MŰEGYETEM 1782

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Networked Systems and Services

Investigation of the impact of cognitive bias on the perceived command execution delay in video games

Bachelor's Thesis

Author
Duy H. Nguyen

Advisor
Dr. Peter A. Kara

May 22, 2025

Thesis task description

Gaming is one of the largest industries of digital entertainment. Modern gaming software may be susceptible to command execution delay, which may be caused by various factors, such as insufficient rendering capabilities or limited network resources. At the time of the thesis, the utilized advances in gaming are often accompanied by brief descriptions when communicated to the users. While such descriptions may be compressed into a couple of words, even a single word may impact user experience. Due to the cognitive bias induced by the labeling effect, the impact of such a word may actually be more significant than what the user genuinely perceives.

Within the scope of the thesis, the tasks to be performed by the student shall include the following:

1. Review the related scientific literature and the relevant international standards.
2. Design the methodology for a subjective study that addresses the aforementioned form of cognitive bias.
3. Implement the experimental setup and carry out the subjective tests.
4. Analyze the collected data and assess the correlation between the test variables and the obtained subjective scores.

Contents

STUDENT DECLARATION

I, *Duy H. Nguyen*, the undersigned, hereby declare that the present thesis work has been prepared by myself and without any unauthorized help or assistance. Only the specified sources (references, tools, etc.) were used. All parts taken from other sources word by word, or after rephrasing but with identical meaning, were unambiguously identified with explicit reference to the sources utilized.

I authorize the Faculty of Electrical Engineering and Informatics of the Budapest University of Technology and Economics to publish the principal data of the thesis work (author's name, title, abstracts in English and in a second language, year of preparation, supervisor's name, etc.) in a searchable, public, electronic and online database and to publish the full text of the thesis work on the internal network of the university (this may include access by authenticated outside users). I declare that the submitted hardcopy of the thesis work and its electronic version are identical.

Full text of thesis works classified upon the decision of the Dean will be published after a period of three years.

Budapest, May 22, 2025.

Duy H. Nguyen
student

Abstract

Gaming is a multi-billion-dollar industry and stands among the most prominent forms of entertainment worldwide. As games become more advanced, player experience is increasingly influenced not only by technical performance factors such as graphics, frame rate, and responsiveness, but also by psychological factors that affect perception. This thesis investigates the impact of the labeling effect—a form of cognitive bias which alters subjective evaluation via descriptive labels—on the player's perception of command execution delay in video games. The experimental setup was developed to assess two types of video game input (single and continuous), with varying game difficulties and different extents of added command execution delay (0 ms, 50 ms, 150 ms, and 250 ms). The subjective study included a total of 48 test conditions, which were evaluated by 60 test participants. The task was to directly compare objectively identical game sequences, which were labeled either as "optimized" or "not optimized". Half of the test participants compared "optimized" sequences to "not optimized" ones, while the other half was presented with the opposite order. The findings reveal that the labeling effect significantly influences perceived responsiveness. The results show that 70% of the ratings reported noticeable differences, with a preference toward the sequences with the "optimized" label. In addition, a strong correlation between the extent of command execution delay and the influence of the labeling effect was observed. When it comes to the game difficulties, the differences measured were statistically significant between easy and hard difficulties; however, the medium difficulty did not differ significantly from either. Regarding input type and label order, no statistically significant differences were measured. The findings suggest that cognitive bias can meaningfully affect how players perceive performance, even in the absence of actual changes. Prior to the submission of this thesis, the work was published in a scientific journal under the title "The Influence of the Labeling Effect on the Perception of Command Execution Delay in Gaming".

Tóm Tắt

Ngành công nghiệp trò chơi điện tử là một lĩnh vực trị giá hàng tỷ đô la và là một trong những ngành giải trí lớn nhất hiện nay. Khi các trò chơi ngày càng trở nên tiên tiến hơn, trải nghiệm của người chơi không chỉ bị ảnh hưởng bởi các yếu tố kỹ thuật như đồ họa, tốc độ khung hình và độ phản hồi, mà còn bởi các yếu tố tâm lý ảnh hưởng đến cảm nhận. Luận văn này nghiên cứu tác động của hiệu ứng gán nhãn—một dạng thiên kiến nhận thức làm thay đổi đánh giá chủ quan thông qua các nhãn mô tả—đối với cảm nhận của người chơi về độ trễ trong thực thi lệnh trong trò chơi điện tử. Một hệ thống thử nghiệm đã được xây dựng để đánh giá hai loại phương thức nhập dữ đầu vào trong trò chơi (đầu vào rời rạc và đầu vào liên tục), với các mức độ khó khác nhau và các mức độ trễ được thêm vào (0 ms, 50 ms, 150 ms và 250 ms). Nghiên cứu chủ quan bao gồm tổng cộng 48 điều kiện thử nghiệm, được đánh giá bởi 60 người tham gia. Nhiệm vụ của họ là so sánh trực tiếp các chuỗi trò chơi giống hệt nhau về mặt kỹ thuật, nhưng được gán nhãn là "tối ưu" hoặc "không tối ưu". Một nửa số người tham gia so sánh chuỗi "tối ưu" với chuỗi "không tối ưu", trong khi nửa còn lại được trình bày theo thứ tự ngược lại. Kết quả cho thấy hiệu ứng gán nhãn ảnh hưởng rõ rệt đến cảm nhận về độ phản hồi. 70% số lượt đánh giá ghi nhận sự khác biệt đáng chú ý, với xu hướng nghiêng về các chuỗi được gán nhãn "tối ưu". Ngoài ra, có mối tương quan mạnh giữa mức độ trễ thực thi lệnh và mức độ ảnh hưởng của hiệu ứng gán nhãn. Xét về độ khó của trò chơi, sự khác biệt được đo lường có ý nghĩa thống kê giữa hai mức dễ và khó; tuy nhiên, mức trung bình không khác biệt đáng kể so với hai mức còn lại. Đối với loại đầu vào và thứ tự gán nhãn, không có ghi nhận nào về sự khác biệt có ý nghĩa thống kê. Các phát hiện này cho thấy thiên kiến nhận thức có thể ảnh hưởng đáng kể đến cách người chơi cảm nhận hiệu suất của trò chơi, ngay cả khi không có sự thay đổi thực tế nào. Trước khi luận văn này được nộp, nghiên cứu đã được công bố trên một tạp chí khoa học với tiêu đề "The Influence of the Labeling Effect on the Perception of Command Execution Delay in Gaming".

Chapter 1

Introduction

1.1 Motivation

Over the past decade, video gaming has become one of the most prominent and widely embraced forms of entertainment. Unlike traditional media, modern video games offer highly immersive and interactive experiences, allowing players to engage with richly-crafted virtual worlds. Through role-playing, mission completion, and narrative exploration, players can inhabit characters and scenarios far removed from everyday life. These experiences are inherently fictional; however, if presented in a proper manner, they can still foster a strong sense of presence and emotional connection. The capacity to blend fiction with perceived reality plays a crucial role in shaping the overall experience.

The user experience in video games is largely shaped by the quality attributes of the game itself. Elements such as high-resolution graphics, engaging narratives, immersive world-building, well-developed characters, and innovative gameplay all contribute to the overall user experience. However, one particularly critical attribute—if not the most important—is the performance of the game. The performance of video games can be measured by different factors [1], yet most of those factors are somewhat relevant to the technological advancement of the system that the game operates on. For instance, the work of Andreev [2] indicates that most players perceive video games that are capable of running at 60 fps or higher to be much better than those games that are frame-locked at 30 fps (i.e., games that run at a maximum of 30 frames per second). However, in my thesis, I investigate the performance from the aspect of responsiveness, which refers to the immediacy of how a game reacts to the player input. In another word, the command execution delay—measured by the time elapsed from the issue of the input command to the visual update of the corresponding action on the display device—is examined in this thesis. Prior studies have shown that even minimal perceptible delays can disrupt immersion, distract players, and significantly degrade their overall enjoyment of the game [3].

Despite the tremendous technological advancements, latency remains a hindrance to the achievement of a smooth, seamless experience of video games. While modern hardware and software can mitigate the command execution delay to mere milliseconds, the perceived responsiveness is not solely dictated by these objective measurements. Human perception is heavily influenced by psychological and cognitive factors [4], which can distort the experience of latency. In some instances, players may perceive minimal delays to be more disruptive than they actually are. This gap between actual system performance and user perception addresses the role of cognitive bias in video game enjoyment.

One such form of cognitive bias is the labeling effect, a phenomenon which can alter the individual's subjective evaluation of an experience or a product through the presence of contextual descriptive labels [4]. On the other hand, we live in a modern society that thrives on trading goods and products, and builds on the provision of services, which essentially makes members of the society consumers. Whenever an individual intends to buy a product, one may notice that product labels usually use certain vocabularies that affirm their characteristics in a manner that encourages purchase. That might be the labeling effect in action [5]; oftentimes, it only takes a single word to enable the labeling effect [6, 7, 8], and video games are no exception to this phenomenon. For instance, in competitive gaming, players are not only playing the game itself, but also competing with others, thus, it is a common practice for players to design and/or follow methods to play games in the most efficient way. For instance, these methods are known as "build order" in multiplayer online battle arena (MOBA) games and real-time strategy (RTS) games, and they are used to gain certain advantages over opponents [9, 10]. Optimizing gameplay also exists in non-competitive games. For example, the activity of attempting to finish a game in the shortest period of time is known as "speedrun". The optimization does not always focus on the way players play the game; occasionally, system settings are also taken into account (e.g., how the keyboard should be bound, how much the sensitivity of the mouse should be adjusted, etc.). Moreover, modern video games are regularly updated with hot fixes and patches. This emphasizes the requirement for the optimization to be updated in order to keep up with the state of the game. Such methods are widely shared and discussed among gaming communities and are usually referred to as "meta". The term "meta" is popular to such an extent that any content if labeled "meta" may be considered as "optimal" or "the best".

Summa summarum, while the advancement of gaming technology provides powerful computer systems that are meant to enable more immersive and complex video games, the impact of cognitive bias on the perception of delay is rather under-studied when it comes to the player experience. Only a few studies on the labeling effect were carried out thus far, and generally no study addressed the impact of the labeling effect on the perceived quality of video games. Therefore, the lack of scientific literature on such impact of cognitive bias is the primary motivation for my thesis.

As a personal motivation, I like playing video games. Over time, I became increasingly interested not just in the gameplay itself, but in the underlying systems that shape player experience—particularly how subtle factors like feedback timing, interface design, and performance optimization affect immersion and satisfaction. I also noticed how often players—including myself—would judge changes in gameplay quality based on patch notes or update labels, even when the differences were not immediately perceptible. This sparked my curiosity about the psychological factors that influence perceived performance, especially the role of cognitive bias, such as the labeling effect. Through this thesis, I aim to combine my personal passion for gaming with academic inquiry to gain a better understanding of how expectations shaped by labels can alter the perception of quality and responsiveness, even when no objective changes are presented.

1.2 Analysis of task description

Within the scope of my thesis, I addressed the tasks in the following manner:

1. "Review the related scientific literature and the relevant international standards."

One of the first steps in this thesis was to dive into the existing scientific literature and international standards related to the topic. This helped me build a strong foundation for understanding how player experience is evaluated and what factors influence it—especially in terms of performance and perception. I looked into previous research on the different forms of cognitive bias, such as the labeling effect, and how these can affect the way entities and characteristics are perceived. Exploring studies on video game performance, user perception, and gameplay optimization gave me a clearer picture of how complex and layered the gaming experience can be. At the same time, I also reviewed international standards that define how user experience and performance should be measured [11, 12], how viewing conditions should be selected [13], and how gaming-related considerations should be addressed [14, 15, 16]. These standards are important because they offer widely accepted guidelines and methods that make it easier to evaluate games in a consistent, meaningful, and reproducible way. Bringing together both academic insights and formal standards helps to ensure that the thesis is not only scientifically sound, but also practically relevant. Due to the interdisciplinary nature and complexity of my research work, I review the related scientific literature separately in the contexts of command execution delay and cognitive bias. Moreover, I address the relevant standards during the elaboration of the employed methodology.

2. "Design the methodology for a subjective study that addresses the aforementioned form of cognitive bias."

   To effectively investigate the impact of the labeling effect on the perception of video game performance, it was essential to design a clear and controlled subjective study. The goal was to create an experimental environment where the only varying factor was the label assigned to identical gameplay sequences. This meant carefully planning how the game sessions would be presented, ensuring that the true nature of the game sequences remained unknown to the test participants. Key elements of the methodology included selecting appropriate game input types (i.e., single-input and continuous-input), defining conditions (i.e., extent of added delay, difficulty, and labels), and constructing a pairwise comparison format that allowed test participants to express preferences—indicating potential cognitive bias. It was also important to take note of the order of presentations (i.e., which assigned label was shown first), which may have an impact on the extent of bias. Overall, the methodology was shaped to isolate the labeling effect and measure its influence on user perception as accurately as possible.

3. "Implement the experimental setup and carry out the subjective tests."

   Once the methodology was finalized, the next step was to implement the experimental setup and conduct the subjective tests. Two different custom game environments were developed to allow controlled manipulation of game conditions—command execution delay and difficulty level—while keeping the visual and gameplay elements identical across different test cases. The key feature of this setup was its ability to assign different labels—"optimized" or "not optimized"—to pairs of identical gameplay sequences. Test participants were recruited to take part in the study, and each was asked to play through multiple pairs of sequences, rating their experience based on perceived performance differences. A brief training phase was carried out to ensure that participants were fully aware of the labels assigned to the sequences, and to familiarize them with the games and the assessment task.

The study sessions were conducted under consistent conditions to reduce external influence, and all responses were collected digitally for later analysis.

4. "Analyze the collected data and assess the correlation between the test variables and the obtained subjective scores."

   Finally, the last part involves the analysis of the collected data and the assessment of the correlation between the test variables and the obtained subjective scores. This step focuses on processing the responses gathered from the test participants during the subjective tests, and applying appropriate statistical methods to identify patterns, trends, and relationships. The goal is to determine how the different test variables—label order, gameplay difficulty, input type, and added delay—affect the perception of gaming quality (i.e., the amount of perceived delay).

1.3 Individual contribution

This thesis represents an independent research effort to explore the effects of cognitive bias—specifically the labeling effect—on perceived delay in video games. I carried out all stages of the research project on my own—of course, with guidance from my thesis supervisor—from the initial review of relevant literature to the final analysis and documentation.

As the only author of this thesis, I conducted an extensive review of the scientific literature and international standards related to command execution delay in gaming, gaming Quality of Experience (QoE), cognitive bias, and subjective testing methodologies. I identified gaps in current research, particularly the lack of empirical studies investigating the labeling effect in the context of gaming.

Based on this knowledge, I proposed research questions, formulated hypotheses, and designed the experimental methodology to address the labeling effect in a controlled environment. I selected appropriate test variables (i.e., assigned labels, difficulty levels, and extents of added delay) that formed test conditions, and created a test scenario where gameplay sequences used for comparisons were identical except for the assigned labels.

I developed and implemented the experimental setup, including the user interface, gameplay sequences, and the rating system. I recruited test participants and conducted the subjective tests, ensuring that ethical standards and data privacy were maintained throughout the study.

