The Effects of Delay and Jitter on Cloud-Based Game Streaming Players

Submitted to the Faculty of
Worcester Polytechnic Institute
in partial fulfillment of the requirements for the Degree of Bachelor of Science.

Clark Apuy, Botao Han
Advisors: Mark. Claypool

Abstract

With the increase in network capacities, the quality of user experience (QoE) of network games with delay and jitter has become important. This paper establishes the quantitative relationship between delay and jitter and QoE via a user study. We applied varied combinations of delay and jitter to the users playing games and asked participants to describe the experience. Analysis of the data shows an inverse linear relationship between jitter and QoE but a varied relationship between delay and QoE. The results may be useful for game designers and network engineers building and maintaining network games.

Table of Contents

List of Figures

List of Tables

1. Introduction

Traditionally, gaming systems have used thick clients to operate. That is to say; gaming has largely remained based on local hardware, manufactured consoles, or computers purchased and used by consumers. The advent of online gaming through connection to the Internet introduced the role of a server, which communicates with "clients" to establish and facilitate an environment where the clients interact in real-time conditions.

Platforms developed for cloud gaming, known as "thin clients," have not yet reached the widespread success that thick clients have been observed to experience but are still growing. Existing and active thin clients include Amazon Luna, Microsoft's Xbox Cloud Gaming, Sony's PlaystationNow, Nvidia's GeForce Now, and the target platform of this project, Google Stadia.

Cloud gaming is a divergence from the standard gaming architecture of thick clients that continues to prevail. Cloud gaming is advantageous in several ways to gaming via thick clients. Thin clients ensure that all necessary and cumbersome processing is left to the cloud, performed on the server rather than on any client hardware. A crucial aspect of the appeal of cloud gaming is that game computations are entirely removed from the client side. The only required devices are those that communicate inputs to and from the player and display the game visuals to the player; the server maintains and updates the game world.

Additionally, thin clients eliminate the barrier between existing platforms with thick clients, as the only hardware required to access cloud-game platforms are those that allow users to interact with the platform. Such hardware is not exclusive to any particular platform. By contrast, individually developed thick clients have remained largely exclusive until recent attempts to implement platform "cross-play" into a select list of games, allowing users of different platforms to interact with each other.

As part of the thin client architecture, processing responsibilities are alleviated from the client-side hardware and allocated to the server on top of its role in relaying information. Due to this configuration, all player inputs are not immediately processed on local machinery; rather, information is delivered over the network. This extra data transmission is known as latency.

While capable of reliably delivering a video with high fidelity, thin clients are more heavily impacted by latency due to data transmission delays. Dealing with higher-quality video consumes more bandwidth, which can add latency for thin clients. Perhaps with video streaming services, this does not pose a major problem, as the users of today are accustomed to awaiting buffering sequences before they can begin viewing the content they wish to stream to their device; in the context of cloud gaming, where latency is present in every action of the player, this is potentially detrimental to a user's quality of experience. In this context, latency corresponds not only to the amount of time the user spends waiting to be able to use the service, as with video streaming but also to the amount of time that passes between game actions being executed and the occurrence of events being communicated as information. This implies strict time parameters for cloud gaming, as users tend to have a negative experience when the game does not respond in both an expected and timely manner. We aim to demonstrate this relationship in our study of the effects of latency in a controlled cloud gaming experience.

A common technique that is employed to mitigate the effects of latency in cloud gaming is adding a playout buffer. This gives the system a larger window to process incoming information and accounts for the inevitable loss of packets during data transmission. This mitigation technique is effectuated by intentionally delaying the output of information from the system by a specified amount of time. In essence, the processing center (i.e., the server) waits for the arrival of data packets. Instead of being processed and transmitted immediately, the packets are stored until the buffer period expires and are then transmitted. In this scenario, if packets are lost during transmission, they can still be reacquired within the given buffer time frame.

If a packet is lost and not reacquired before it is due for transmission to the client, an interruption occurs, resulting in a loss of frames during playback. As large buffer sizes minimize interruptions during playback, small buffer sizes allow there to be more interrupts when packets are lost. Either outcome has a negative impact in terms of the experience of the user due to either the feeling of a lack of responsiveness in the game or due to the presence of interruptions and being unable to respond properly to in-game stimuli.

This trade-off presents the need to balance game responsiveness and deliver a smooth visual experience to maintain the player's immersion in the game. Since determining the best buffer size depends upon the user's interactions and experiences with these factors, a way to evaluate the kind of experience a user has with a system is needed. Our study employs the Mean Opinion Score (MOS) metric in a short survey with a simple format that gauges the changes in a user's quality of experience in response to increases or decreases in buffer size and network effects jitter after each round of gameplay.

Games of different genres and fundamental gameplay mechanics may have different relative sensitivities to a given network parameter. For example, a popular type of game is a first-person shooter (FPS), where gameplay is highly dependent on quick and precise responses on behalf of both the game and the players. In contrast, latency in a turn-based strategy game often has considerably less impact on the user performance and quality of experience. Two games were selected for use in the study. The first game included is a rhythm game known as Osu!, where players must time the inputs of four keys on the keyboard with musical "notes" falling down the screen in 4 columns. The second is an arcade-reminiscent game called Geometry Wars, where the player controls shapes in a closed two-dimensional space and must survive by eliminating enemy shapes.

Playout delay and network jitter were simulated using a modified internet router during the study. Participants were exposed to delays that directly impacted their mouse and keyboard inputs. Likewise, network jitter was induced to cause visual aberrations on participants' display monitors. Participants were prompted to fill out a 3-question survey after each period of gameplay with a randomly generated combination of buffer size and network jitter.

The values for delay and jitter were generated from a predetermined set of values: delay was administered strictly in increments of 0, 30, 60, or 90 milliseconds, and jitter in increments of 0, 20, 40, or 60 milliseconds. Delay and jitter were never simultaneously present in any game round played by participants except in a single "crossover" round of each game, where the delay was set to a value of 60 milliseconds, and jitter was set to 40 milliseconds. This routine was followed for each of the 33 users that participated in the study.

The chapters that follow contain all relevant information about this project and the facilitation of the user study conducted. This report of our findings includes the project background (Chapter 2), the methodology of our study (Chapter 3), the analysis of our results (Chapter 4), pertinent conclusions (Chapter 5), and insights into potential future research (Chapter 6).

