Only age, gender and video game habit information will be collected. Email addresses for the raffle will be collected separately from data on the consent form. No identifying personal information will be tied to experimental data.

D.) Where will the data be stored and how will it be secured?

All data will be recorded on qualtrics and only the investigators listed on this document will be given access.

E.) What will happen to the data when the study is completed?

We will state our findings in our IQP report. After that, the data will be turned over to the faculty advisor, who will securely archive it where it will not be publicly available.

F.) Can data acquired in the study adversely affect a subject's relationship with other individuals? *(i.e. employee-supervisor, student-teacher, family relationships)*

No.

G.) Do you plan to use or disclose identifiable information outside of the investigation personnel?
No ☒ Yes ☐ *(Please explain.)*

H.) Do you plan to use or disclose identifiable information outside of WPI including non-WPI investigators?
No ☒ Yes ☐ *(Please explain.)*

8.) Incidental findings: *In the conduct of information gathering, is it possible that the investigator will encounter any incidental findings? If so, how will these be handled? (An incidental finding is information discovered about a subject which should be of concern to the subject but is not the focus of the research. For example, a researcher monitoring heart rates during exercise could discover that a subject has an irregular heartbeat.)*

Incidental findings are not expected because not much personal information will be gathered.

9.) Deception: *(Investigators must not exclude information from a subject that a reasonable person would want to know in deciding whether to participate in a study.)*

Will the information about the research purpose and design be withheld from the subjects?
No ☒ Yes ☐ *(Please explain.)*

10.) Adverse effects: *(Serious or unexpected adverse reactions or injuries must be reported to the WPI IRB within 48 hours using the IRB Adverse Event Form found out at http://www.wpi.edu/offices/irb/forms.html. Other adverse events should be reported within 10 working days.)*

What follow-up efforts will be made to detect any harm to subjects and how will the WPI IRB be kept informed?
A closing survey will be held at the end of the study for each participant and the appropriate forms will be filled out if there is an incident.

11.) Conflict of Interest: *(A conflict of interest occurs when an investigator or other key personnel in a study may enjoy material benefits based on study results. Relationships that give rise to a conflict of interest or the appearance of a conflict of interest must be disclosed in the informed consent statement provided to study subjects. More information, including examples of relationships that require disclosure and those that do not, can be found here.)*

A.) Do any of the investigators listed on this application have a potential or actual conflict of interest with regard to this study?
a. Investigator (name) _Robert Dabrowski_______________________________ No ☒ Yes ☐

b. Investigator (name) _Robert Smieja_______________________________ No ☒ Yes ☐
c. Investigator (name) _Christian Manuel_______________________________ No ☒ Yes ☐

B.) If any of the answers to 11A. are "Yes," please attach an explanation of the nature of the conflict to this application and identify appropriate language for use in the consent form. Examples of consent language are found on the IRB website, here.

C.) Does each investigator named above have a current WPI conflict of interest disclosure form on file with the appropriate supervisor/department head? No ☐ Yes ☐

D.) Do any of the investigators' COI forms on file with WPI contain information regarding this research?
No ☒ Yes ☐
a. If "Yes," identify the investigator(s) _benign_________________________________

12.) Informed consent: (*Documented informed consent must be obtained from all participants in studies that involve human subjects. You must use the templates available at http://www.wpi.edu/offices/irb/forms.html to prepare these forms.* ***Informed consent forms must be included with this application.*** *Under certain circumstances the WPI IRB may waive the requirement for informed consent.)*

Investigator's Assurance:

I certify the information provided in this application is complete and correct.

I understand that I have ultimate responsibility for the conduct of the study, the ethical performance of the project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the WPI IRB.

I agree to comply with all WPI policies, as well all federal, state and local laws on the protection of human subjects in research, including:

- ensuring the satisfactory completion of human subjects training.
- performing the study in accordance with the WPI IRB approved protocol.
- implementing study changes only after WPI IRB approval.
- obtaining informed consent from subjects using only the WPI IRB approved consent form.
- promptly reporting significant adverse effects to the WPI IRB.