I performed a comprehensive statistical analysis of the results. This included calculating means, identifying distribution patterns, and assessing statistical significances between conditions using appropriate tests. Then from the observations, I drew conclusions on the impact of cognitive bias on the perceived command execution delay.

In addition to the research and analysis, I wrote the entirety of the thesis document without the help of any form of assistance based on artificial intelligence. This project has not only allowed me to deepen my understanding of research related to user experience, but also provided me with valuable experience in independent scientific work, critical thinking, as well as technical writing.

I hereby state that according to my best knowledge, the work introduced in this thesis is the first to investigate the impact of the labeling effect on the perception of command execution delay. While I relied on earlier works on QoE to decide the methodology of my research effort, my individual contributions separate well from the scientific literature.

The scientific contents of my thesis have already been published in an international peer-reviewed journal at the time of thesis submission, under the title "The Influence of the Labeling Effect on the Perception of Command Execution Delay in Gaming" [17]. My thesis supervisor and myself are the only authors of the published paper—no other student of this university or any other institution contributed to the work.

1.4 Hypothesis declaration

The hypothesis of this work is built around several research questions that aim to explore how labels such as "optimized" and "not optimized" influence the perception of delay.

(i) The first question asks whether these labels alone can significantly change how test participants rate identical game sequences. Is there a statistically significant difference between the theoretical ratings and the obtained results that are influenced by the labels "optimized" and "not optimized"?

(ii) Next, the study looks at the role of label order. Is there a statistically significant difference between pairs where the "optimized" is the first label and "not optimized" is the second label and pairs with the opposite order?

(iii) The study also examines whether players respond differently when they interact with different types of game input. Is there a statistically significant difference between the results obtained for single-input and continuous-input games?

(iv) The fourth question focuses on game difficulty. Are there statistically significant differences between the results obtained for easy, medium, and hard game instances?

(v) The last question addresses the amount of added delay. Are there statistically significant differences between game sequences with different extents of added delay?

In all of these cases, the null hypothesis is that any difference between the aforementioned clusters of the obtained results is simply due to random error.

1.5 Thesis structure

The remainder of my thesis is structured as follows. The related scientific literature is reviewed in Chapter 2, separately for command execution delay and cognitive bias. The methodology of the research effort is detailed in Chapter 3, including test environment, test variables, software (i.e., the two video games I created for this work), test conditions, and test protocol. The analysis of the obtained subjective ratings is presented in Chapter 4, addressing overall results, label order, input types, game difficulty, and added command execution delay. The chapter also provides additional discussions related to the implications, limitations, and potentials of such work. Finally, my thesis is summarized in Chapter 5, which lists the potential future continuations of the research effort as well.

Chapter 2

Related Work

In this chapter, the published scientific achievements relevant to the study are discussed. The research efforts on the topic of command execution delay and cognitive bias are analyzed separately.

2.1 Command execution delay

The classic study of Shniederman [18] shows that maintaining the computer system response time under one second is crucial for user engagement and it keeps user frustration at a low level, whereas any longer delay may disrupt the cognitive flow. Moreover, delay that exceeds 10 seconds may result in task abandonment. However, it is worth noting that those findings date back more than four decades; human-computer interaction (HCI) has evolved considerably since then. A more recent study of Raaen and Eg [19] indicates that some users can detect—and are irritated by—latencies as short as 66 ms. Delay tolerance can nonetheless increase when users know that the technology in use inherently introduces delays. For instance, in satellite-based space communications, Wang [20] found that latencies up to 10 seconds may actually be tolerated.

In the context of video games, the work of Metzger et al. [21] proposes a lag model for video games. The presented model treats total latency as the sum of network delay, processing delay, and playout delay (i.e., command execution delay). The model also includes additional factors such as input command rate, server tick rate in online titles, and codec latency in cloud gaming scenarios. The survey of Pantel and Wolf [22] shows that players' tolerance for latency depends not only on network delay but also on the game genre—a finding that the study of Schmidt et al. [23] also concludes later on.

Claypool and Claypool [24] observed that latency sensitivity varies across the distinct phases of gameplay. The "play phase"—in which real-time player interaction occurs—is the most vulnerable to latency in online games. Even so, a thoughtful game design can keep the overall experience acceptable under substantial network lag. The work of Fritsch et al. [25] demonstrates this with the analysis of the game EverquestII, showing that players still evaluated the gameplay as tolerable despite the high latency during the "play phase".

In the field of cloud gaming, Chen et al. [26] measured the command execution delay when streaming video games to a personal computer (PC) by different cloud services. The study shows that latencies range from 135 ms up to 500 ms. A more recent study [27] focuses on mobile cloud gaming measured lower but still noticeable delays. Depending on the type of network connection of the device—which was either Wi-Fi or long-term evolution (LTE)—the latency varied between 70 ms and 177 ms.

In the context of virtual reality (VR), the work of Allison et al. [28] measured the latency of the technology itself (i.e., the time interval between a user's head movement and the corresponding update of visual representation). Delays up to 200 ms mark the tolerance threshold before users become uncomfortable. Task performance, however, may demand far stricter responsiveness. Huang et al. [29] asked test participants to complete a balance task while varying the delay of the visual representation. The study concludes that once the delay surpassed 25 ms, the task became impossible. Together, these findings signify the difference between general latency tolerance and task performance.

There are numerous research efforts that examine the impact of influence factors on QoE in gaming scenarios. Multiple studies [30, 31, 32] suggest that age, gender, personality, previous gaming experience, and user performance are some of the main factors that contribute to one's perception of QoE. To quantify their effect, Moller et al. [14] created a seven-dimensional Gaming Experience Questionnaire (GEQ), while Depping and Mandryk [33] developed the Game Specified Attributing Questionnaire (GSAQ) to measure the correlation between the players' characteristics (e.g., emotion, motivation, and behavior) and QoE. Among the aforementioned factors, player performance——which is of particular interest for game designers and developers——is heavily dependent on command execution delay. Numerous studies on networked-gaming [8, 34, 35, 36, 37] focus on the impact of delay on the player's performance. The results of these studies converge and indicate that games with a delay around 100 ms are acceptable; when reaching 200 ms games are playable but annoying; and above 500 ms they were reported to be extremely difficult and scored poorly in terms of QoE. The work of Chanel et al. [38] indicates that performance is linearly affected by game difficulty. If a game is too easy, boredom sets in; if it is too hard——especially under delayed conditions——frustration rises. Note that QoE drops in both cases. Researchers therefore proposed Dynamic Difficulty Adjustment (DDA) models [39, 40, 41, 42] to reduce the impact of the player's performance on QoE. Sabet et al. [43, 44] introduced an improved DDA model for cloud-based gaming which operates in real time and does not require the tracking of player performance. By modifying in-game mechanics, the model could mitigate the negative effect of delay on QoE [45]. On the other hand, the work of Lee and Chang [46] introduces Advance Lag Compensation (ALC) for first-person shooters (FPS), while the finding of Liu et al. [47] indicates that as long as player's self-evaluated performance was fulfilled, the game still received positive QoE ratings under delayed conditions. Only when these conditions were not met, the games became laggy and annoying to play. Finally, studies [48, 49] also suggest that delay can degrade experience through visual-audio asynchrony. Experiments show that players may tolerate visual-audio offsets up to roughly 250 ms, although the exact threshold depends on the characteristics of the game.

A great number of research efforts on the topic of command execution delay focuses only on network delay, while local delay is overlooked—even though it also contributes to the total lag experienced by the player. The work of Raaen and Petlund [50] shows that local system delay may reach up to 100 ms—making it comparable to network delay. The study of Long and Gutwin [51] reports a variety range of local lag ranging from 50 ms to 500 ms. The study also introduces a predictive model called "Time to React" with the aim to help developers estimate and counteract the impact of such local delay.

Claypool et al. [52] examined how the lag of the cursor affects user performance. In the study, test participants were asked to use a mouse with added delay to choose and click on fast-moving targets across varying target speeds and mouse latency levels. The results indicate that both user selection times and error rates increase linearly with added local delay, cutting overall performance to roughly 25% of the performance without lag. In a follow-up study using the same methodology, Claypool et al. [53] concluded that perceived QoE is mainly governed by the magnitude of delay rather than by target speed. These findings suggest that local latency also affects modern FPS games. Liu et al. [7] examined the effect of local delay on the player's performance in the FPS game Counter Strike: Global Offensive. During the test, participants were asked to fire different weapons at targets under different conditions of delay. The results indicate that reducing latency from 125 ms to 25 ms improved player performance by 20%. The work of Ivkovic et al. [54] also focuses on first-person-view computer tasks and the impact of delay on user performance. The authors measured the real-world local latency by using a high-speed camera—frame by frame—from the moment the input command was issued to the actual visualization update on the user screen. The study reveals that the average local delay ranges from 23 ms to 243 ms which may cause a significant and substantial degradation in the performance of the target tracking and target acquisition tasks—which are fundamental tasks in FPS games.

The scientific literature suggests that FPS games are more sensitive to delay in general than other games from different genres. The work of Quax et al. [55] reveals that fast-paced games such as FPS, fighting games (FTG), and racing games—which require intense interactions—suffer more from latency than puzzle or strategy games. Other studies reveal similar findings in the field of cloud-based gaming [56], traditional online gaming [57], and mobile gaming [58]. On the other hand, the study of Claypool [59] focuses on the impact of latency in RTS games. The study concludes that such strategy games revolve around long-term strategy rather than instantaneous reaction, thus, making the game genre less dependent on responsiveness. This finding aligns with the findings from the work of Pedri and Hesketh [60]. The authors also concluded that fast-interaction computer tasks demand greater attentional resources, which in turn reduce the user's ability to process temporal information. As a result, time is perceived to pass more quickly during these tasks, making users more likely to tolerate longer delays without awareness or discomfort.

Sabet et al. [43] conducted an experiment in which he periodically injected a local lag to the input while the test participants played different games. The authors developed three games: Shooting Range (aiming and shooting at targets with the mouse), T-Rex (jumping over obstacles on a 2D platform), and Rocket Escape (horizontal dodging randomly spawned obstacles). The results reveal that players may mitigate delay by anticipating upcoming events whenever the game behavior was based on rules and thus, was predictable. Consequently, Shooting Range and T-Rex still received favorable QoE ratings. On the other hand, obstacles in Rocket Escape are randomly spawned which makes the game less adaptable; thus, the added latency sharply lowered the QoE ratings. The study of Normoyle et al. [3] concludes that players may adapt well to constant delays up to 300 ms; however, a small jitter (i.e., the variation of delay) may be noticed by all users and cause annoyance.

These findings resonate with the finding of the work of Kohrs et al. [61]. The authors used functional resonance imaging (fMRI) to examine how the brain responds to local delay during computer-based tasks. The results reveal that when delay occurs frequently, the brain adapts over time, gradually returning neural activity to the baseline level. In contrast, unexpected or irregular delays may disrupt user focus and cognitive flow, indicating that predictability plays a critical role in how delay is perceived by users.

Command execution delay can affect both single-player and multi-player gaming experiences, but its impact is generally more concerned in multi-player settings. Multi-player games often involve real-time competition or cooperation, where delays can introduce a sense of unfairness and lead to frustration—particularly when a player feels disadvantaged against others [62, 63]. As a result, subjective QoE ratings tend to degrade more in multi-player contexts [64]. Siu et al. [65] observed that players perceive single-player games to be less challenging than multi-player games, even when the tasks are objectively identical. This perception amplifies the impact of delay in multi-player environments, where players often expect tighter responsiveness. Among multi-player game genres, FTGs are especially delay-sensitive because the gameplay of such game is frame-accurate—even minimal lag can shift the outcome of a match. Thus, netcode and delay compensation methods are applied in such games to ensure fair play and competitive integrity of the game [66, 67, 68].

The influence of local delay on other input devices was also studied. Claypool [69] evaluated user performance in a selection task using a thumbstick controller under varying levels of delay. The study concludes that completion time increased exponentially as delay grew, but players with higher skill levels were less affected—highlighting that player ability may contribute to delay tolerance and perceived QoE. The study of Long and Gutwin [70] concludes that similar behaviors can be observed in the case of pointing devices such as drawing tablets.

Table 2.1: Perceptual and tolerance thresholds of delay for each game genre reported in the scientific literature.

Scientific work	**Game genre**	**Perceptual threshold**	**Tolerance threshold**
Beigbeder et al. [37]	FPS	75 ms	200 ms
Quax et al. [71]	FPS	60 ms	–
Xu et al. [72]	FTG	67 ms	–
Fritsch et al. [25]	MMORPG	–	1250 ms
Tan et al. [73]	MOBA	50 ms	200 ms
Beznosyk et al. [57]	Puzzle platform	60 ms	200 ms
Pantel et al. [22]	Racing	50 ms	500 ms
Claypool [59]	RTS	100 ms	500–800 ms
Hohlfeld et al. [74]	Survival RPG	170 ms	–
Nichols and Claypool [75]	Team sports	500 ms	–

Table 1 summarizes the research efforts that investigate the perceptual and tolerance threshold of delay across various gaming genres. In the table, the perceptual threshold refers to the just noticeable difference (JND) [76]. This threshold typically ranges between 50 ms and 75 ms for genres such as FPS, FTG, MOBA, puzzle platform, and racing games. For RTS games, the perceptual threshold is around 100 ms, while it may reach up to 170 ms in survival role-playing (RPGs), and even 500 ms in team sports games. Note that the high extent of delay was measured in an American football game, in which gameplay was partially automated, thus requiring less reactivity from the player. The tolerance threshold—the delay level at which gameplay remains acceptable—varies more widely, ranging from 200 ms to 1250 ms. This latter great value was measured in a massively multi-player online role-playing game (MMORPG). Game genres presented in Table 1 involve player reactivity to some extent, which limits their tolerance to delay; however, genres such as simple puzzle games or turn-based strategy games may withstand significantly higher delays. Moreover, in online multi-player settings, games with low perceptual and tolerance thresholds tend to be most susceptible to performance degradation due to latency.

In summary, extensive research highlights the significant impact of command execution delay on player experience. Studies consistently show that constant delays up to approximately 200 ms are generally tolerable; however, delays exceeding 200 ms tend to impair player performance and degrade overall QoE. Additionally, even low levels of jitter are typically noticeable and frequently lead to frustration. Within the scope of this thesis, the work focuses exclusively on constant command execution delays.

2.2 Cognitive bias

The study of Wilke and Mata [4] identifies confirmation bias as one of the most common forms of cognitive bias. Classic works of Wason [77, 78] and Klayman [79] define confirmation bias as the tendency for individuals to actively seek information that supports their existing beliefs, even when contradictory evidence is available. Such reasoning errors are not necessarily the result of a lack of knowledge or intelligence, but rather come from a complex interaction between cognitive limitations and motivational factors.

The work of Darley and Gross [80] investigates the relationship between confirmation bias and the labeling effect. The study concludes that individuals tend to seek out evidence before forming judgments; however, this process is often compromised when their initial expectations are shaped by invalid or biased information. Jones and Sugden [81] conducted an experiment and came to the conclusion that people are willing to pay for information that confirms their prior beliefs—even when such evidence holds no objective information value.