2. Background

The information in this chapter is presented to enforce the reader's understanding and perception of the concepts that are essential to our user study (see Chapter 3) and its results (see Chapter 4). This chapter provides additional context to cloud gaming systems, the network parameters of delay and jitter, and the quality of experience as relevant to this publication - primarily in the form of information originating from other related work.

2.1: Cloud Gaming Framework - Google Stadia

As their label suggests, cloud gaming services only receive and send data to the virtual client ("in the cloud") while the game is run on a remote server, meaning that the user does not have to download the game to their local device. This framework differentiates the thin clients that offer cloud gaming from traditional thick clients.

However, with some added conveniences come a few drawbacks. Since data is being transferred between local hardware and the server through the Internet, the quality of the Internet plays a significant role in the cloud gaming experience. Network performance parameters such as delay, jitter, and bitrate is essential for maintaining an acceptable QoE for users.

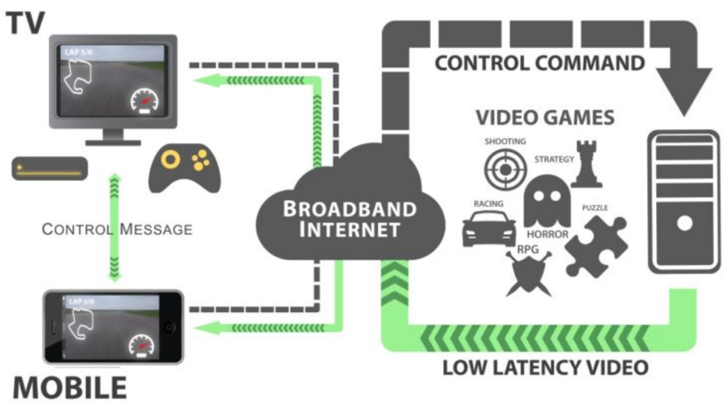


Figure 2.1: Graphical illustration of connection involved in cloud gaming

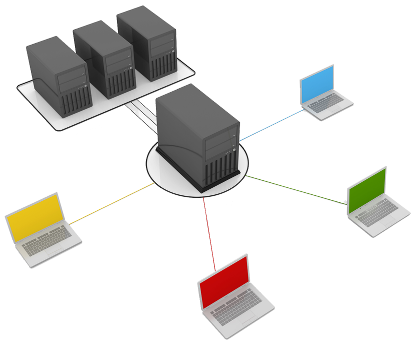
A study of network traffic associated with systems such as Google Stadia seeks and garners information that provides insight into the performance of a cloud gaming system (Stadia) [1].

A section of the analysis from this publication examines Stadia's performance under various conditions commonly encountered by users of cloud gaming services, which are also comparable to those simulated in our study: "degraded network conditions" as well as conditions of "reduced network capacities, packet losses, and added delays." In our study, delay and jitter are implemented to simulate a gameplay experience under the effects of buffers and interrupts (respectively); their procedure implements delay, packet loss, and a restricted network capacity to examine and analyze the response from Stadia's inner workings. Through their data analysis, it was determined that Stadia boasts a "resilience to [packet] loss" in terms of game visual quality and game "playability."

While their work does not have a particular focus on QoE, this information can certainly support the notion that Stadia preserves QoE under specific network quality degradations. The same study could be conducted just as easily on other cloud game platforms such as Amazon Luna or NVIDIA GeForce Now. Similar research analyzing different systems can contribute towards developing a better optimized QoE for its users.

2.2: Delay and Jitter

Delay and jitter are two possible standards to indicate network quality [2]. High delay corresponds to "lag" in the video stream and may be noticed by users. For instance, if there is a high delay in a live televised sports game, what is observed is what happened several milliseconds, or even whole seconds ago. High jitter leads to interruptions in the live stream, which appear as "frozen frames" - playback "freezes" on a particular frame when new frames cannot be played on time. Both of them adversely affect the quality of experience of the user, in different ways and to different degrees [3].



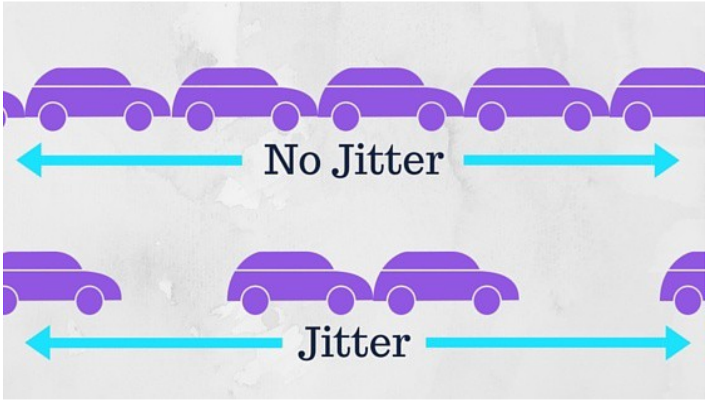


Figure 2.2: Graphical depictions of network components (left) and jitter (right) [2]

The QoE research challenge is to find a quantitative relationship between QoE and (a) delay and (b) interrupts. This is the challenge that our work targets, in the context of cloud gaming via thin clients, by conducting a user study. As it turns out, there may be a generally acceptable range of delay for most individuals. A series of brain activity experiments revealed that neural activation of brain regions, known to be involved in attentional and action control [5], is closely associated with behavioral indicators of potentially adverse effects of delays in human-computer interaction [3], just like our neurons interact with different activities involving senses like vision every day. Whereas delays that on average are below a just noticeable threshold (200 ms) do not lead to a significant increase in activation, those above such a threshold (400 ms, 600 ms) lead to strong recruitment of the bilateral anterior insular cortex, posterior medial frontal cortex, inferior parietal lobule, and inferior frontal junction. Thus, the observed increase of activity in these brain regions can be taken as an indicator of an unsuccessful human-computer interaction at the moment of the occurrence of unexpected delays [6].

2.3: Quality of Experience (QoE)

In our study, QoE is measured subjectively and indicates user satisfaction with the experience presented to them [7]; the variables that influence the QoE perceived by our users are delay and jitter. Network performance typically fluctuates during service and there are times where this turbulence causes incoming packets to consume the entirety of the network's bandwidth. If the playout stream is not delayed by a sufficiently large buffer, QoE inevitably suffers in some way or another.