Signature of Principal Investigator ________________________________ Date ____________

Print Full Name and Title ____

Please return a signed hard copy of this application to the WPI IRB c/o Ruth McKeogh 2nd Floor Project Center
Or email an electronic copy to irb@wpi.edu
If you have any questions, please call (508) 831-6699.

Informed Consent Agreement for Participation in a Research Study

Investigators: Robert Dabrowski, Robert Smieja, Christian Manuel
Contact Info: onlive-iqp@wpi.edu
Title of Research Study: The Effects of Lag on Thin Client Gaming

Introduction You are being asked to participate in a research study. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

Purpose of the study: The purpose of this experiment is to determine player satisfaction while playing video games using a game providing technology known as a "thin client".

Procedures to be followed: You will play a video for a series of 15 short game sessions and take surveys between each session. Each game session will last approximately 1 to 2 minutes.

Risks to study participants: There are no foreseeable risks from this experiment.

Benefits to research participants and others: The overall results of the experiment will provide video game providers using thin client technology insight into how they should optimize their products.

Record keeping and confidentiality: All data collected will be only accessible by the above stated investigators. Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or it's designee and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you.

Compensation or treatment in the event of injury: There is no risk of injury so no compensation will be provided in the event of injury. You do not give up any of your legal rights by signing this statement.

Cost/Payment: Refreshments are provided as you play and you may enter to win an $50 Visa gift card.

For more information about this research or about the rights of research participants, or in case of research-related injury, contact: See top of page for reference to the investigators.

IRB Chair: Professor Kent Rissmiller, Tel. 508-831-5019, Email: kjr@wpi.eduand
University Compliance Officer: Michael J. Curley, Tel. 508-831-6919, Email: mjcurley@wpi.edu

Your participation in this research is voluntary. Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

By signing below, you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

______________________________ Date: ____________________
Study Participant Signature

____
Study Participant Name (Please print)

____ Date: ____________________
Signature of Person who explained this study

____
E-mail address if you would like to be entered in the raffle.

Before The Subject Arrives

- Connect OnLive box through laptop network control workstation
- Load Crazy taxi on the OnLive box
- Ensure proper difficulty settings in Crazy Taxi
- Have consent forms ready
- Have refreshments available
- Make sure the 3-D printer is turned off

When The Subject Arrives

"Hello! You are about to participate in a study to determine player satisfaction of a video game technology. Here is your consent form ~~~ Hand them the consent form ~~~ please to read it, then sign the back. You may include your email address if you wish to be entered in the raffle for the gift card.

Please start by filling out the first page of the survey only. It is for demographic purposes. Before you begin, please make yourself comfortable and adjust the TV volume. Refreshments are on the table in the center of the room. If you need to use the restroom, now is the time.

For the experiment, you will be playing Crazy Taxi, in which you will be a taxi driver and the goal is to find customers, pick them up, and deliver them to their destinations. The controls can be seen on the paper next to the controller.

The first part is a practice round of Crazy Taxi with the **best game conditions, later the conditions may vary.** Now is the time to practice the controls."

Let the user play a regular game of < 3 minutes = 1 normal game

The Experiment

"For the rest of the experiment, you will play ten 35-60 second game sessions and fill out a short survey immediately after each session. The short surveys are each associated with one page of the survey that you started before. Please wait while I set up the first test. After which you may begin."

Get the first latency of the schedule ready, run it, and tell them to start playing.

At the 35 second mark, take a screenshot of the score and have the user stop playing and fill out the survey. While they fill out the survey, start the next latency in the schedule AND set up the next game.

Repeat until all scheduled latencies/sessions have been played out.

Anticipated Questions

Depending on the question, the answer will usually be: “we will explain that aspect of the experiment once you have finished all ten trials.”

What game conditions will be changing for each test?

We will tell you at the end of the experiment.

What is the best way to play the game? How do I win?

Pick up customers and deliver them to the destinations.

What is OnLive?

A video game provider.

Can I go to the bathroom?