The labeling effect significantly alters how individuals perceive their environment. Sakai et al. [82] conducted an experiment investigating the impact of visual cues on the sense of smell. Test participants were asked to assess the intensity of odors while being presented with various colors. The results reveal that odor intensity ratings were significantly higher when the color matched the participants' expectations (e.g., a dark brown color paired with a Coca-Cola scent). This finding suggests that confirmation bias, based on the participants' preconceptions, influenced their sensory judgment. In this case, the shown color acted as a label that reinforced the belief in stronger odor intensity—even though there was not. Bentler et al. [83] examined the effect of labeling on human hearing. Participants were provided with identical hearing aids labeled as "conventional" or "digital". Despite the device being the same, the majority favored the "digital" version, with some even reporting perceptible improvement. The study illustrates how the labeling effect can shape subjective experiences. In a different context, Iglesias [84] surveyed customer perceptions in the banking sector. The author interviewed participants before and after their work at the bank. The study reveals that though preconceptions did not significantly impact the overall evaluation of the service experience, they did influence the assessment of specific quality dimensions.

The labeling effect also shapes how individuals process and interpret information. Gao et al. [85] examined the influence of stance labels on readers' selection of news articles. The study indicates that humans are not neutral information processors; their pre-existing beliefs, bias, and emotional inclinations significantly influence how they engage with and interpret news content. Readers are more likely to select articles aligned with their viewpoints, reinforcing the effect of confirmation bias in media consumption. In a related study, De Graaf et al. [86] investigated how experts search and process information in given documents. Participants—selected professionals in their fields—were tasked with answering a set of questions using information available in the provided texts. The study reveals that experts frequently rely on their prior knowledge, even when the document contains relevant information. This reliance, in some cases, led to the disregard of available evidence, resulting in incomplete or inaccurate responses. The study highlights the impact of cognitive bias in the selective nature of information processing when expectations or labels are presented.

The labeling effect may also significantly influence consumer behavior and product perception. Chovanova et al. [87] found that brand labels can bias consumers' decision-making processes, regardless of age or gender, suggesting that brand identity alone can strongly affect purchasing behavior. Similarly, Gao et al. [85] demonstrated that consumers are willing to pay a premium for products, such as beef, when additional attribute information—such as the product's origin—is provided. This aligns with the findings of Stylidis et al. [88], which argues that product attribute information can be a decisive factor in shaping consumers' perceived quality of a product. However, there are limitations to this effect. Fitzgerald et al. [89] observed that excessive labeling and conflicting attribute information can overwhelm and confuse consumers, impairing decision-making and ultimately negatively impacting businesses. In a different context, the work of Christandl et al. [90] explores price perception bias and shows that customers overestimated price increases following changes in Value Added Tax (VAT), even for products unaffected by the tax change. These results reinforce the notion that consumer expectations, influenced by labeling and prior beliefs, can distort objective judgment.

In the domain of software and video game development, several studies have examined the impact of cognitive bias on the development process. Research findings suggest that cognitive bias most often manifest during the testing and debugging phases of software development [91, 92, 93] as developers are often influenced by their anticipation of how the software is supposed to behave. The study of Calikli and Bener [94] shows that developers may test software with incomplete specifications, yet still conclude the implementation is correct. This phenomenon, often referred to as positive testing—where testers seek to confirm rather than challenge the functionality of the software—has been widely investigated [95, 96, 97]. The work of Rainer and Beecham [98] on evidence-based software engineering (EBSE) shows that participants frequently recommended requirement management tools (RMT) with which they were familiar, despite objective evidence supporting the superiority of alternative tools. Additionally, the work of Jorgensen [99] addresses the impact of cognitive bias on the cost and effort estimation in software development. In this case, the estimation is prone to error due to confirmation bias in selecting what is perceived as relevant project cost information.

Cognitive bias also plays a significant role in the perception of video quality and VR experiences. Kara et al. [100, 101] investigated the impact of the labeling effect on perceived video quality in the context of high-definition (HD) and ultra-high-definition (UHD) streaming. Participants were asked to compare videos labeled either accurately or misleadingly. The findings indicate that perceptions of visual quality were heavily influenced by expectations created by the labels, rather than by actual resolution differences. In a more recent study, Kara et al. [102] examined the labeling effect in high dynamic range (HDR) video streaming. The study concludes that labeling visual stimuli as "Premium HDR" led to higher subjective ratings for visual aspects such as luminance, color, and image quality. However, a notable trade-off in perceived frame rate emerged, suggesting that participants expected some form of compromise—which is likely due to preconception of the "cost" of HDR enhancements. The work of Geyer et al. [103] explores similar cognitive bias in the perceived quality of "rugged" smartphones (i.e., highly durable devices). In one experiment, rugged phones were directly compared to conventional smartphones;

in others, visual content was shown on either a smartphone or a computer monitor. In all instances, statistically significant differences were found: rugged phones were generally perceived as less capable, presumably due to the assumption that durability comes at the expense of performance—similar to the frame rate trade-off observed in the HDR study. The study of Bouchard et al.[104] focuses on the impact of preconceptions on the sense of presence. Participants were deliberately misled to believe that the virtual environment they experienced was real, with additional cues reinforcing the illusion that the events were occurring in real time. The results demonstrated a marked increase in the participants' sense of presence, confirming that belief and expectation can enhance the immersive quality of virtual experiences.

In summary, numerous studies have examined how cognitive bias influences human perception and decision-making processes. Confirmation bias commonly arises when individuals seek out information that supports their existing beliefs and preconceptions, while disregarding conflicting evidence. The labeling effect can further intensify this bias by shaping perception based on expectations created by labels. Research indicates that this phenomenon occurs regularly across various areas of daily life. This is especially relevant in consumer behavior, which is fundamentally driven by perceived quality. In the scope of this thesis, quality is defined in terms of video game responsiveness—specifically, command execution delay. I explored how the labeling effect can influence the perception of this performance metric. The methodology of the subjective tests—detailed in the next chapter—builds on the experimental setup of earlier works [101, 102, 103].

Chapter 3

Methodology

In this chapter, the selected methodology is explained in detail, covering the test environment (i.e., the location where the subjective study was conducted), the test variables (i.e., the characteristics that varied across different stimuli), the test conditions (i.e., the specific combinations of these characteristics), the software (i.e., the games developed to investigate the hypothesis), and the test protocol (i.e., the procedures for data collection and the order in which stimuli were presented).

3.1 Test environment

The test environment for the subjective study was designed to ensure consistency, reliability, and minimal external influence on perception. All experiments were conducted in a quiet, controlled indoor setting to reduce distractions and external noise. Every participant assessed the visual stimuli through the same hardware configuration which included a mid-range gaming laptop equipped with a 15.6-inch screen. The display resolution was set to 1920×1080 pixels. The display settings, resolution, and brightness levels were kept consistent across all test sessions. Test participants were shielded from the external light, and ambient lighting in the room was moderated to avoid glare or strain on the eyes. Participants were seated at a fixed distance from the screen to standardize visual conditions. As the corresponding standard [13] recommends 3.2 H (3.2 times the screen height), the initial viewing distance was set to approximately 75 cm—based on a screen height of 23.8 cm. However, as visualization quality was not the focus of the study, test participants could slightly adjust this distance. The games were designed and built on the same local machine, thus, no Internet connection was required, which was important to avoid performance fluctuations. Prior to the experiment, participants were given a short tutorial to familiarize themselves with the games, the controls, the assessment task, and the labels, but no information about the study's true intent was disclosed to prevent bias.

3.2 Test variables

The subjective study was designed to accommodate three variables. The combinations of the variables constituted the test conditions, which were applied to both input types.

The first test variable was the label assigned to each stimulus with the intention of inducing potential cognitive bias. Based on previous scientific literature [83, 102], labels such as "Digital" and "Premium" may have a significant impact on the users' perception of a product. Following the same manner, I chose a relevant term in the context of video gaming that may induce bias in the perceived performance of the game. The label was either “optimized” or “not optimized”; note that no information was disclosed regarding the target of optimization to the test participants.

The second test variable was the extent of added command execution delay, which refers to the time elapsed between the user input (i.e., the user presses a button on the keyboard of the laptop) and the execution of the corresponding action. As shown in Table 2.1 in Section 2.1, 50 ms was the JND for the most delay-sensitive game. In addition, the games used in this study were designed to be fast-paced games as well. Thus, the value of 50 ms was chosen to be the lowest level of added command execution delay. On the other hand, 200 ms is a common tolerance threshold among different game genres, including the most reactive genres. Thus, by deviating from the tolerance threshold by the amount of the aforementioned JND (i.e., 200 ms plus/minus 50 ms), two other extents of delay were chosen for the study. Furthermore, the work of Ikovic et al. [54] indicates that real-world local latency in video gaming is typically up to 250 ms. Summa summarum, the four extents of added command execution delay were the following: 0 ms (i.e., no added delay), 50 ms, 150 ms, and 250 ms.

The third test variable was the difficulty settings of the game (i.e., the speed of progression in the game). Three different difficulties were present in the study: easy, medium, and hard. These difficulties are elaborated in the description of the software.

3.3 Software

To investigate the research questions, I developed two custom video games—a single-input game and a continuous-input game. Both games were developed and built using the Unity Real-Time Development Platform [105]. The test variables (i.e., the labels, the added delays, and the difficulty settings) were also implemented using Unity. The common programming language in Unity is C#, thus, I used this language in the development of the two games.

3.3.1 Single-input game

The single-input video game used in the study was a 2D rhythm game. A screenshot of the game is shown in Figure 3.1.

The core gameplay revolved around a single action—in which the player had to provide the correct input within a certain window of time. Four keys—the arrow keys on the keyboard—were assigned to different colors (i.e., blue for left arrow key, red for up, yellow for down, and green for right). At the beginning of each stimulus, all the target arrows were randomly generated. Note that all of the target arrows had the same color as the buttons that the player needed to press (e.g., right arrows were always green). The game started when the player pressed Enter, after which the target arrows traveled down vertically from their original positions, and as the arrows approached the buttons at the bottom of the screen, the player’s task was to press the correct buttons to gain scores. Once a correct input was made, the arrow disappeared, preventing participants from exploiting the task by spamming or holding down keys. There were four different interactions between the buttons and target arrows: “Miss”, “Hit”, “Good”, and “Perfect”. If a button was pressed while an arrow was overlapping the corresponding button, the distance from the center of the arrow to the center of the button was calculated. In the development environment,

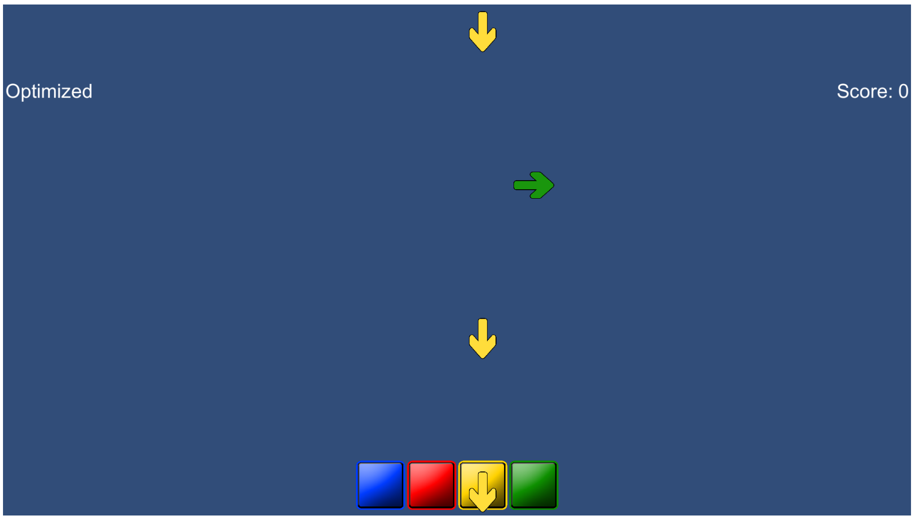

Figure 3.1: Screenshot of the single-input video game used in the subjective tests.

both the arrow and the button were 1 unit long. If the overlapping proportion was more than 0 and less than 50% (i.e., which means the aforementioned calculated distance was between 0.5 and 1), then the game provided visual feedback of a "Hit" interaction. Similar calculations were applied for other interactions as well: "Good" when the overlapping length was in the range of 50% and 80%; "Perfect" when the overlapping length was more than 80%; and a negative percentage (i.e., the arrow does not overlap with the button) resulted a "Miss" interaction. If any arrow key was pressed before the arrow reached the hitbox, no action was carried out. The mechanism of the hitbox is illustrated in Figure 3.2. As the player successfully provided the correct input, the score was updated. The score that the player achieved was displayed in the top-right corner of the screen, and the assigned label was present in the top-left corner.

Game difficulty was controlled by adjusting the vertical speed of the arrows. As difficulty increased, arrows moved faster, reducing the time they remained within the target zone (i.e., the areas of the buttons). Specifically, the arrows were within the target zone for 1500 ms for easy difficulty, 750 ms for medium, and 500 ms for hard. Note that all the arrows were 1 unit in length and consecutive arrows were 2 units apart. This means that in easy difficulty, two consecutive arrows were 3000 ms apart, and the corresponding values for medium and hard difficulties were 1500 ms and 1000 ms, respectively. Each gameplay sequence lasted a fixed 60 seconds, regardless of performance, meaning that higher difficulties featured more arrows rather than shorter sequences. This ensured that performance was measured through scoring rather than sequence duration. It is important to note that a fixed 250 ms delay made the hard difficulty particularly challenging, as the arrow remained in the target area for only 500 ms. However, since the delay was constant, players could adapt by adjusting their timing.

3.3.2 Continuous-input game

The continuous-input video game was a 3D driving game. A screenshot of the driving game is shown in Figure 3.3.

The gameplay required players to control a vehicle, which was steered by the left and right arrow keys. In this setup, "continuous input" meant that holding down a key would

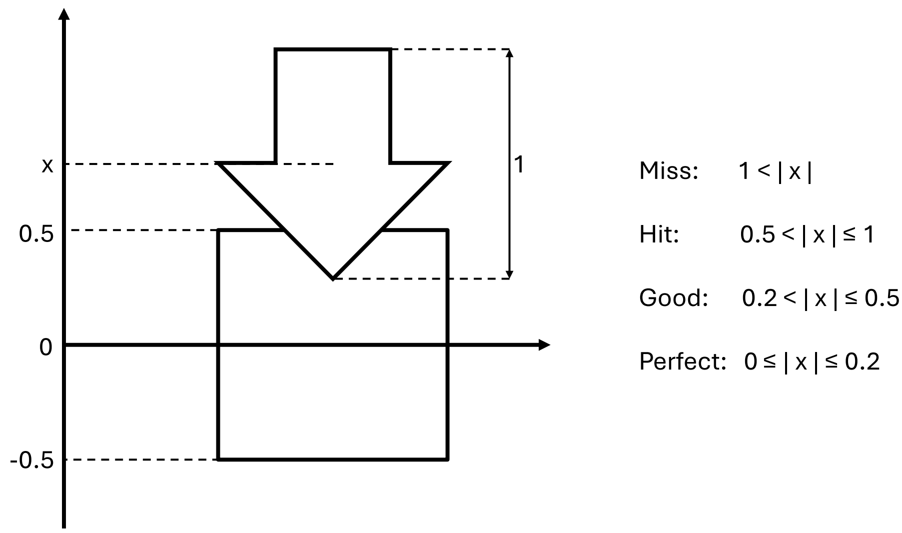

Figure 3.2: Scoring mechanism of the single-input game.

continuously steer the vehicle in that direction while the car was moving forward at a constant speed. The road was divided into three parts: left, right, and center. Two of these parts were regularly blocked by obstacles, thus, the task of the player was to avoid collision. All the obstacles and the 3D car model included the built-in "Collider" component of the development environment [105], which allowed the detection of collision between two objects in the game environment. Each game session provided the player with a total of five "lives"—allowing player to collide with obstacles four times. After five collisions, the game session ended immediately. Additionally, the game would also end if the player over-steered the vehicle, causing it to rotate more than 90 degrees in either direction, thus facing the wrong way. The obstacles were designed to spawn randomly, and the implementation ensured that at least one of the three lines was blocked.