In video streaming (such as Netflix), excessive bandwidth consumption may not be a rare occurrence: the solution that many may be familiar with is for the application to cease playout and display a message notifying the user that the stream is “buffering” until seamless playback can be guaranteed. Video streaming platforms such as Netflix are playout buffers to smooth over the variable delay and display a message notifying the user that the stream is “buffering” until seamless playback can be guaranteed. The alternative outcome has in the previous studies [3] [9] been proven to be a greater detriment to QoE: in the absence of a buffering protocol, the displayed video instead experiences an interruption (“freezes”) until the incoming packets can be processed to form a new image, resulting in the loss of some packets entirely. In this scenario, frames of the video are lost because they can not be processed quickly enough in time for their own scheduled playout in sequence and thus the continuity of the data stream is disrupted, resulting in a “choppy” playout from the perspective of the user.

2.3.1: Evaluation of QoE

The QoE associated with cloud gaming is sensitive to factors arising from the network; delay, jitter, bitrate, and other parameters are essential in trying to draw conclusions on the QoE relating to any such service [8]. Several parameters that may influence and be used to quantify QoE are discussed in [7].

Many studies have been carried out in efforts to make evident a tangible relationship between QoE and network performance in terms of parameters such as delay, jitter, packet loss, and bitrate [9]. In this endeavor, metrics have been conceptualized to create an objective evaluation of QoE in terms of physical values of parameters that indicate network performance. Our own work differs in that we instead evaluate QoE in terms of subjective responses submitted by participants of a user study. Both methods for evaluating QoE are valid and have been adopted by different research groups depending on the objective and scope of the work.

2.3.2: Correlation between QoE and its Influencers

A study by Allard et al. [3] that is similar to ours sought to explicitly model a relationship between QoE and a prescribed set of factors. As in our case, they facilitated a user study to examine the impact of both buffers and interrupts (in isolation) on QoE, in the context of video streaming. There are fundamental differences between cloud-based applications that naturally cause our separate studies to diverge, the main difference being that the QoE associated with cloud gaming is inherently more fragile and sensitive to delays and interrupts than cloud video streaming during service. From their work, they came to a conclusion of high relevance to our own work: in general, the effects of interrupts on QoE tend to be substantially higher than that of delays caused by buffering (roughly a ratio of 2:1). The results of our work may support a conclusion that corroborates this finding.

The work done by Hossfeld et al. [4] considers other factors impacting QoE and conducts subjective studies such as our own to do so. Their work addresses the possible idea that QoE can vary depending on the application being employed by users; they found that "user rating diversity remains remarkably application-invariant." However, it is important to note that the applications tested were limited to video streaming, wireless Internet connection setup, and social network authentication; cloud game streaming is not included among these.

3. Methodology

A platform from which to stream gameplay to a client device, as well as the selected games for streaming on the platform, were carefully chosen as part of the extensive design process for a user study. Through this study, we aim to obtain insight into the relative effects of visual interrupts and playout buffers on a user's quality of experience (QoE). The games were chosen based on their expected sensitivities to each parameter, and each user played the games while enduring the effects of various levels of induced delay and network jitter, implemented by a customized router.

3.1 Hardware/Software

In order to simulate spontaneous delays and network jitter on our game-streaming platform of choice, we obtained a custom router that is able to manipulate these parameters for prescribed values. The router was connected directly to the client and monitored all data departing from the client ("upstream"?).

The script we used could make the router induce any of the following effects for positive real-valued inputs: delay, packet loss, and jitter. Throughout the course of the study, we only utilized the delay and jitter functions. While the input for delay commands consists of a single value denoting the size of the buffer, the jitter input consists of two values: a base delay (equivalent to the delay command input) and a variation value (the jitter). All jitter commands were implemented with a base delay of 10ms for consistency and because it lies within the range of delay most commonly experienced when on a cloud-gaming platform.

3.2 Hardware/Software

One of the first tasks our team took on in preparation for the study was scouting a platform to stream gameplay from the server to the client. For this decision, there were a handful of options available in the public domain. During our investigation into potential platforms, two became the focus of our attention - Steam Link and Gaming Anywhere. In the end, it was for a multitude of reasons that we opted to use Gaming Anywhere to facilitate our study.

At first glance, Steam Link seemed to be a promising choice. The image streamed to the client was nearly as impressive as what was displayed on the server. It has many customizable settings, including allowed framerate and resolution, as well as the added bonus of having access to a large array of games - the entire Steam library (that the user's Steam account has ownership of).

It was when we began to change the parameters of the network environment that it became clear that Steam Link would not be as viable as we had hoped. We found Steam Link to have an acceptable and expected response to induced delays. The outputs of the router directly translated the display of the client and from what was observable on screen, there seemed to be no other actions on Steam Link's end. This was not the case when Steam Link detected impulses of jitter generated by the router. Whereas Steam Link would not compensate for delays to any clearly noticeable degree, it seemed that any amount of jitter placed on the system would generate an equivalent response from Steam Link, in the form of reduced bitrate. That is, instead of experiencing a loss of frames on screen as an expected result of network jitter, a degradation in visual quality and resolution is observed. This was the outcome of the vast majority of internal testing conducted with Steam Link and network jitter. However, there were a few rare cases where Steam Link would not actively interfere with the image being streamed with the client and perfectly transcribe visual aberrations. Unfortunately, this occurrence was not guaranteed and was thus ruled out as a reliable choice for a study involving network jitter as a parameter.

Despite our final decision to employ Gaming Anywhere in favor of Steam Link, we were initially inclined to be able to adopt Steam Link. A downside of Gaming Anywhere that we wished to avoid was the inherent low visual quality. Unlike Steam Link, Gaming Anywhere was not designed with a large base of potential consumers in mind but instead designed for use in other studies such as this one. As such, Steam Link's software prioritizes the smoothness of the visuals above much else and generally maintains the highest resolution possible; that is, the resolution of the image is variable under Steam Link. Gaming Anywhere instead provided the consistency and reliability that was needed for this study. A consequence of this is that the resolution of the image put out by Gaming Anywhere on the client was of considerably lower quality even with no external influences (delay or jitter), especially compared to that obtained from Steam Link under the same parameters. A possible and likely result of this is that the user quality of experience we observed during the study was skewed towards lower scores. However, this effect would apply equally over the aggregate of data points and the possibility of relative skew between scores is considered null.

Aside from concerns of visual fidelity, an upside to choosing Gaming Anywhere to provide a basis for our study was the support available to us in setting up the platform and optimizing it for our study.