Please wait until the end of your current session

Debriefing statement/Closing Statement

“In this study, we tested the effects internet latency has on player performance and experience. You played Crazy Taxi because of its simplicity and inherent precision and deadline characteristics, meaning that it can require many commands very quickly to do well.

From your answers and scores, we aim to draw a correlation between user experience and performance against internet latency of game streaming. If you have any questions about the study or procedure, feel free to ask them now. Because other students may be participating in this study in the future, we ask that you not discuss the details of this study with your friends or classmates”

Demographic Information

Browser Meta Info

This question will not be displayed to the recipient.
Browser: **Chrome**
Version: **31.0.1650.63**
Operating System: **Windows NT 6.1**
Screen Resolution: **1920x1080**
Flash Version: **11.9.900**
Java Support: **1**
User Agent: **Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/31.0.1650.63 Safari/537.36**

The following questions are optional.

What is your gender?

What is your major? Select all that apply.

- Aerospace Engineering
- Biomedical Engineering
- Chemical Engineering
- Civil & Environmental Engineering
- Computer Science
- Electrical & Computer Engineering
- Environmental Engineering
- Fire Protection Engineering
- Industrial Engineering
- Interactive Media & Game Development

What is your age?

How often do you play video games? (choose one)

- Never
- 1 to 5 hours per week
- 5 to 10 hours per week
- 10 to 15 hours per week
- 15 to 25 hours per week
- 25+ hours per week

XXI

What kinds of video games do you play the most?

What kinds of video games do you play the most?

- FPS
- MOBA
- MMORPG
- RPG
- RTS
- Racing
- Strategy
- Sports
- Simulation
- Other

How much past experience do you have with video games?

No Experience	Small Amount	Average	More Than Average	A Lot of Experience

How did you hear about this experiment?

- Email
- Posters
- Psychology Department
- Word of Mouth
- Other

Trial 1

Trial #1

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 2

Trial #2

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 3

Trial #3

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 4

Trial #4

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 5

Trial #5

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)

Trial 6

Trial #6

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)

Trial 7

Trial #7

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 8

Trial #8

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 9

Trial #9

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Trial 10

Trial #10

How would you rate the quality of the game's graphics and visuals?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How well do you think you preformed?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How would you rate the responsiveness of the controls?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

How enjoyable was this experience?

1 (Worst)	2	3	4	5	6	7 (Best)
○	○	○	○	○	○	○

Application for Access to the
Social Science Participant Pool

To apply to have access to the Social Science Participant Pool you must be affiliated a faculty member, graduate student, postdoctoral fellow, visiting researcher, visiting scholar, IQP or MQP student. If you fall into one of these categories, please complete and include:

☐ **copy of IRB certificate of protocol approval (must submit a new copy with application each quarter)**
☐ **a short debriefing statement** (Debriefing statements will be made available for the participants after they've completed their experiment requirement. You still need to debrief your participants at the end of each session.)

Return this form TYPED to the Participant Pool Coordinator, Jeanine Skorinko, SL 317C
You may email any questions and the application to Skorinko@wpi.edu

Term/Year Using Pool: ____

Name: ____

Phone:

Email:

Standing: ☐ **UG** ☐ **G** ☐ **Post-Doc** ☐ **Faculty** ☐ **Visiting Scholar** ☐ **Other**

Faculty Sponsor/ Principle Investigator:

Protocol Title & # (can be found on IRB approval):

____

Protocol Expiration Date:

Length of Study (How long it actually takes to complete in minutes or hours): ___________________

Less than 30 minutes: ☐**YES** ☐**NO**
Between 35-45 minutes: ☐**YES** ☐**NO**

Research Assistant(s):

Research Assistant(s) email addresses: ____

Desired Experiment Name(s):

Experiment Location(s):

Brief Experiment Description(s) that will be posted on Sona:
____
____

Do You Need to Participate in Pre-Selection: ☐**YES** ☐**NO**

If yes, what are you selecting for?
(Please check the demographics form to see if it can be covered there instead.)

If yes, a copy of your questionnaire must be attached to this form. (You must include your protocol #, length of time to take your questionnaire (under two minutes for two sides).