Difficulty levels were determined by the vehicle's speed—higher difficulty meant faster movement. Since the road length remained constant across difficulties, game durations varied: 60 seconds for easy, 40 seconds for medium, and 20 seconds for hard difficulty. These durations were only applicable if there were four or fewer collisions. Note that the distance between subsequent obstacles was the same across difficulties, thus, the time window for collision avoidance was reduced as the difficulty increased. To be more specific, in the case of easy difficulty, two consecutive obstacles were 4000 ms apart, and the corresponding values for medium and hard were 2000 ms and 1000 ms, respectively. This time window was longer than in the case of the single-input game, as more time was required to execute the action (e.g., to steer from the center lane to the left lane).

3.3.3 Delay implementation

To simulate command execution delay, a custom script was included in both games. In many other games developed in Unity, the controlled objects—relying on inputs from peripheral devices—are commonly implemented with built-in functions directly without delay. In order to introduce delay, the actual inputs were provided to the "Delay Input" class with an extra parameter, "delayAmount" (i.e., a float number that represents the desired delay time in milliseconds), which then generated the delayed input signal. In the implementation of the two games, the input signal was retrieved from the "Delay Input" class instead of the "Input" class.

Figure 3.3: Screenshot of the continuous-input video game used in the subjective tests.

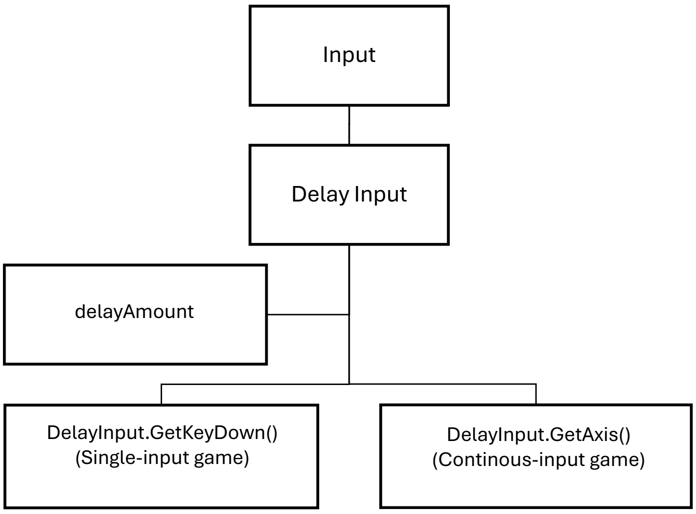

Figure 3.4: Hierarchy structure of input functions in two games.

Note that in the single-input game, it was only necessary to read the value of the input the moment when the key was pressed down, and not the whole duration when the key was held. However, in the continuous-input game, this duration was also taken into consideration. Thus, different input handling methods were used for the two games. Figure 3.4 shows that the methods used to handle the input in both games were called from the custom library "Delay Input". The output of the method "GetKeyDown()" was set to be a boolean value "true" only at the moment the key was pressed (i.e., the value was "false" not only when the key was not pressed, but even when the key was not yet released), thus, I used this method for the single-input game. In contrast, the method "GetAxis()" always returned "true" as long as the key was pressed. Since in the continuous-input game, only the left and right keys were meaningful to the gameplay, the method "GetAxis()" was used instead of "GetKey()". For example, with a 250 ms delay, the vehicle continued steering for 250 ms after the player released the key. Figure 3.5 illustrates this with an input example of 500 ms. Adjusting to this delay (e.g., releasing the key earlier) added an extra layer of difficulty compared to the single-input game.

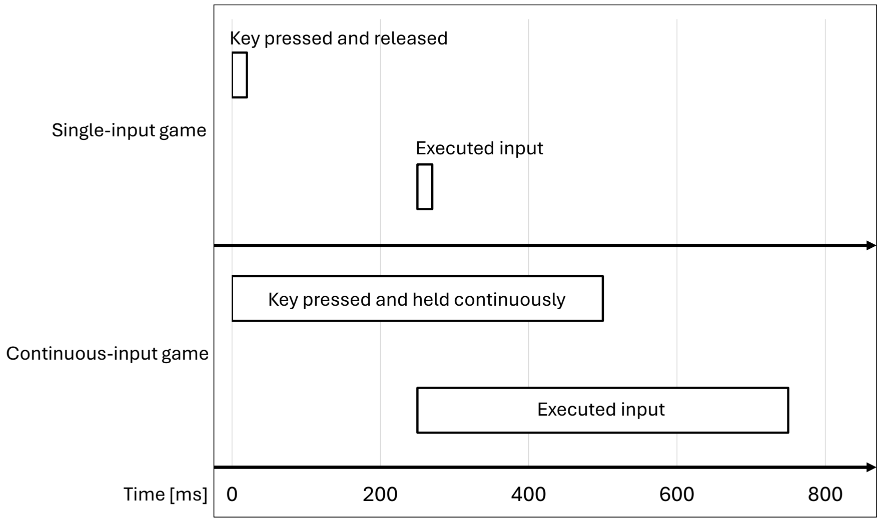

Figure 3.5: Input events and their executions under 250 ms of added delay in the single-input game and the continuous-input game.

3.4 Test conditions

My approach of test condition determination involved testing every possible combination of the test variables (i.e., two label orders, four levels of added delay, and three difficulty settings) on the two games. Table 3.1 summarizes all 24 test conditions used in the subjective study. As stated earlier, these were applied to both the single-input game and the continuous-input game.

Table 3.1: Test conditions used in the single-input game and the continuous-input game.

Game difficulty	Added delay	First label
Easy	0 ms	Optimized
Easy	0 ms	Not Optimized
Easy	50 ms	Optimized
Easy	50 ms	Not Optimized
Easy	150 ms	Optimized
Easy	150 ms	Not Optimized
Easy	250 ms	Optimized
Easy	250 ms	Not Optimized
Medium	0 ms	Optimized
Medium	0 ms	Not Optimized
Medium	50 ms	Optimized
Medium	50 ms	Not Optimized
Medium	150 ms	Optimized
Medium	150 ms	Not Optimized
Medium	250 ms	Optimized
Medium	250 ms	Not Optimized
Hard	0 ms	Optimized
Hard	0 ms	Not Optimized
Hard	50 ms	Optimized
Hard	50 ms	Not Optimized
Hard	150 ms	Optimized
Hard	150 ms	Not Optimized
Hard	250 ms	Optimized
Hard	250 ms	Not Optimized

3.5 Test protocol

In each segment of the study, test participants played the exact same game sequence twice and were asked to compare the second instance to the first one. The only difference between the two was the label: either the first was labeled "optimized" and the second "not optimized", or vice versa. Each label was shown on the screen before the corresponding sequence and also embedded within the gameplay itself, as illustrated in Figures 3.1 and 3.3. Label order was fixed for each test participant, meaning that for a given individual, the "optimized" label was consistently applied either to the first or second sequence throughout the entire test. For half of the test participants, the "optimized" label was always applied to the first sequence, and for the other half, the second. The sequences were clustered by input type, but their order was randomized within a cluster.

The task of the test participants was to directly compare the sequences, which were shown in pairs. The comparison was carried out via a standardized rating scale. The classic study of McKelvie [106] concludes that a rating scale should have at least five to six categories; a smaller number of categories may result in the loss of discriminative power and validity. In addition, the study of Myers [107] suggests that the language used in the rating scale may also induce changes in the test participants' responses. The work of Moller [108] addresses the rating scale standardization methods and their limitations. In this subjective study, double-stimulus tests were conducted (i.e., a single feedback was

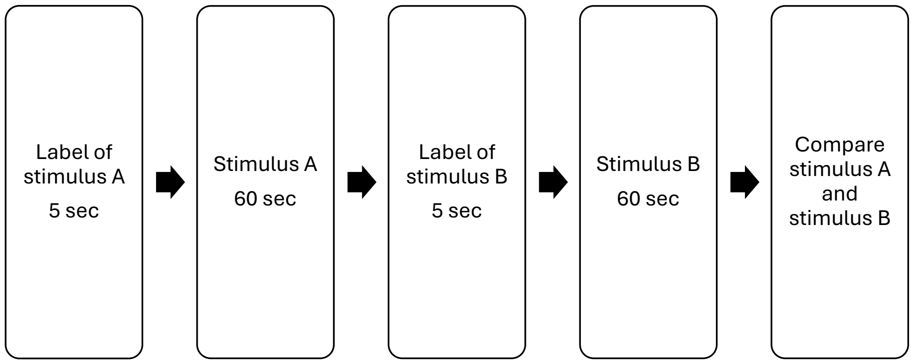

Figure 3.6: Temporal structure of the subjective tests for the single-input game.

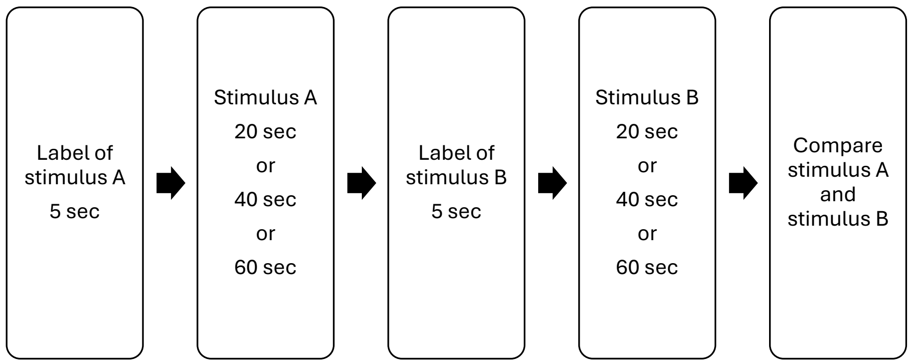

Figure 3.7: Temporal structure of the subjective tests for the continuous-input game.

provided for two stimuli), thus, the absolute category rating (ACR) scale—albeit being the most frequently used scale for QoE studies—was not applicable, as it may only provide indirect comparisons via means and distributions. ACR typically requires test participants to rate a single stimulus in isolation—which is less effective when the stimuli have no real technical differences. The degradation category rating (DCR) scale can be implemented for the double-stimulus test; however, DCR focuses on the perceptibility of differences and annoyance—which would limit the study to the potential negative impact of the labeling effect. Thus, a 7-point paired comparison rating scale was used. The results provided by such a symmetrical scale can assess both the negative and positive impact of the labeling effect, and the three categories in each direction enables the differentiation of various extents of perceived differences. Using an odd rating scale was essential to the study, as such is required to express the absence of perceived differences; an even scale would demand either a positive or negative result. On the level of standards, the conventional 5-point ACR and DCR quality assessment scales are standardized by the ITU-T P.910 recommendation [12], and the 7-point paired comparison scale is standardized by the ITU-R BT.500 recommendation [13]. The rating options of the scale are the following: "Much worse" (−3), "Worse" (−2), "Slightly worse" (−1), "The same" (0), "Slightly better" (+1), "Better" (+2), and "Much better" (+3).

The timing and structure of the subjective tests are illustrated in Figure 3.6 and Figure 3.7, respectively. Each comparison included five steps. First, the label for the first game sequence (i.e., stimulus A) was displayed for 5 seconds. This was simply a dark gray screen with either "optimized" or "not optimized" written in bold white text at the center.

Then, the test participant played the first sequence. As previously mentioned, the single-input game sequences always lasted 60 seconds, while the duration of the continuous-input sequences depended on the difficulty—either 20, 40, or 60 seconds. These durations assumed that the test participant did not make more than four collisions in a single sequence, as such ended the game, and thus, made the sequence proportionally shorter. After the first sequence, the same two steps of label and game sequence were repeated for the second sequence (i.e., stimulus B). Again, it needs to be highlighted that there were no differences between these sequences, only the label differed. Finally, the test participant was asked to compare the second sequence to the first one (i.e., choose one of the seven rating options), for which a time window of 30 seconds was provided. If a test participant completed every sequence without any continuous-input sequence ending prematurely, the entire test lasted around 1 hour and 52 minutes. Because of this long duration, test participants were given the option to take a short break during the session. Before the test began, they were also given time to get used to the games at different difficulty levels.

Chapter 4

Results

A total of 60 people took part in the study. For half of them, the “optimized” version was shown first, while for the other half, it was the “not optimized” one. Of the participants, 39 were male and 21 were female, with ages ranging from 20 to 28 and an average of 23. This age group is a relevant demographic for video game testing. According to The Entertainment Software Association [109], over 190 million people in the U.S. aged from 5 to 90 play video games, of which 75% are Gen Z (i.e., people in the age from 13 to 28). Similarly, in five major European markets [110], about 124 million people play video games, and 78% of them are between 15 and 24 years old. Thus, choosing participants within the age from 20 to 28 ensures that the results reflect the target audience. It is important to note that while gender and age were recorded, they were not analyzed as part of this study. The ratings and demographic data were collected separately.

A total of 1440 comparison scores were collected. This is based on the full set of combinations, resulting from four added delay values, three difficulty levels, and two game types, each assessed by 60 test participants. Each participant completed 24 comparisons. The collected data were analyzed separately according to each test variable—delay, difficulty, and game type—as well as the label order. For the purpose of the analysis, the rating options from the comparison scale were treated as their corresponding numerical values. These numerical descriptors were also visible to participants during the test. Furthermore, for cases in which the “not optimized” stimulus was presented first, the numerical scores were inverted. As a result, all comparisons are interpreted from the perspective of the “optimized” stimulus—meaning each rating reflects how the “optimized” version was perceived relative to the “not optimized” one.

4.1 Overall results

Figure 4.1 presents the theoretical distribution of comparison ratings. Since there were no actual performance differences between the paired stimuli, the expected outcome would be that all ratings indicate “The same” (0), with no deviations toward either positive or negative values.