3.3 Choice of games

Many games would have been viable choices to incorporate into the user study. There are primarily two parameters that are of interest to our study in regard to the quality of experience - delay and jitter. Hence we decided to select games on the basis of their relative sensitivities to these parameters.

Our first choice was Osu!, a rhythm game that tasks players with timing inputs to match tabs representing musical notes that fall down the screen. It has been shown in previous studies that users playing this genre of game may experience an increase in performance in response to induced delay; we instead maintain the hypothesis that Osu! gameplay will be more sensitive to visual interrupts than delay. If a player is waiting until the tab is in the right position to make an input and the screen experiences a spontaneous loss of frames, we expect player performance to be impacted.

Conversely, we chose Geometry Wars with the expectation of a relatively higher sensitivity to delay. Spiritually an arcade game, Geometry Wars has the player actively surviving by shooting enemy shapes in a confined 2D rectangular area of space. The gameplay exhibits an increasingly faster pace that matches the rising difficulty as time progresses in the game. As the enemy spawns become more rapid and inclusive of faster enemy types, we expect player performance to be affected when the player's reactions are not immediately implemented.

In summary, we hypothesize that Osu! Gameplay will be interrupt-sensitive for the player performance's reliance on consistent visual output, and likewise, a relative delay sensitivity of Geometry Wars for requiring timely reactions on the player's behalf.

An additional consideration in the selection of these games was the simplicity of the gameplay. We designed our study so as to be able to obtain user data for a variety of delay and jitter values, which requires several runs of the same games; time was a resource that had to be properly managed for the study. By choosing games that are relatively simple, we limited the time needed to explain the gameplay to our users as well as the time needed for them to familiarize themself with the game.

3.4 Pilot Studies, choice of delay, and jitter values for study

Two pilot studies were conducted before the first week of the user study. During the first pilot study, we implemented delay values of 0, 20, 40, and 60 milliseconds and jitter values of 0, 25, 50, and 75 milliseconds. We observed that the response to the resulting interrupts was drastically negative for every jitter value, with little differentiation between each. On the other hand, the various delay values seemed to not follow any observable pattern with respect to the measured QoE. We attributed this to the range of selected values being too low and tight. For the second pilot study, we tested a different set of values: delay was raised to values of 0, 30, 60, and 90 milliseconds while jitter values were reduced to 0, 20, 40, and 60 milliseconds. This choice also reflects the greater impact of interrupts on QoE than delays. With these values in place during the second pilot study, we observed a more varied pattern of results that better fell in line with expectations. This is the set of values we maintained throughout the entire course of the user study in the weeks that followed. The particular permutations are listed; note that a permutation containing both a value for the delay and for jitter was incorporated into the study.

Delay (ms)	Jitter (ms)
0	0
0	20
0	40
0	60
30	0
60	0
90	0

Delay (ms)	Jitter (ms)
60	40

Table 1: Presets of Delays and Jitters

3.5 Approval from the Institutional Review Board (IRB)

In order to gain approval to conduct the user study, we submitted an application for review by the IRB. As our study posed minimal to no risk to any of our potential users, we opted for an exemption application that significantly reduced the amount of information needed by the IRB for approval. Our submission to the IRB included a copy of the consent forms and surveys to be used, a mission statement, and a procedure and methodology to be followed during the study.

3.6 User Recruitment

The week before the study took place, a link to a scheduling page on Slottr was sent via email to WPI undergraduate students. Through this link, any student could reserve an available time to come to our lab and participate in the study. The initial interest and attendance of many of the users can be attributed to the IMGD program requiring students to obtain play-testing credit for participating in research studies such as the one we have conducted. Email recipients were notified that participation in our study would make them eligible to receive credit. Additionally, a raffle for a $25 gift card was arranged in order to recruit for the user study. Participants who signed up for the study using the Slottr link were sent a reminder via email either the day before their appointment or the morning of the appointment.

3.7 Impact of COVID-19

As per the guidelines set out by the IRB during the review process for our initial application, an additional section had to be appended to our consent form to accommodate COVID-19. Luckily, without the requirement of social distancing, we were able to facilitate the user study in the Zoo Lab at the sub-basement level of Fuller Laboratories as normal. Masks were worn by the student investigators at all times, and the immediate area and equipment accessed by each user were cleaned with disinfectant wipes after their participation in the study.

3.8 Procedure

Upon arrival at room A21 of Fuller Labs, each participant was greeted and presented with a consent form in a digital format. Participants were asked to carefully peruse the form and sign their names at the bottom of the page if satisfied with the terms of the agreement.

Each person was subsequently seated at the designated client computer and asked to provide answers to a preliminary demographic study. Shortly thereafter, the users were briefed on the first game and the procedure to be taken throughout the duration of the study. The object of the game as well as the controls was explained to the user in detail, after which the user was allowed to play a full round where no data was taken; during this time, the user could practice playing the game and develop familiarity with the game controls. This step was also taken upon switching to the second game after the first 8 rounds of gameplay.

After this first “practice run”, users were notified that data would begin to be collected, and the first generated delay/jitter permutation was manually entered into the program used to manipulate the custom router. After the first run experienced delay and/or jitter (or neither, in the case of the control setting), the user was asked to fill out the corresponding set of post-game questions on a laptop. Once the survey was completed by the user, the next delay/jitter permutation would be selected and then the next round of gameplay would be started. The same set of questions was provided after every round of gameplay. Each round of Osu! consisted of one full playthrough of a predetermined song that has a runtime of approximately 90 seconds. This was the same song level played by every user that participated in the study, as well as for each round of Osu! played.

Rounds of Geometry Wars could not be segregated into short 90-second increments, however. In this game, each round lasts as long as the player can survive. Instead, we made use of a timer set to ring after 60 seconds, at which point the game was manually paused from the server so that the user could answer the post-game survey.

3.8.1 Demographic Survey

In order to provide relevant and valuable information on the data we procured during the study, five questions were curated and presented to all users who participated in the study. From these questions, we aimed to gather information regarding the relative skill of the participant and familiarity with the games and platforms that would be presented to them, as well as general demographic information. The questions that were presented are shown in Appendix A.

3.8.2. Post-game Survey

After each round of gameplay, users were presented with a set of three post-game questions for them to answer. Each question inquired about a particular aspect of the user's experience during the previous round of gameplay. The same answer choices were offered for each question, five in total. This format is based on the Mean Opinion Score (MOS) metric, which quantifies answers to questions on the basis of a five-point scale. The questions that were presented are shown in Appendix B.