Figure 4.2 illustrates the actual distribution of the obtained subjective ratings, based on all the 1440 collected scores. It is important to note that fewer than 30% of the responses indicate that the paired stimuli were perceived as “the same”. In other words, more than 70% of the ratings show a perceived difference in the command execution delay. Among the ratings, 57.36% were positive—suggesting that the “optimized” stimulus was

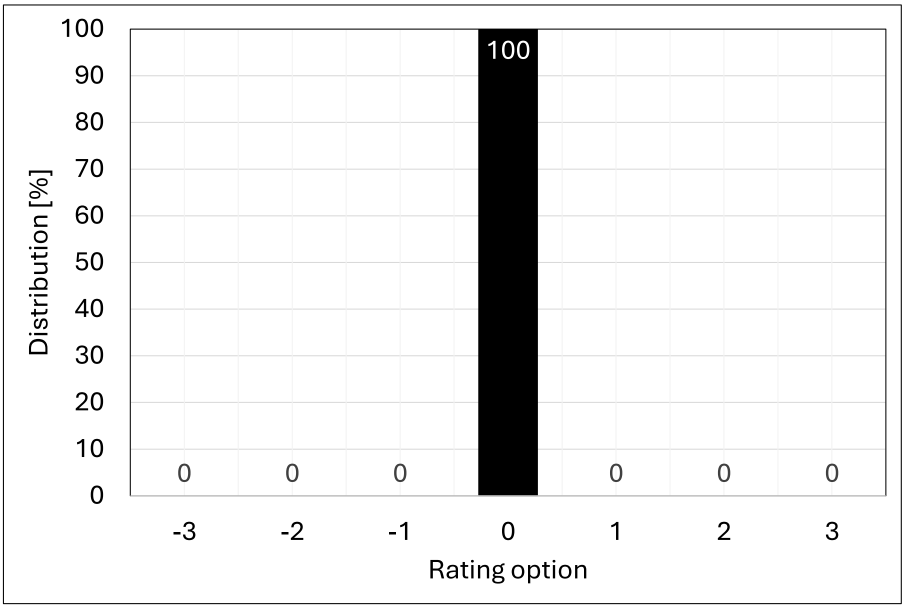

Figure 4.1: Theoretical rating distribution based on the lack of objective differences.

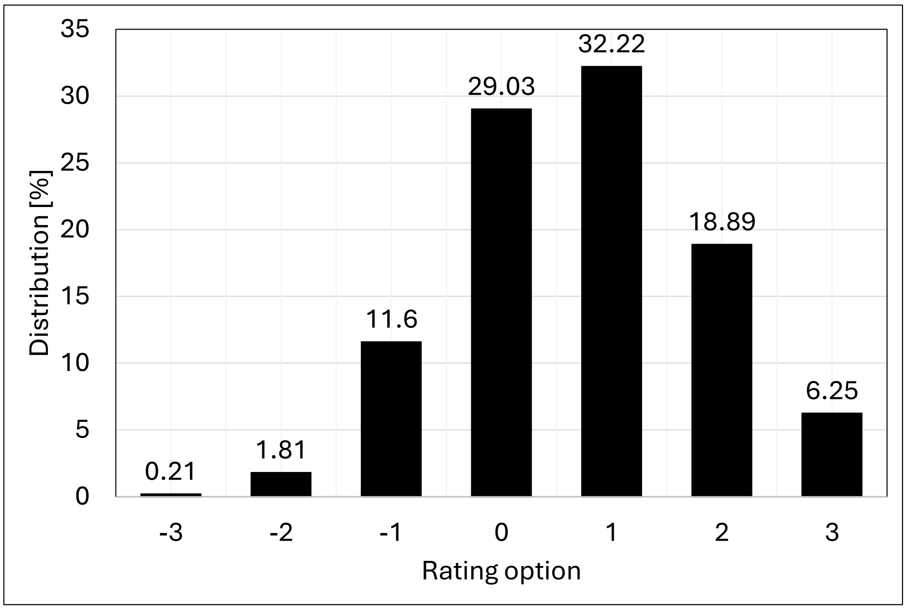

Figure 4.2: Overall rating distribution of the subjective study.

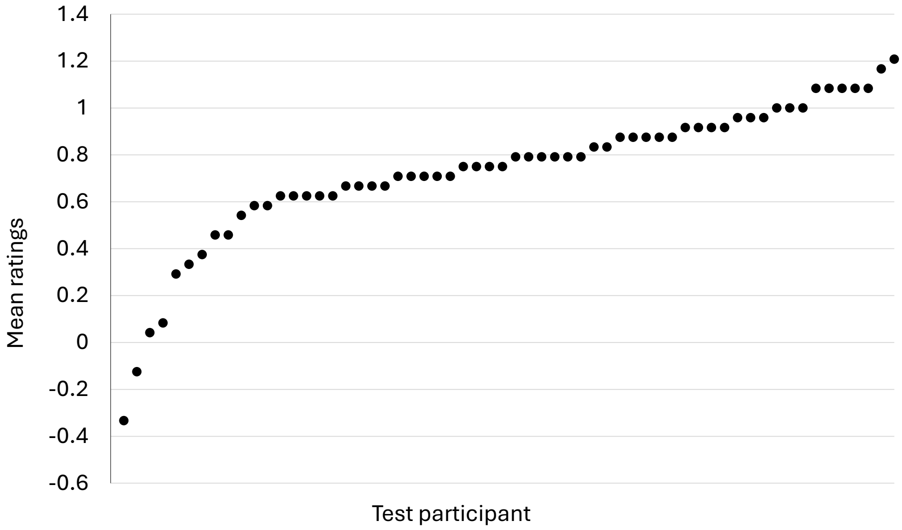

Figure 4.3: Mean ratings of the test participants.

perceived to perform better than the "not optimized" stimulus; however, 13.61% of the ratings were negative. The most frequently selected option was "Slightly better" (+1), while the least selected rating option was "Much worse" (−3), with only three instances recorded. Moreover, the "Better" (+2) rating was selected more often than the "Slightly worse" (−1) rating, showing a more favorable subjective evaluation of the "optimized" stimuli. It may be expected that when such a rating scale is applied, the two slight options are the most frequently selected among the rating scores. However, distribution similar to this one can also occur when a strong preference is exhibited [101, 102, 103]. The overall results reflect a positive assessment of the "optimized" stimulus compared to the "not optimized" one—despite the fact that the pair of stimuli were objectively identical. In comparison with the theoretical distribution, in which all ratings would be zero, Student's t-test provides $p < 0.01$, meaning that the labeling effect in the experiment resulted in statistically significant differences.

Regarding the findings from similar experiments in the scientific literature, specifically those comparing identical stimuli and evaluated by a 7-point comparison scale, a comparison of differentiating scores can be made. In this research, 70.97% of the ratings were differentiating, non-zero scores. In the study on HDR [102], the corresponding proportion was 77.75% for quality-related evaluations and 77.43% for stalling duration. In the context of UHD content [101], this value was 72.4%. In the study on rugged smartphones [103], the proportion of differentiating ratings was 56.06% when multiple devices were used, 66.91% with a single device, and 38.83% when the content was presented on a computer. These comparisons suggest that the labeling effect observed in the present study is among the greater instances of cognitive bias reported in the scientific literature.

The mean ratings of the test participants are illustrated in Figure 4.3. Each data point represents the mean rating of a test participant. The graph reveals that 56 out of 60 test participants had positive rating means, and 10 of them had means of 1 or higher. However, there were 2 participants who had positive mean ratings less than 0.2, and there were another 2 participants who had negative mean ratings. The average mean rating was 0.73.

The number of "The same" ratings are illustrated in Figure 4.4. Each data point represents the number of comparisons in which the test participant did not perceive any difference

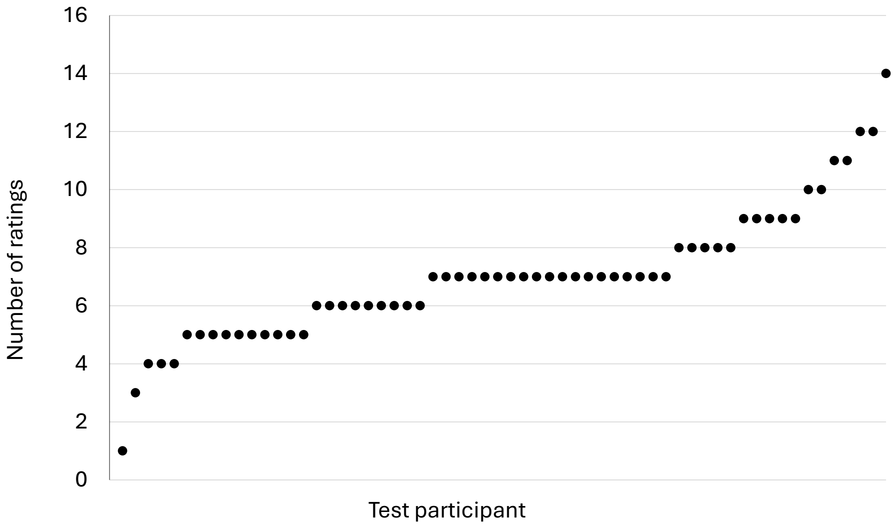

Figure 4.4: Number of “The same” (0) ratings provided by the test participants.

between the pair. One of the test participants throughout the experiment found only 1 out of 24 pairs to be “The same” (0); however, another test participant reported that there were 14 pairs that did not differ from each other. The average number of “The same” (0) ratings was 6.97.

4.2 Impact of label order

Figure 4.5 presents the rating distribution for the cases in which the first label was “not optimized” and the second was “optimized” (i.e., the “optimized” stimulus was compared to the “not optimized” stimulus). Figure 4.6, on the other hand, illustrates the distribution for the reverse case. In this analysis, the results of the latter configuration are inverted so that both distributions consistently reflect the perceived command execution delay of the “optimized” stimulus relative to the “not optimized” one. The mean ratings for these two groups were 0.77 and 0.69, respectively. For both order of labels, the order of rating frequency is the same as in the case of the overall results; however, notable differences can be observed. Specifically, when “optimized” was presented as the first label, the frequencies of the “The same” (0) and “Slightly better” (+1) ratings were reduced, meanwhile the frequencies of the other rating options increased. In addition, the proportions of “Slightly worse” (-1) and “Much better” (+3) ratings increased by approximately 5%. Despite these visible shifts in distribution, the difference between the two conditions is not statistically significant, as indicated by $p = 0.1$.

4.3 Impact of input types

The rating distributions for the single-input and continuous-input games are illustrated in Figures 4.7 and 4.8, respectively. The corresponding mean values were 0.72 for the single-input game and 0.74 for the continuous-input game. The most notable difference between the two distributions is the slightly higher frequency of “Slightly better” (+1) ratings and the lower frequency of “The same” (0) ratings in the continuous-input game—

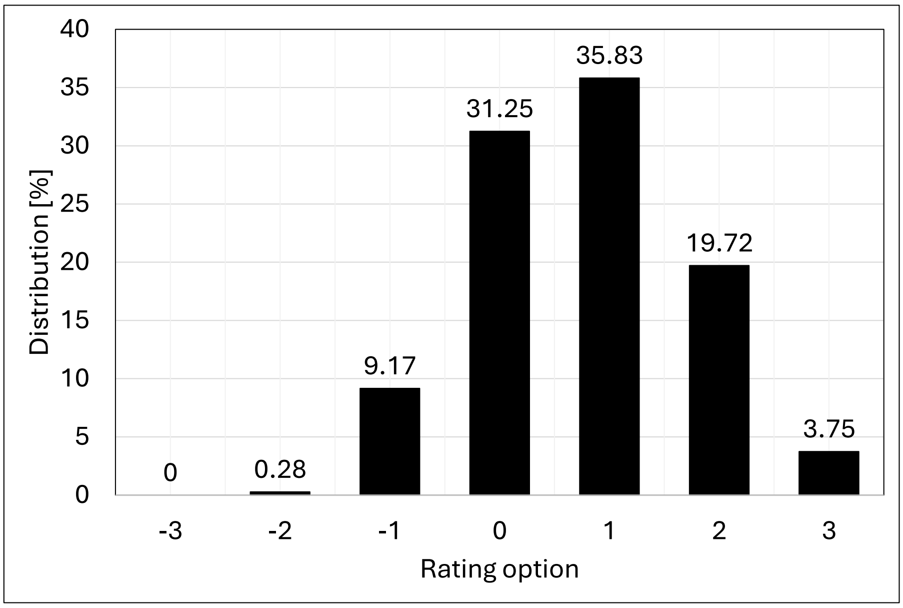

Figure 4.5: Rating distribution of tests where “not optimized” was the first label and “optimized” was the second label.

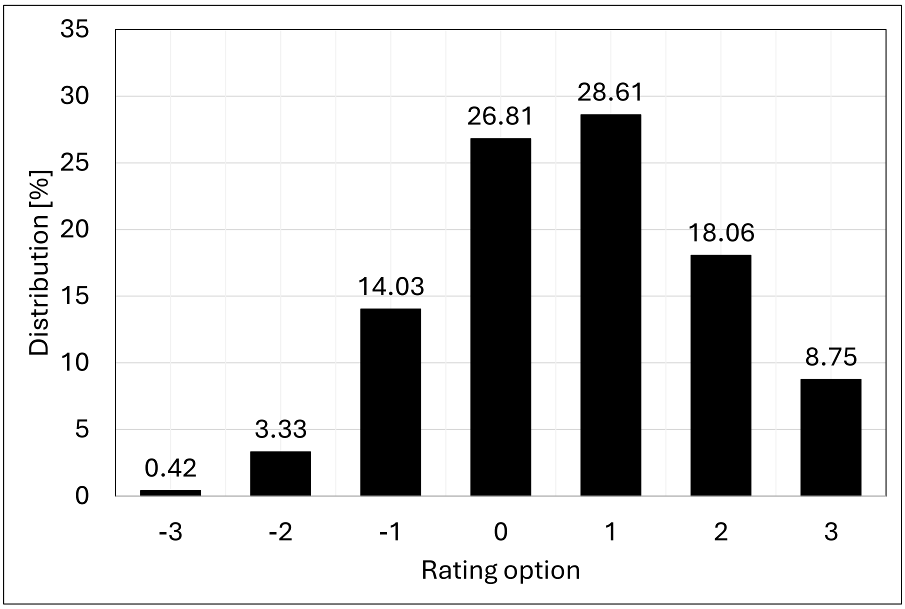

Figure 4.6: Rating distribution of tests where “optimized” was the first label and “not optimized” was the second label.

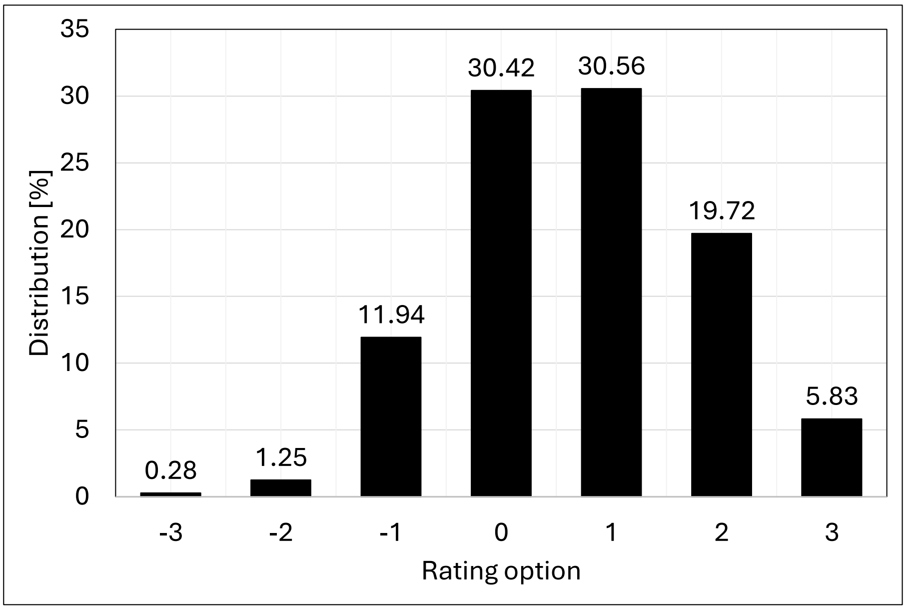

Figure 4.7: Rating distribution of the single-input game.