The first question asks about the user's general experience playing the game. It serves as a baseline question to provide insight into no particular aspect of the user's experience, but rather as an indicator of the general level of enjoyment experienced while playing the game. The second question in the survey was designed to gather the user's perceptions of visual effects due to network jitter, which causes visual aberrations as previously discussed. The third and final question serves to indicate the extent to which the induced delay impacted the quality of experience of the user. In this way, the second and third questions inquire about individual aspects of the user's quality of experience.

Both surveys were created and filled out by the users on a laptop signed in to Google Forms. All user responses were automatically recorded and exported to a spreadsheet, accessible via the same account on Google Sheets.

4. Analysis

In this chapter, we overview the results of the user study. The study results are extracted from submitted responses to the two surveys presented to participants via Google Forms, which exports submissions to a spreadsheet in Google Sheets. Once the sequence of preset values for delay and jitter was randomly generated for all 16 rounds of gameplay, the values were entered into 16 corresponding empty rows in the linked spreadsheet, along with the user's identification number and the game played during that round. Google Forms then automatically enters the time of submission and the answers selected for each of the three questions, beginning with the first row of the spreadsheet for the first submission.

The answers given to the survey taken at the start of the user study are of note, where users gave insight into the relative amount of time spent playing computer games in a week and the types of games they may be previously familiar with. Here we also present the analysis of our post-game survey results from the user study. In total, 33 users participated in the user study and responded to the surveys. All graphical figures corresponding to post-game survey data are presented with 95% confidence intervals as a standard measure.

4.1 Demographic Survey Results

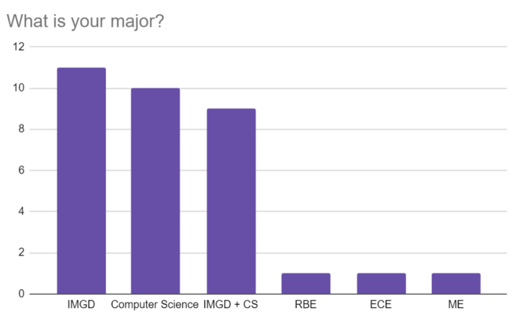


Figure 4.1: Ages of participants

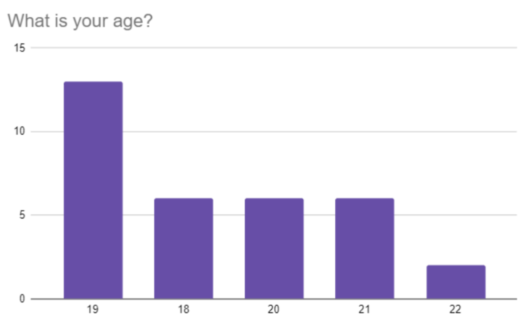


Figure 4.2: Majors of participants

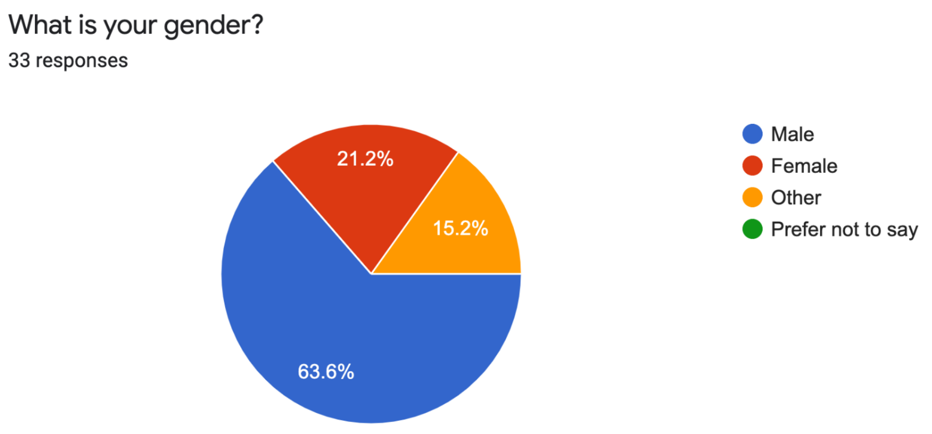


Figure 4.3: Gender identities of participants

The majority of participants in the study identified as male in their submitted preliminary survey responses. All participants were between 18 and 22 years of age at the time of the study. According to the survey results, 64% of the participants identified as male, 21% identified as female, and the remaining 15% of respondents identified as other; note that none of the participants opted for the choice of “prefer not to say.” All but three participants specified themselves as Computer Science and/or IMGD majors, the rest being engineering majors. We make no conclusions in this report regarding the quality of experience based on our participants' gender identities, ages, or declared majors. However, we simply include the data here to provide any potentially relevant context to future related studies.

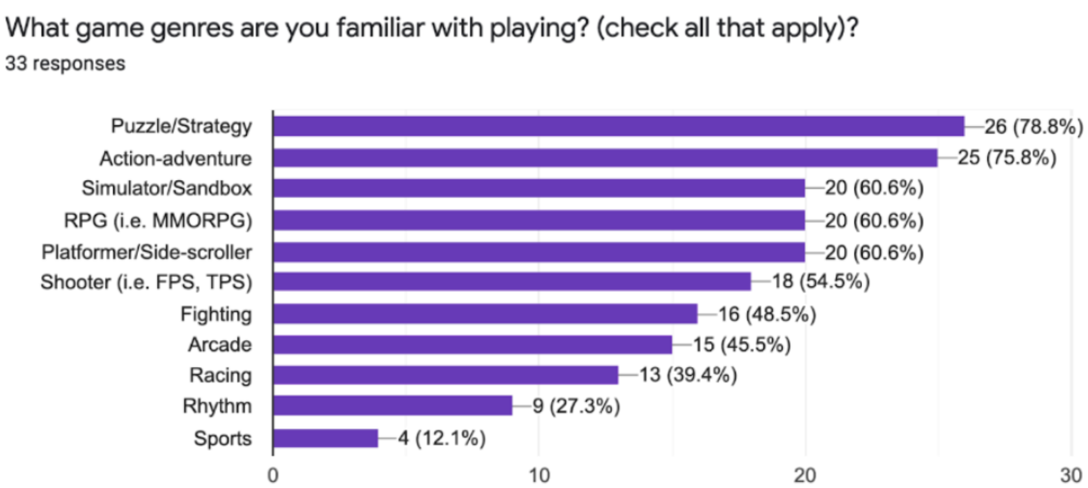


Figure 4.4: Game genres familiar to participants

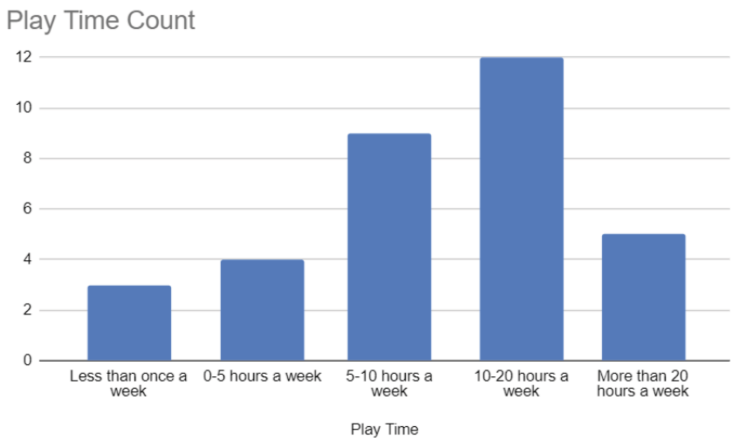


Figure 4.5: Gaming frequency of participants (weekly average)

More than half of the participants spend at least 10 hours a week playing video games of some sort, which happens to be more than the average amount of time spent by more than 60% of Americans [10]. As shown in Figure 4.4, the highest number of participants considered themselves familiar with the action-adventure and puzzle/strategy genres (more than 75% of participants each). Conversely, participants identified the least with the sports and rhythm genres, to which Osu! belongs; these genres exhibited a selection rate of less than 30% each. Interestingly enough, the arcade and shooter genres lie at the exact center in terms of user familiarity as the fifth- and sixth-most selected genres, respectively, out of the 11 total options available. Coincidentally, "arcade" and "shooter" most accurately define and categorize the second chosen title for the study, Geometry Wars. These data suggest that the users who participated in the study are more likely to have previous experience playing games like Geometry Wars than they are to have experience playing rhythm games such as Osu!

4.1 Post-Gameplay Survey Results

For each round of gameplay influenced by different combinations of jitter and delay, the participants were asked to submit responses to a short post-game survey about their experience in a quantified way. The survey consisted of three multiple-choice questions that remained the same when presented to participants after each game round. Here we use a Mean Opinions Score (MOS) metric, where each question has five answer choices that correspond to a rating on a scale of 1-5. These answer choices ordered from most negative to most positive are “poor,” “bad,” “fair,” “good,” and “excellent.”

The first question (Q1) inquires about participants’ overall experience (i.e., level of enjoyment) while playing the game. The second question (Q2) prompts participants to rate the level of “smoothness” of the game visuals, while the third question (Q3) extracts a rating of the perceived “responsiveness” of game controls. Since jitter causes interruptions to the visual feed that reduce observed “smoothness,” question 2 was chosen to gauge the effect of jitter on user experience. Likewise, since visual buffers cause a delay in the playout of frames on screen, question 3 was chosen to gauge the effect of buffers on user experience. Interest patterns may be observed from the distribution of the answer choices selected in responses to these questions.

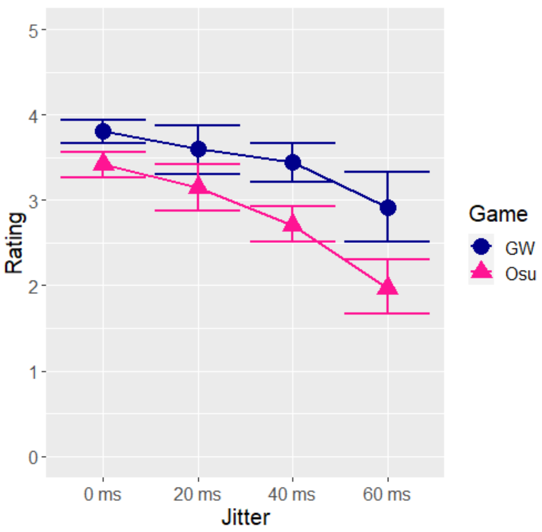


Figure 4.6: QoE vs. Jitter

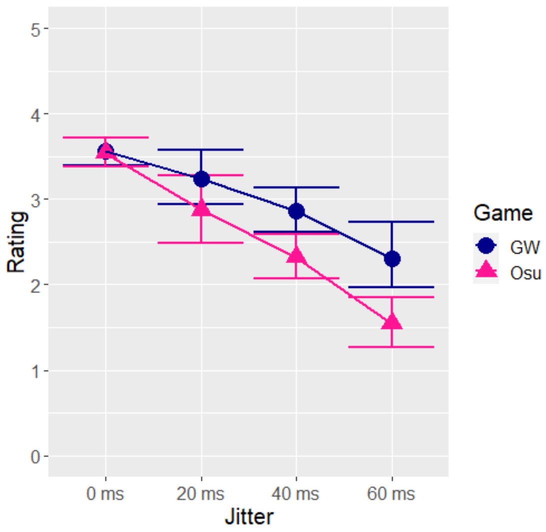


Figure 4.7: Smoothness vs. Jitter

It can be observed from Figures 4.6 & 4.7 that, generally, lower response scores tend to be given for games of Osu! played; this may also be observed in the figures presented below as well. This is perhaps attributable to the fact that Osu! may be considered a more challenging game to play by users, considering that fewer of our participants stated that they are familiar with playing rhythm games than almost all other genres (see Figure 4.4 from 4A). The lack of familiarity with this type of game may be a possible factor that has affected users’ experience and ratings.

Furthermore, while the ratings associated with the first two survey questions decrease in value for increases in jitter experienced during gameplay, it can be seen that the slope of this downward trend is steeper for Osu! compared to Geometry Wars. Observe the slopes of both the blue (Geometry Wars) and pink (Osu!) lines connecting each point on the graph: the gap between the rating values of Osu! and Geometry Wars increases with the jitter of the difference in slope between the two games. This suggests that jitter may have a more drastic effect on Osu! Gameplay than Geometry Wars gameplay, as increases in jitter seem to cause more significant degradations to the Osu! gameplay experience. The jump from 40ms to 60ms of jitter produces the most severe degradation in ratings for both games. In general, each subsequent increase in jitter has an even more significant impact on ratings. Perhaps what is also of note in the figures above is that ratings become more varied for increases in jitter, denoted by the spacing between the bars at each point in the graph. Additionally, there is an interesting occurrence of note in Figure 4.7: the range of rating scores confined within the bars for Geometry Wars and Osu! begins with near-full overlap for the 0 ms jitter case. Then, the overlap diminishes for further increases in jitter until total visible separation at 60 ms of jitter. Overall, the distribution of ratings visible in both figures above is confined to the range of 1.0 - 4.0.