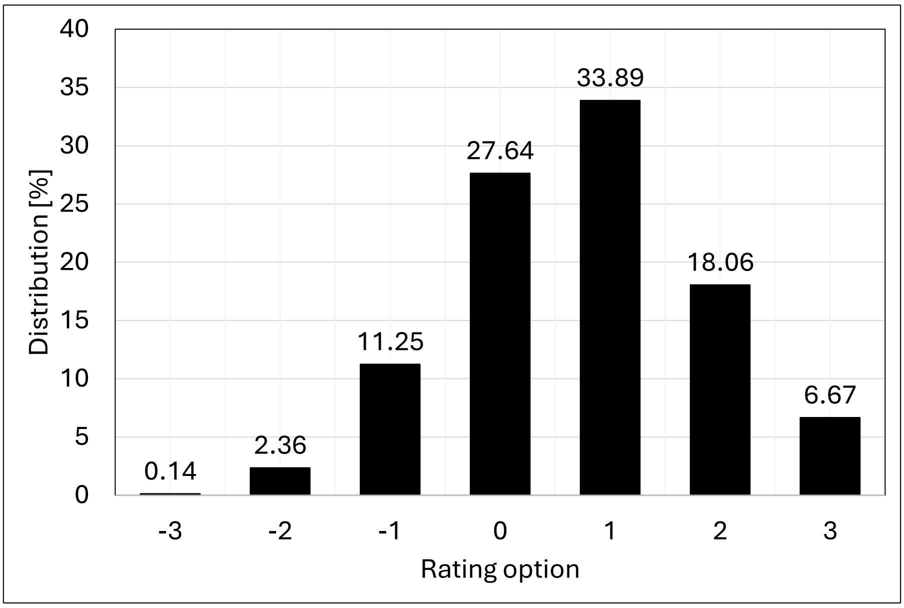

Figure 4.8: Rating distribution of the continuous-input game.

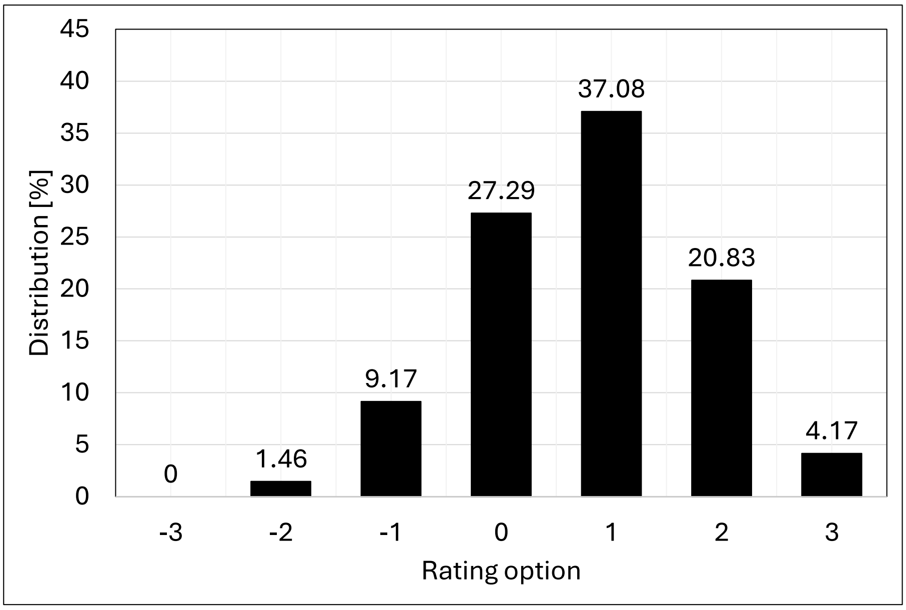

Figure 4.9: Rating distribution of games at easy difficulty.

each differing by approximately 3% when compared to the single-input game. Nonetheless, these differences are not statistically significant, as highlighted by $p = 0.41$.

4.4 Impact of game difficulty

The rating distributions for the easy, medium, and hard difficulty levels are presented in Figures 4.9, 4.10, and 4.11, respectively. The corresponding mean values were 0.79 for easy, 0.75 for medium, and 0.66 for hard difficulty. Although there was an approximate 10% difference in the frequency of "Slightly better" (+1) ratings between easy and medium difficulties, the mean values are relatively close, with $p = 0.26$ indicating no statistically significant difference. Similarly, the difference between medium and hard difficulties is not statistically significant, with $p = 0.15$; however, the comparison between easy and hard difficulties yields a statistically significant difference, with $p = 0.04$. Note that although the total proportion of differentiating (i.e., non-zero) ratings changed by only 2.5%, the distribution of those ratings varied considerably. Previous research on the influence of difficulty on player experience [30, 39] suggests that video games are most enjoyable when difficulty levels are kept within certain optimal thresholds. Deviating from these thresholds—either by increasing or decreasing difficulty too much—can negatively impact the player experience. Other studies emphasize the cognitive demands imposed by increased game difficulty [111, 112, 113], including those focusing on educational games [114]. For instance, Large et al. [112] investigated players of a MOBA game using cognitive tasks, and identified correlations between various cognitive traits (e.g., processing speed) and in-game performance (i.e., player rankings). Building on such findings, future research should explore the relationship between individual cognitive abilities—and the cognitive load induced by varying game difficulties—and rating behavior within the context of the labeling effect.

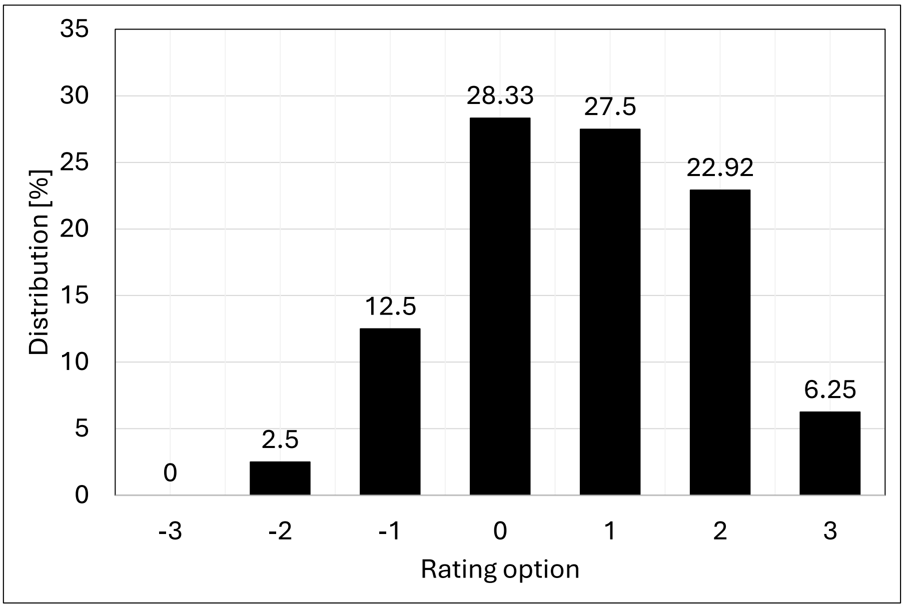

Figure 4.10: Rating distribution of games at medium difficulty.

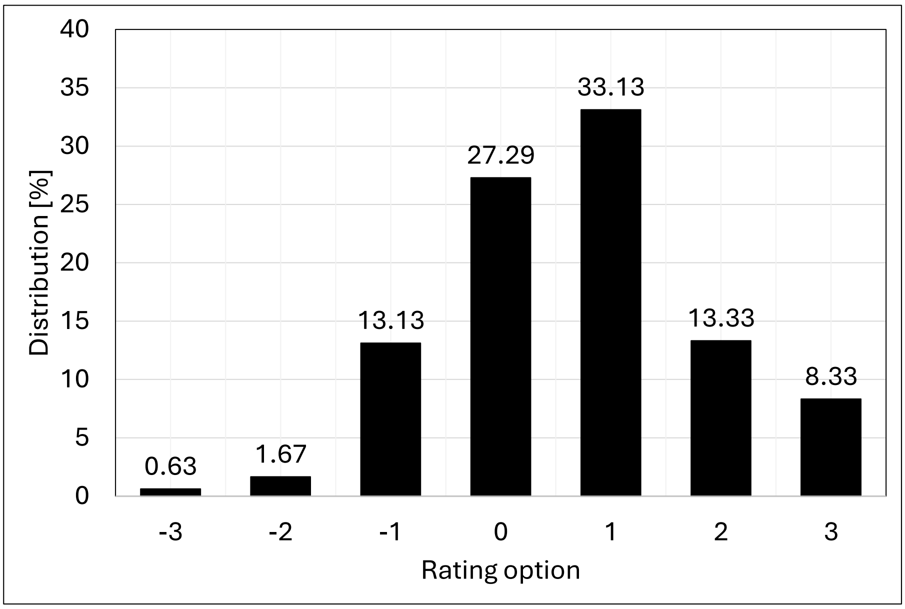

Figure 4.11: Rating distribution of games at hard difficulty.

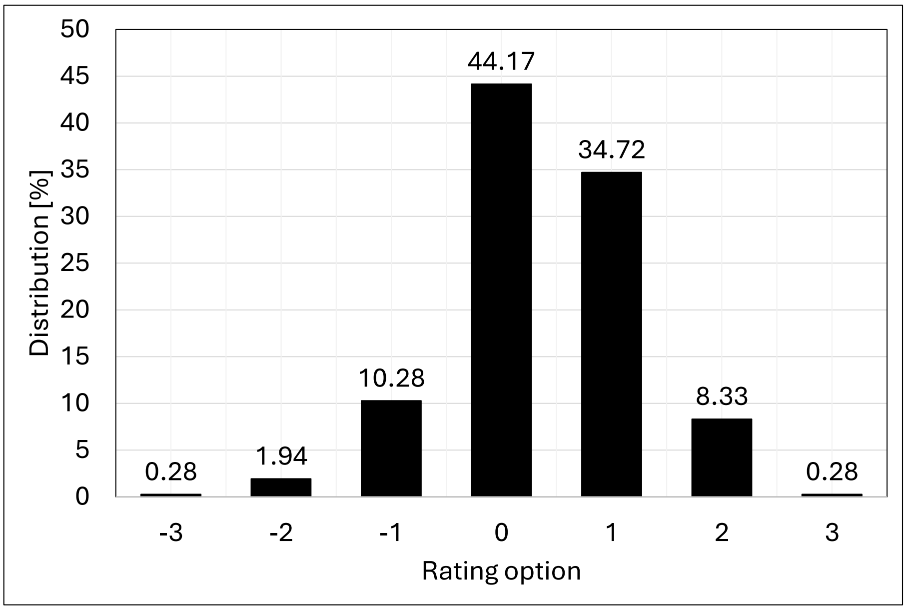

Figure 4.12: Rating distribution at no added command execution delay.

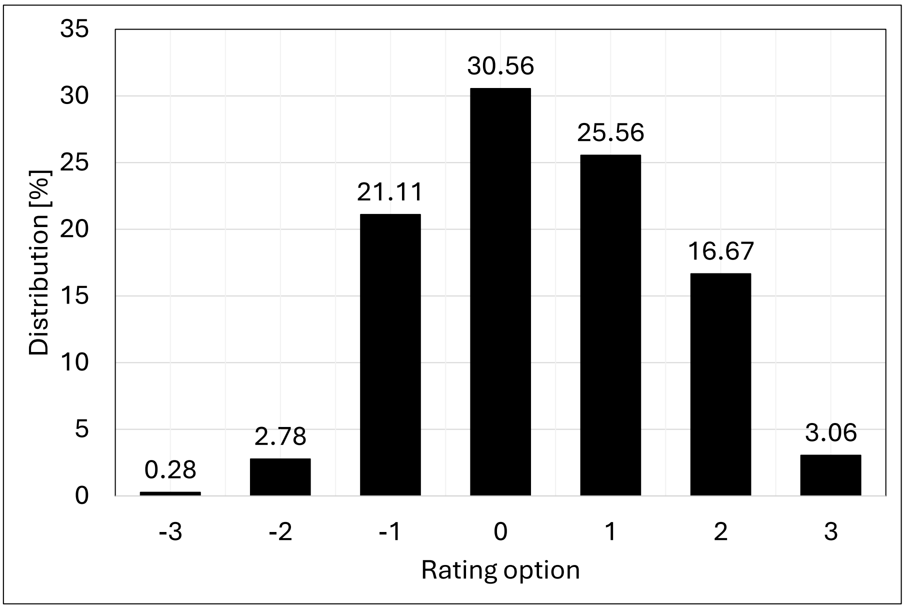

Figure 4.13: Rating distribution at 50 ms added command execution delay.

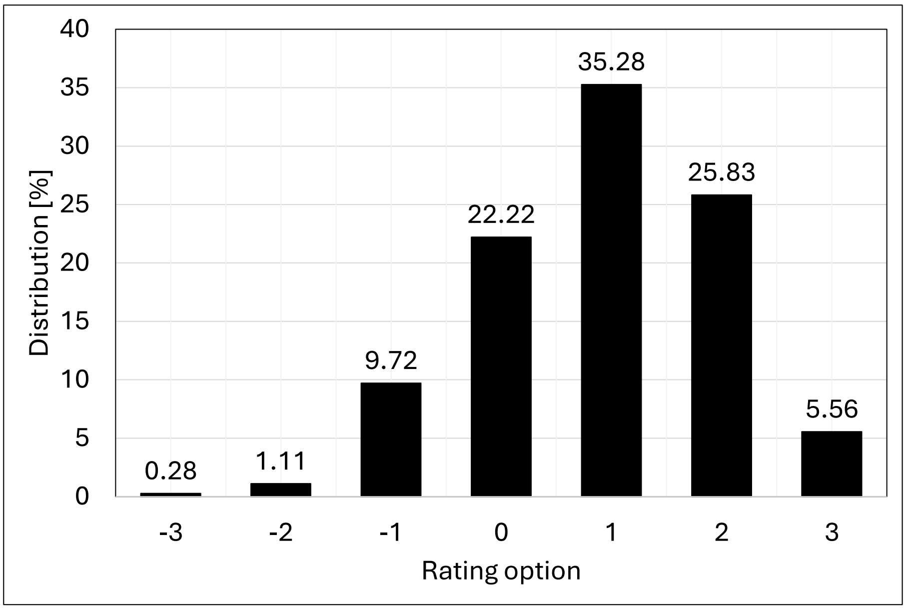

Figure 4.14: Rating distribution at 150 ms added command execution delay.