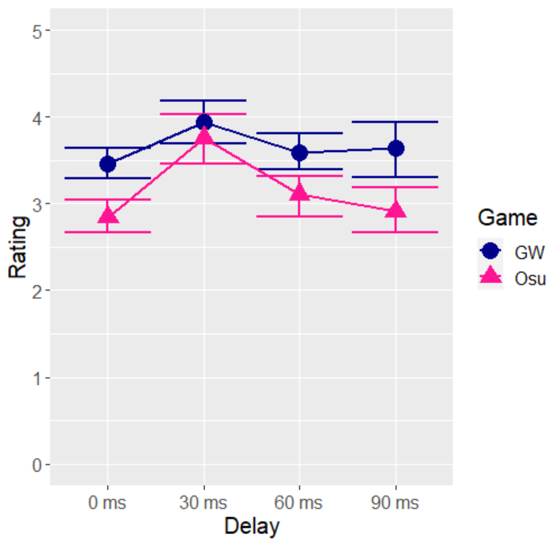


Figure 4.8: QoE vs. delay

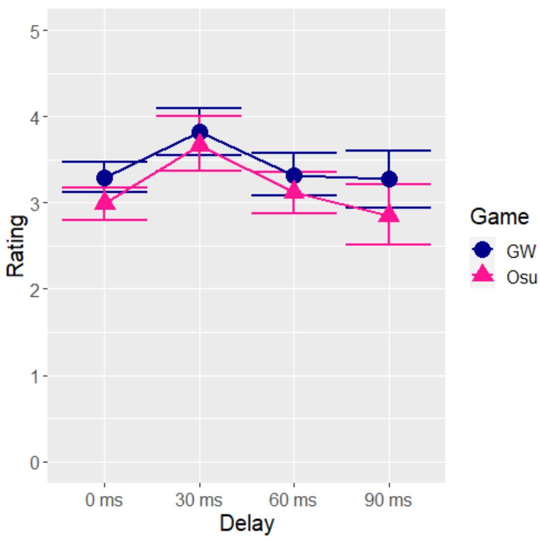


Figure 4.9: Responsiveness vs. delay

Figures 4.8 and 4.9 show that ratings associated with the 30ms delay stand out among the rest as some of the highest presents, corresponding to a peak in the data trend. On average, there is an overall decrease in rating scores for buffer sizes surpassing 30 ms, as may be expected in similarity to the trends arising from jitter; however, there is an increase in scores when comparing the case of no delay to the 30 ms buffer size.

Were we to isolate the three non-zero delay cases (i.e., exclude 0 ms delay), we would observe a negative trend of ratings for increasing buffer sizes, but one which begins to plateau (i.e., decreases less with increases in delay). It can be seen in Figure 4.8 that there is surprisingly an increase in rating scores between the 60 ms and 90 ms cases for game rounds of Geometry Wars played. This is perhaps an anomaly in the data, but the overall pattern discussed may suggest a diminishing effect of delay on gameplay, as opposed to the previously examined data trends that arise with jitter.

In Figure 4.8, the data corresponding to 0 ms of delay consists of the lowest scores on the graph for both games. However, the reason to believe that delay should increase a user's quality of experience is that the data arising from question 1 of the survey are perhaps inconclusive regarding the effect of delay on a user's overall experience. Generally, the ratings associated with delays fall within a much narrower range than those associated with jitter; scores tend to fall within the range of 2.5 - 4.5 for both games. Additionally, the bars enclose smaller ranges of scores in these figures. What is more, is that there is minor variation and a more notable overlap in scores between Geometry Wars game rounds and Osu! game rounds, which is particularly evident in Figure 4.9; the most significant amount of overlap between the data of the two games is observed in this particular graph.

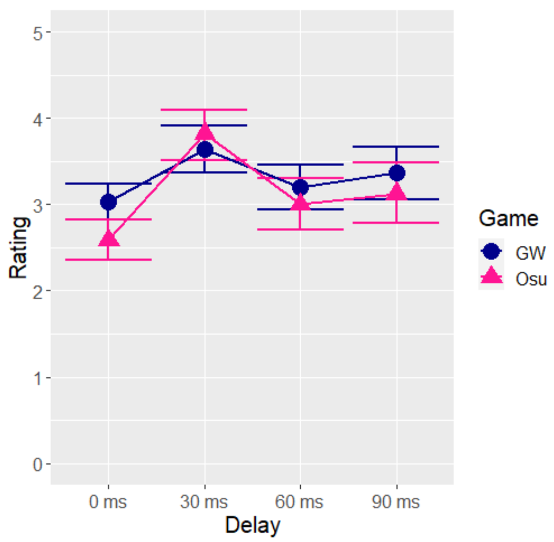


Figure 4.10: Smoothness scores vs. delay

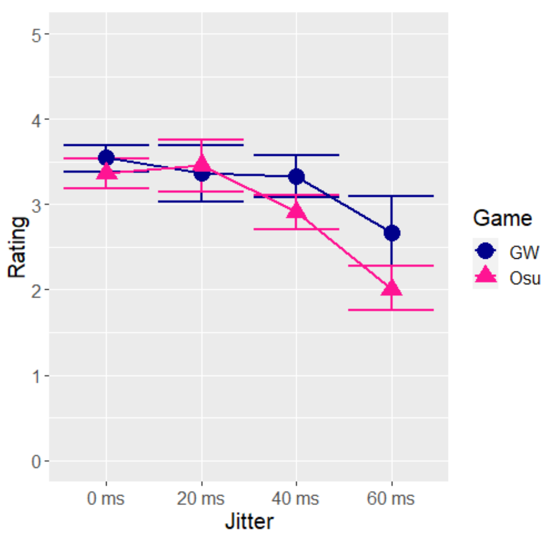


Figure 4.11: Responsiveness vs. Jitter

The last improvements in graphical representations of the data are presented in Figures 4.10 and 4.11 above. We intended to understand the effects of Delay and Jitter on “Smoothness” and “Responsiveness.” We assumed that higher delay and higher jitter would cause worse experiences. From Figure 4.11, our prediction on the jitter’s effect is correct because of the downward trend. However, Figure 4.10 indicates that delay and smoothness have no strict linear relationship. The highest rating appears in the 30ms testing group and the lowest rating appears in the 0ms group. The best experience might be brought by a certain value of delay for different games, which is worth further researching.

5. Future Work

We analyzed the relationship between user feedback and a specific delay or jitter value. One potential study direction might be to narrow down the range of the combinations of jitter and delay values that provide the best user experiences. This might provide a more accurate relationship.

Figure 4.8 and Figure 4.9 showed a pattern different than the decreasing pattern we expected. This could be because the size of the experimental sample was not big enough to generate a confident relationship. Future works to make improvements in this area should include inviting more users to enlarge the sample size. Also, future experiments could apply more different combinations that contain delay values to find the best QoE combination. During the experiment, some users did not know how to play rhythm games and had no previous experience. Before each game began, we gave participants a practice run, but one minute of practice run might not have been enough for participants who had never played rhythm games before. It may be better to use games of a more common type in future experiments such as First Person Shooting games or Multiplayer Online Battle Arena games. We could also invite users to test games that they are familiar with, which should reflect more accurate feedback.

To sum up, we could improve our experiment data by (1) tracking certain users' responses to further study, (2) narrowing down the range of the combinations of jitter and delay values that provide the best user experiences, (3) inviting more users to enlarge the sample size, and (4) increasing the types of testing games to allow participants to test on their familiar games.

6. Conclusion

Cloud gaming is advantageous in saving the computing processes of the client side and extending the gaming experience everywhere. With the necessity of the existence of cloud gaming, the QoE of it is worth noticing. We consider buffer size is a key factor in cloud gaming QoE. This project aims to find the quantitative relationship between buffer size and QoE. Since the buffer size is a critical factor for the quality of user experience, a bigger buffer usually causes less jitter and more delay, and vice versa for a small buffer size. Finding a balance point for the appropriate size for the buffer is important. In our experiments, we applied different values of delay and jitter as variables on two different types of games to collect data. Having these values as variables allows us to understand which factor has a relatively bigger influence on user experience.

Based on the user study, we found that with increasing the jitter value, the QoE and smoothness tend to have an inversed linear relationship, which is in line with our previous expectations. However, with the increasing delay value, the QoE and responsiveness have a different pattern than we assumed, such games study seem to be less vulnerable to increasing delay values. The highest rating happened when the delay reached 30ms for both games. Based on our data, we conclude that higher jitter causes users to have lower QoE and smoothness, while higher delay does not necessarily lead to lower QoE and responsiveness. The specific value of delay that gives the user the best experience, may vary among different games.

References

[1] X. Xu and M. Claypool, “A First Look at the Network Turbulence for Google Stadia Cloud-based Game Streaming,” in IEEE INFOCOM - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Vancouver, BC, Canada, May 2021, pp. 1–5. doi: 10.1109/INFOCOMWKSHPS51825.2021.9484481.

[2] J. Fossenbell, “Understanding Internet Speed and WiFi | CenturyLink,” Apr. 14, 2021. https://discover.centurylink.com/understanding-internet-speed.html (accessed Oct. 11, 2022).

[3] J. Allard, A. Roskuski, and M. Claypool, “Measuring and Modeling the Impact of Buffering and Interrupts on Streaming Video Quality of Experience,” in Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia, Chiang Mai Thailand, Nov. 2020, pp. 153–160. doi: 10.1145/3428690.3429173.

[4] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, “Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea,” in 2012 Fourth International Workshop on Quality of Multimedia Experience, Jul. 2012, pp. 1–6. doi: 10.1109/QoMEX.2012.6263849.

[5] J. Roach and K. Parrish, “What is Cloud Gaming?,” Digital Trends, Mar. 29, 2021. https://www.digitaltrends.com/gaming/what-is-cloud-gaming-explained/ (accessed Nov. 18, 2022).

[6] C. Kohrs, N. Angenstein, and A. Brechmann, “Delays in Human-Computer Interaction and Their Effects on Brain Activity,” PLoS ONE, vol. 11, no. 1, p. e0146250, Jan. 2016, doi: 10.1371/journal.pone.0146250.

[7] H. E. Dinaki, S. Shirmohammadi, E. Janulewicz, and D. Cote, “Deep Learning-Based Fault Localization in Video Networks Using Only Client-Side QoE,” IEEE Trans. Artif. Intell., vol. 1, no. 2, pp. 130–138, 2020, doi: 10.1109/TAI.2020.3041816.

[8] S. S. Sabet, S. Schmidt, S. Zadtootaghaj, C. Griwodz, and S. Möller, “Delay Sensitivity Classification of Cloud Gaming Content,” in Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems, Istanbul Turkey, Jun. 2020, pp. 25–30. doi: 10.1145/3386293.3397116.

[9] C. Moldovan and T. HoßFeld, “Impact of Variances on the QoE in Video Streaming,” 28th International Teletraffic Congress (ITC 28), Sep. 2016, vol. 03, pp. 19–24. doi: 10.1109/ITC-28.2016.311.

[10] J. Clement, “Weekly Time Spent Playing Video Games in the U.S. 2021,” Statista, Oct. 11, 2022. https://www.statista.com/statistics/936654/weekly-time-spent-playing-video-games-usa/ (accessed Sep. 20, 2022).

Appendices

Appendix A: Demographic Survey Questions

1. What is your age?
2. What is your gender?
    - Male
    - Female
    - Other
    - Prefer not to say
3. What is your major?
4. How often do you play video games?
    - Less than once a week
    - 0-5 hours a week
    - 5-10 hours a week
    - 15-20 hours a week
    - More than 20 hours a week
5. What game genres are you familiar with playing (check all that apply)?
    - Shooter (i.e. FPS, TPS)
    - RPG (i.e. MMORPG)
    - Action-adventure
    - Rhythm
    - Racing
    - Fighting
    - Sports
    - Puzzle/Strategy
    - Simulator/Sandbox
    - Platformer/Side-scroller
    - Arcade

Appendix B: Post-game Survey Questions

The questions were posed as follows:

How was your game experience

How would you rate the “smoothness” of the game visuals during this last round?

How would you rate the responsiveness of the game controls during the last round?

Answer choices for each question were:

- Excellent
- Good
- Fair
- Poor
- Bad