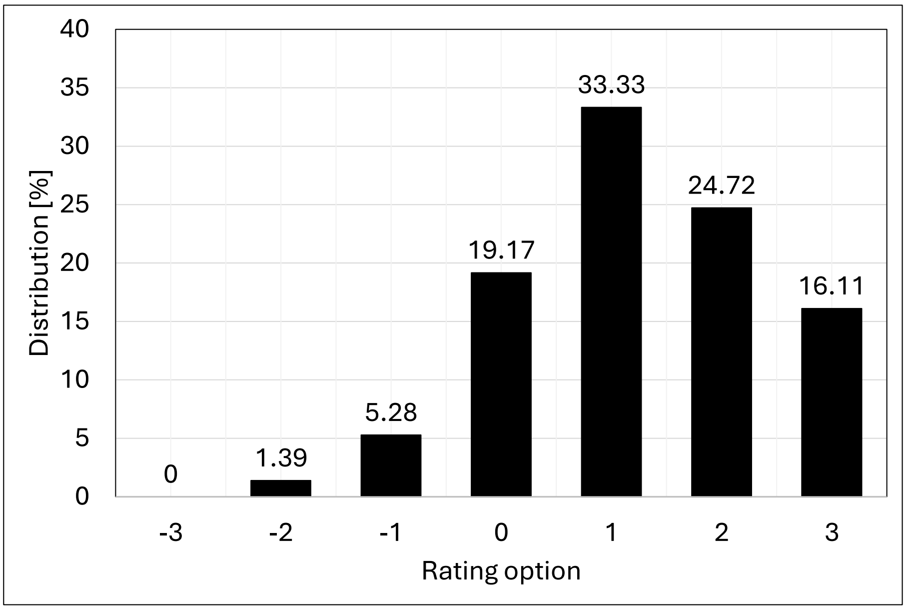

Figure 4.15: Rating distribution at 250 ms added command execution delay.

4.5 Impact of added command execution delay

The rating distributions at 0 ms, 50 ms, 150 ms, and 250 ms added command execution delay are shown in Figures 4.12, 4.13, 4.14, and 4.15, respectively. The means for these delay values were 0.37, 0.41, 0.91, and 1.23, respectively. When comparing added delays of 0 ms and 50 ms, the result is not statistically significant, as $p = 0.33$. However, all other comparisons yield $p < 0.01$, indicating statistically significant differences. The distribution of the ratings reveals a trend: as the amount of added delay increases, the proportion of differentiating, non-zero ratings also rises. Specifically, the percentages of non-zero ratings were 55.83% for 0 ms, 69.44% for 50 ms, 77.78% for 150 ms, and 80.83% for 250 ms. This consistent increase results in a strong positive correlation of 0.92 between the amount of added command execution delay and the frequency of non-zero ratings. Note that at the highest delay, "Much better" (+3) ratings reached 16.11%, nearly equal to the "The same" (0) ratings at 19.17%. It is likely that the term "optimized" was directly associated with the amount of delay. Consequently, as delay became more noticeable, it may have reinforced a confirmation bias—leading participants to perceive the "optimized" gameplay as superior. However, even when no delay was presented, more than half of the ratings indicated perceivable differences.

4.6 Discussion

In live service game environments, developers typically strive to meet player expectations by balancing gameplay, addressing bugs and issues, and introducing new contents. These additions are commonly delivered under labels such as "patch", "update", or "downloadable content"—the latter is more relevant when the additions and changes are more substantial. Zhong and Xu [115] found that frequent updates help maintain player engagement and satisfaction—a conclusion also supported by Liu and Samiee [116]. Furthermore, research by Claypool et al. [117] highlights that new patches can influence player behavior and even inspire entirely new playstyles [118]. However, Del Gallo [119] argues that in the case of the MOBA game League of Legends, developers may use patches strategically to divert attention from persistent, unresolved issues within the game. Patches do not always introduce new content; they may instead focus on technological enhancements or security improvements. Arora et al. [120] examined the effects of software security patches, while Lin et al. [121] proposed a model for the automatic generation of such patches. Security-related updates are frequently implemented to prevent cheating and other forms of exploitation within games [122]. Thus, these patches are often positively received by players, as suggested by Jussila et al. [123] and Truelove et al. [124]. The study of Mertens [125] concludes that games with long lists of bugs and errors can be considered "broken"; however, such a negative reputation may vanish as the developer releases new patches, updates, and downloadable contents. The author also argued that these "broken" games usually receive backlash at launch and player count eventually drops. Yet after receiving updates—the public notes of which usually describe the updated version of the game to have better performance and/or improved gameplay—players may be attracted to the game again. In games where low network latency is essential to gameplay (e.g., FTGs), the improvement of the utilized netcode is crucial [24, 66, 67, 68]. At the time of writing this thesis, rollback netcode is widely used. Its main idea is to compensate delay by predicting player input. Of course, if there is an actual difference between the predicted and the real player input, then the game session is reverted to an earlier state, and the correct input is carried out. The relevance of such research is that for certain video games, even a few frames worth of delay may have a significant impact on QoE and player performance. Therefore, any change to netcode communicated in patches or updates may have a noteworthy prior, preliminary influence on players that understand such mechanisms—and thus, may have preconceptions about the updated gameplay performance.

The result of the experiment presented in this thesis highlights the influence of the labeling effect within the context of gaming. The methodology did not introduce objective differences within the pairs of the pair comparison tests (i.e., the two game sequences were always identical), only between pairs, in accordance with the test conditions. According to my best knowledge, no such study has been carried out so far that addresses any aspect of gaming QoE where a video game patch or update is communicated to have improved performance yet there is zero difference. The findings of such a subjective evaluation are potentially significant for current practices in video game development and maintenance. One may hypothesize that for a portion of the test participants, the mismatch between expectations and the lack of real improvement may negatively affect certain aspects of QoE as shown in earlier research [126]; however, the labeling effect may instead yield an overall enhancement in perceived quality for other participants.

Chapter 5

Conclusions

In order to conclude the work performed within the scope of this thesis, first I shall address the completion of the thesis tasks, followed by a summary of the obtained results, and then finally I shall detail the potential future continuations of the research effort.

5.1 Completion of the thesis tasks

I hereby confirm that each and every thesis task was completed successfully. Their completion is summarized as follows:

1. "Review the related scientific literature and the relevant international standards."

   At the beginning of the thesis work, I conducted an extensive review of the related scientific literature and relevant international standards. This review focused on prior studies exploring user experience in video games, the impact of system performance on perceived quality, and the influence of cognitive bias—particularly the labeling effect. The investigation of international standards provided the knowledge to design practical, technically sound, and reproducible experimental setups for subjective QoE evaluation.

2. "Design the methodology for a subjective study that addresses the aforementioned form of cognitive bias."

   Based on the knowledge and expertise acquired during my studies at the university as well as during the preparation for this research, I designed a detailed methodology for a subjective study. The goal was to evaluate how descriptive labels—"optimized" and "not optimized"—influence the perception of video game performance, even when no objective differences exist. The methodology included two self-developed video games to simulate the test conditions that varied in terms of label order, input type, difficulty level, and command execution delay. Four extents of added delay were presented: 0 ms, 50 ms, 150 ms, 250 ms; these values were chosen based on the measured values of delay in real-world gaming scenarios, as well as JND and thresholds of delay from prior research efforts published in the scientific literature. The three difficulties were: easy, medium, and hard; as the difficulty increased, the game became faster—which required faster reaction from the players, and increased the relevance of added delay. A user interface was designed to present the label of the test sequence, which was also clearly communicated in a brief segment prior to each sequence. The first label in each comparison was either “optimized” or “not optimized”, consistently for each participant throughout the study.

3. “Implement the experimental setup and carry out the subjective tests.”

   The experimental setup was successfully implemented, including all necessary components for conducting the subjective tests. The results were collected digitally via a standardized 7-point rating scale—which was used in subjective studies sharing similar methodological characteristics [101, 102, 103]. The subjective tests followed the best practices of the scientific community, following the detailed experimental setup. A total of 60 test participant successfully completed the subjective study, and provided valid results.

4. “Analyze the collected data and assess the correlation between the test variables and the obtained subjective scores.”

   The data collected from the subjective tests was thoroughly analyzed to identify patterns and correlations between the test variables and the participants’ subjective ratings. Analytical and statistical methods, including Student’s t-test, were applied to determine the significance of the impact of label order, input type, game difficulty, added command execution delay, and the labeling effect in general. The obtained results are summarized below.

5.2 Summary of the obtained results

The results of the subjective study demonstrated the statistically significant influence of the labeling effect on the perception of command execution delay in video games. In general, the null hypothesis can be rejected confidently as the measured differences were not due to random error. Although the compared game sequences were objectively identical, over 70% of the participants’ ratings indicated perceived differences, revealing a clear impact of the “optimized” versus “not optimized” labels. Specifically, 57.36% of the ratings favored sequences with the “optimized” label, while only 13.61% perceived better responsiveness for sequences with the “not optimized” label. The influence of the labels remained consistent across different label orders and input types (single-input and continuous-input games), with no statistically significant differences found between them. However, statistically significant differences were observed between the easy and hard difficulty levels, as well as across varying extents of added command execution delay. A strong positive correlation was identified between the extent of the added delay and the impact of the labeling effect, suggesting that greater delay values may strengthen the effect of cognitive bias.

5.3 Future work

In this thesis, I investigated the impact of the labeling effect on the perception of command execution delay, the added values of which ranged from 0 ms up to 250 ms. However, the chosen range of added delay in this study was limited based on the tolerance levels measured from prior research efforts. Thus, greater extents of added delay may result different rating behavior.

One notable limitation of my study is the absence of objective user performance measurement (e.g., the score that players obtained when they finish each stimulus). Future research should also take into account objective performance in order to enable a better understanding of the potentially associated factors. Additionally, such gameplay-related characteristics could offer a deeper insight into user behavior. For instance, at higher difficulty levels, sequence abandonment may occur—where the player either intentionally provides incorrect inputs to end the session early or simply stop providing input. Detecting such instance could support data curation and enable more accurate data analysis. Another potential phenomenon is that players may project their own operational errors onto the stimulus—such as pressing the wrong key or reacting too late. This form of projection could substantially influence subjective assessment. Note that experiments that last for extended periods of time may result fatigue, which is particularly applicable to studies that rely on the reaction time of the individual. Fatigue may result more instances of operational error, making the aforementioned phenomenon even more relevant. Regarding command execution delay, while the work completely neglected the variation of delay, future research could benefit from examining the perception of jitter in similar experimental contexts.

Beyond command execution delay, other game-related characteristics—such as frame rate—should also be examined, as they may significantly impact user experience. Future research should further explore potential correlations involving viewing conditions, display properties, levels of immersion, and perceptual fatigue—all of which may affect the influence of the labeling effect, as well as the overall QoE. For example, different combinations of display size, display resolution, and viewing distance may influence the extent to which the labeling effect impacts perceived visual quality. While the effect of viewing distance on perceived quality has been thoroughly investigated in prior work [127, 128, 129], its relationship to the labeling effect and other forms of cognitive bias remains largely unexplored. Physiological measurements could offer deeper insights into the impact of cognitive bias. Such measurements are also commonly used to assess immersion [130, 131, 132], particularly in relation to emerging immersive technologies like VR [133, 134, 135]. Future research should explore the correlation between the level of immersion and susceptibility to cognitive bias. Finally, it would be particularly relevant for the gaming industry to investigate the impact of performance patches or updates that introduce no objective change, in order to quantify their potential positive or negative influence on gaming QoE.

Acknowledgements

First and foremost, I would like to thank all the test participants who took part in the subjective study. Their time, effort, and cooperation were essential to the success of this research, and I am deeply grateful for their contribution.

I would like to express my sincere gratitude to my thesis supervisor, Dr. Peter A. Kara, for his continuous guidance, support, and encouragement throughout the course of this work. His expertise, constructive feedback, and patience have been invaluable to the completion of this thesis and to my academic development.

Lastly, I would also like to extend my heartfelt thanks to my parents for their unwavering love, support, and for believing in me. Their encouragement and sacrifices have been the foundation of my education and personal growth. Without their support, this achievement would not have been possible.

List of Figures

List of Tables

List of Abbreviations

ACR	absolute category rating
ALC	Advance Lag Compensation
DDA	Dynamic Difficulty Adjustment
DCR	degradation category rating
EBSE	evidence-based software engineering
fMRI	functional magnetic resonance imaging
FPS	first-person shooter
FTG	fighting game
GEQ	Gaming Experience Questionnaire
GSAQ	Gaming Specified Attributing Questionnaire
HCI	human-computer interaction
HD	high-definition
HDR	high dynamic range
ITU	International Telecommunication Union
JND	just noticeable difference
LTE	long-term evolution
MMORPG	massively multiplayer online role-playing game
MOBA	multiplayer online battle arena
PC	personal computer
QoE	Quality of Experience
RMT	requirement management tool
RPG	role-playing game
RTS	real-time strategy
UHD	ultra-high-definition
VAT	Value Added Tax
VR	virtual reality

Bibliography

[1] Xun Xi, Baoxing Xi, Chenglin Miao, Rongjian Yu, Jie Xie, Rong Xiang, and Feng Hu. Factors influencing technological innovation efficiency in the Chinese video game industry: Applying the meta-frontier approach. *Technological Forecasting and Social Change*, 178:121574, 2022.

[2] Dmitry Andreev. Real-time frame rate up-conversion for video games: or how to get from 30 to 60 fps for “free”. In *ACM SIGGRAPH 2010 Talks*, pages 1–1. 2010.

[3] Aline Normoyle, Gina Guerrero, and Sophie Jörg. Player perception of delays and jitter in character responsiveness. In *Proceedings of the ACM Symposium on Applied Perception*, pages 117–124, 2014.

[4] Andreas Wilke and Rui Mata. Cognitive bias. 2017.

[5] Philipp Scharfenberger, Jan R Landwehr, Claire I Tsai, Jenny L Zimmermann, Andreas Herrmann, Ann L McGill, et al. How consumption vocabulary directs product discussions: the guiding influence of feature labels on consumers’ communication and learning about products in online communities. *Journal of Marketing Behavior*, 4(2-4):173–202, 2020.

[6] Mark Claypool, Kajal Claypool, and Feissal Damaa. The effects of frame rate and resolution on users playing first person shooter games. In *Multimedia Computing and Networking 2006*, volume 6071, page 607101. SPIE, 2006.

[7] Shengmei Liu, Mark Claypool, Atsuo Kuwahara, Jamie Sherman, and James J Scovell. Lower is Better? The Effects of Local Latencies on Competitive First-person Shooter Game Players. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021.

[8] Kuan-Ta Chen, Polly Huang, and Chin-Laung Lei. Effect of network quality on player departure behavior in online games. *IEEE Transactions on Parallel and Distributed Systems*, 20(5):593–606, 2008.

[9] Athanasios Kokkinakis, Peter York, Moni S Patra, Justus Robertson, Ben Kirman, Alistair Coates, Alan P Chitayat, Simon Demediuk, Anders Drachen, Jonathan Hook, et al. Metagaming and metagames in esports. *International Journal of Esports*, 1(1), 2021.

[10] Stephanie Boluk and Patrick LeMieux. Metagaming: Videogames and the Practice of Play. *Computer Games and New Media Cultures*, 2017.

[11] P.800: Methods for subjective determination of transmission quality. `https://www.itu.int/rec/T-REC-P.800-199608-I`, 2019. [Online; accessed 22-05-2025].

[12] P.910: Subjective video quality assessment methods for multimedia applications . https://www.itu.int/rec/t-rec-p.910, 2024.
[Online; accessed 22-05-2025].

[13] BT.500: Methodologies for the subjective assessment of the quality of television images.
https://www.itu.int/rec/R-REC-BT.500/en, 2023.
[Online; accessed 22-05-2025].

[14] Sebastodes Moller, Steven Schmidt, and Saman Zadtootaghaj. New ITU-T Standards for Gaming QoE Evaluation and Management. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2018.

[15] Sebastian Möller, Jan-Niklas Antons, Justus Beyer, Sebastian Egger, Elena Núñez Castellar, Lea Skorin-Kapov, and Mirko Sužnjević. Towards a new ITU-T recommendation for subjective methods evaluating gaming QoE. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2015.

[16] SG12 - Performance, quality of service (QoS) and quality of experience (QoE).
https://www.itu.int/en/ITU-T/studygroups/2022\protect\@normalcr\relax-2024/12/Pages/default.aspx, 2024.
[Online; accessed 22-05-2025].

[17] Duy H Nguyen and Peter A Kara. The influence of the labeling effect on the perception of command execution delay in gaming. *Multimodal Technologies and Interaction*, 9(5), 2025.

[18] Ben Shneiderman. Response time and display rate in human performance with computers. *ACM Computing Surveys (CSUR)*, 16(3):265–285, 1984.

[19] Kjetil Raaen and Ragnhild Eg. Instantaneous human-computer interactions: Button causes and screen effects. In *Human-Computer Interaction: Users and Contexts: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III 17*, pages 492–502. Springer, 2015.

[20] Enlie Wang. User's Delay Perception and Tolerance in Human-Computer Interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 46, pages 651–655. SAGE Publications Sage CA: Los Angeles, CA, 2002.

[21] Florian Metzger, Albert Rafetseder, and Christian Schwartz. A comprehensive end-to-end lag model for online and cloud video gaming. In *Proc. 5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*, pages 15–19, 2016.

[22] Lothar Pantel and Lars C Wolf. On the impact of delay on real-time multiplayer games. In *Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, pages 23–29, 2002.

[23] Steven Schmidt, Saman Zadtootaghaj, and Sebastian Möller. Towards the delay sensitivity of games: There is more than genres. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[24] Mark Claypool and Kajal Claypool. Latency can kill: precision and deadline in online games. In *Proceedings of the first annual ACM SIGMM conference on Multimedia systems*, pages 215–222, 2010.

[25] Tobias Fritsch, Hartmut Ritter, and Jochen Schiller. The effect of latency and network limitations on mmorpgs: a field study of everquest2. In *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*, pages 1–9, 2005.

[26] Kuan-Ta Chen, Yu-Chun Chang, Po-Han Tseng, Chun-Ying Huang, and Chin-Laung Lei. Measuring the latency of cloud gaming systems. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1269–1272, 2011.

[27] Teemu Kämäräinen, Matti Siekkinen, Antti Ylä-Jääski, Wenxiao Zhang, and Pan Hui. A measurement study on achieving imperceptible latency in mobile cloud gaming. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 88–99, 2017.

[28] Robert S Allison, Laurence R Harris, Michael Jenkin, Urszula Jasiobedzka, and James E Zacher. Tolerance of temporal delay in virtual environments. In *Proceedings IEEE Virtual Reality 2001*, pages 247–254. IEEE, 2001.

[29] Pingguo Huang, Ryo Arima, and Yutaka Ishibashi. Influence of network delay on human perception of weight in virtual environment. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pages 1221–1225. IEEE, 2017.

[30] Christoph Klimmt, Christopher Blake, Dorothée Hefner, Peter Vorderer, and Christian Roth. Player performance, satisfaction, and video game enjoyment. In *Entertainment Computing–ICEC 2009: 8th International Conference, Paris, France, September 3-5, 2009. Proceedings 8*, pages 1–12. Springer, 2009.

[31] Mona Erfani, Magy S El-Nasr, David Milam, Bardia Aghabeigi, Beth A Lameman, Bernhard E Riecke, Hamid Maygoli, and Sang Mah. The effect of age, gender, and previous gaming experience on game play performance. In *Human-Computer Interaction: Second IFIP TC 13 Symposium, HCIS 2010, Held as Part of WCC 2010, Brisbane, Australia, September 20-23, 2010. Proceedings*, pages 293–296. Springer, 2010.

[32] Toby Hopp and Jolene Fisher. Examination of the Relationship Between Gender, Performance, and Enjoyment of a First-Person Shooter Game. *Simulation & Gaming*, 48(3):338–362, 2017.

[33] Ansgar E Depping and Regan L Mandryk. Why is this happening to me? how player attribution can broaden our understanding of player experience. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1040–1052, 2017.

[34] Mark Claypool and David Finkel. The effects of latency on player performance in cloud-based games. In *2014 13th Annual Workshop on Network and Systems Support for Games*, pages 1–6. IEEE, 2014.

[35] Victor Clincy and Brandon Wilgor. Subjective evaluation of latency and packet loss in a cloud-based game. In *2013 10th International Conference on Information Technology: New Generations*, pages 473–476. IEEE, 2013.

[36] Shaolong Li, Changja Chen, and Lei Li. Evaluating the Latency of Clients by Player Behaviors in Client-Server Based Network Games. In *2008 3rd International Conference on Innovative Computing Information and Control*, pages 375–375. IEEE, 2008.

[37] Tom Beigbeder, Rory Coughlan, Corey Lusher, John Plunkett, Emmanuel Agu, and Mark Claypool. The effects of loss and latency on user performance in unreal tournament 2003®. In *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games*, pages 144–151, 2004.

[38] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1052–1063, 2011.

[39] Justin T Alexander, John Sear, and Andreas Oikonomou. An investigation of the effects of game difficulty on player enjoyment. *Entertainment Computing*, 4(1):53–62, 2013.

[40] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529, 2009.

[41] Gustavo Andrade, Geber Ramalho, Hugo Santana, and Vincent Corruble. Challenge-Sensitive Action Selection: an Application to Game Balancing. In *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 194–200. IEEE, 2005.

[42] Alexander Baldwin, Daniel Johnson, and Peta A Wyeth. The effect of multiplayer dynamic difficulty adjustment on the player experience of video games. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1489–1494. 2014.

[43] Saeed S Sabet, Steven Schmidt, Carsten Griwodz, and Sebastian Möller. Towards the Impact of Gamers' Adaptation to Delay Variation on Gaming Quality of Experience. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2019.

[44] Saeed S Sabet, Steven Schmidt, Saman Zadtootaghaj, Babak Naderi, Carsten Griwodz, and Sebastian Möller. A latency compensation technique based on game characteristics to mitigate the influence of delay on cloud gaming quality of experience. In *Proceedings of the 11th ACM Multimedia Systems Conference*, pages 15–25, 2020.

[45] Cheryl Savery, Nicholas Graham, Carl Gutwin, and Michelle Brown. The effects of consistency maintenance methods on player experience and performance in networked games. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative work & Social Computing*, pages 1344–1355, 2014.

[46] Steven WK Lee and Rocky KC Chang. Enhancing the experience of multiplayer shooter games via advanced lag compensation. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pages 284–293, 2018.

[47] Shengmei Liu, Mark Claypool, Bhuvana Devigere, Atsuo Kuwahara, and Jamie Sherman. 'Git Gud!'–Evaluation of Self-Rated Player Skill Compared to Actual Player Performance. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, pages 306–310, 2020.

[48] Hay MMM Lwin, Yutaka Ishibashi, and Khin T Mya. Influence of Voice Delay on Human Perception of Group Synchronization Error for Remote Learning: One-way communication case. In *2020 IEEE Conference on Computer Applications (ICCA)*, pages 1–5. IEEE, 2020.

[49] Phisit Pornpongtechavanich, Pongpisit Wuttidittachotti, and Therdpong Daengsi. QoE modeling for audiovisual associated with MOBA game using subjective approach. *Multimedia Tools and Applications*, 81(26):37763–37779, 2022.

[50] Kjetil Raaen and Andreas Petlund. How much delay is there really in current games? In *Proceedings of the 6th ACM Multimedia Systems Conference*, pages 89–92, 2015.

[51] Michael Long and Carl Gutwin. Characterizing and modeling the effects of local latency on game performance and experience. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, pages 285–297, 2018.

[52] Mark Claypool, Ragnhild Eg, and Kjetil Raaen. Modeling user performance for moving target selection with a delayed mouse. In *International Conference on Multimedia Modeling*, pages 226–237. Springer, 2016.

[53] Mark Claypool, Andy Cockburn, and Carl Gutwin. The impact of motion and delay on selecting game targets with a mouse. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(2s):1–24, 2020.

[54] Zenja Ivkovic, Ian Stavness, Carl Gutwin, and Steven Sutcliffe. Quantifying and mitigating the negative effects of local latencies on aiming in 3d shooter games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 135–144, 2015.

[55] Peter Quax, Anastasiia Beznosyk, Wouter Vanmontfort, Robin Marx, and Wim Lamotte. An evaluation of the impact of game genre on user experience in cloud gaming. In *2013 IEEE International Games Innovation Conference (IGIC)*, pages 216–221. IEEE, 2013.

[56] Saeed S Sabet, Steven Schmidt, Saman Zadtootaghaj, Carsten Griwodz, and Sebastian Möller. Delay sensitivity classification of cloud gaming content. In *Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems*, pages 25–30, 2020.

[57] Anastasiia Beznosyk, Peter Quax, Karin Coninx, and Wim Lamotte. Influence of network delay and jitter on cooperation in multiplayer games. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 351–354, 2011.

[58] Justus Beyer and Sebastian Möller. Assessing the Impact of Game Type, Display Size and Network Delay on Mobile Gaming QoE. *PIK-Praxis der Informationsverarbeitung und Kommunikation*, 37(4):287–295, 2014.

[59] Mark Claypool. The effect of latency on user performance in real-time strategy games. *Computer Networks*, 49(1):52–70, 2005.

[60] Sue Pedri and Beryl Hesketh. TIME PERCEPTION: EFFECTS OF TASK SPEED AND DELAY. *Perceptual and Motor Skills*, 76(2):599–608, 1993.

[61] Christin Kohrs, Nicole Angenstein, and André Brechmann. Delays in Human-Computer Interaction and Their Effects on Brain Activity. *PloS one*, 11(1):e0146250, 2016.

[62] Kaliappa Ravindran, Ali Sabbir, and Balachandran Ravindran. Impact of network loss/delay characteristics on consistency control in real-time multi-player games. In *2008 5th IEEE Consumer Communications and Networking Conference*, pages 1128–1133. IEEE, 2008.

[63] Anh Le and Yanni E Liu. Fairness in Multi-Player Online Games on Deadline-Based Networks. In *CCNC*, pages 670–675, 2007.

[64] Sebastian F Lindström, Markus Wetterberg, and Niklas Carlsson. Cloud gaming: A QoE study of fast-paced single-player and multiplayer gaming. In *2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)*, pages 34–45. IEEE, 2020.

[65] Kristin Siu, Matthew Guzdial, and Mark O Riedl. Evaluating singleplayer and multiplayer in human computation games. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, pages 1–10, 2017.

[66] Alain Lioret, Lior Diler, Sami Dalil, and Marion Mota. Hybrid Prediction for Games' Rollback Netcode. In *ACM SIGGRAPH 2022 Posters*, pages 1–2. 2022.

[67] Huynh, E.; Valarino, F. An analysis of continuous consistency models in real time peer-to-peer fighting games.
`https://www.diva-portal.org/smash/get/diva2:1322881/FULLTEXT01.pdf`, 2019.
[Online; accessed 22-05-2025].

[68] Ehlert, A. Improving input prediction in online fighting games.
`https://www.diva-portal.org/smash/get/diva2:1560069/FULLTEXT01.pdf`, 2021.
[Online; accessed 22-05-2025].

[69] Mark Claypool. Game Input with Delay—Moving Target Selection with a Game Controller Thumbstick. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(3s):1–22, 2018.

[70] Michael Long and Carl Gutwin. Effects of local latency on game pointing devices and game pointing tasks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[71] Peter Quax, Patrick Monsieurs, Wim Lamotte, D.De Vleeschauwer, and Natalie Degrande. Objective and subjective evaluation of the influence of small amounts of delay and jitter on a recent first person shooter game. In *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games*, pages 152–156, 2004.

[72] Jingxi Xu and Benjamin W Wah. Concealing network delays in delay-sensitive online interactive games based on just-noticeable differences. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2013.

[73] Chi E Tan, Wee H Tan, Siti F Shamsudin, Sujata Navaratnam, and Yiing Y Ng. Investigating the Impact of Latency in Mobile-Based Multiplayer Online Battle Arena (MOBA) Games. *International Journal of Creative Multimedia*, 3(1):1–16, 2022.

[74] Oliver Hohlfeld, Hannes Fiedler, Enric Pujol, and Dennis Guse. Insensitivity to Network Delay: Minecraft Gaming Experience of Casual Gamers. In *2016 28th International Teletraffic Congress (ITC 28)*, volume 3, pages 31–33. IEEE, 2016.

[75] James Nichols and Mark Claypool. The effects of latency on online madden NFL football. In *Proceedings of the 14th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pages 146–151, 2004.

[76] Melissa K Stern and James H Johnson. Just noticeable difference. *The corsini encyclopedia of psychology*, pages 1–2, 2010.

[77] Peter C Wason. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3):129–140, 1960.

[78] Peter C Wason. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281, 1968.

[79] Joshua Klayman. Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32:385–418, 1995.

[80] John M Darley and Paget H Gross. A Hypothesis-confirming Bias in Labeling Effects. *Journal of Personality and Social Psychology*, 44(1):20, 1983.

[81] Martin Jones and Robert Sugden. Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50:59–99, 2001.

[82] Nobuyuki Sakai, Sumio Imada, Sachiko Saito, Tatsu Kobayakawa, and Yuichi Deguchi. The effect of visual images on perception of odors. *Chemical Senses*, 30(suppl_1):i244–i245, 2005.

[83] Ruth A Bentler, Diane P Niebuhr, Tiffany A Johnson, and Gregory A Flamme. Impact of digital labeling on outcome measures. *Ear and Hearing*, 24(3):215–224, 2003.

[84] Victor Iglesias. Preconceptions About Service: How Much Do They Influence Quality Evaluations? *Journal of Service Research*, 7(1):90–103, 2004.

[85] Zhifeng Gao and Ted C Schroeder. Effects of Label Information on Consumer Willingness-to-pay for Food Attributes. *American Journal of Agricultural Economics*, 91(3):795–809, 2009.

[86] Klaas-Andries D Graaf, Peng Liang, Antony Tang, and Hans V Vliet. The impact of prior knowledge on searching in software documentation. In *Proceedings of the 2014 ACM Symposium on Document Engineering*, pages 189–198, 2014.

[87] Henrieta H Chovanová, Aleksander I Korshunov, and Dagmar Babčanová. Impact of brand on consumer behavior. *Procedia Economics and Finance*, 34:615–621, 2015.

[88] Kostas Stylidis, Casper Wickman, and Rikard Söderberg. Perceived Quality of Products: A Framework and Attributes Ranking Method. *Journal of Engineering Design*, 31(1):37–67, 2020.

[89] Paula M Fitzgerald, Donovan K Russo, Jeremy Kees, and John Kozup. How confusion impacts product labeling perceptions. *Journal of Consumer Marketing*, 36(2):306–316, 2019.

[90] Fabian Christandl, Detlef Fetchenhauer, and Erik Hoelzl. Price Perception and Confirmation Bias in the Context of a VAT Increase. *Journal of Economic Psychology*, 32(1):131–141, 2011.

[91] Webb Stacy and Jean MacMillan. Cognitive bias in software engineering. *Communications of the ACM*, 38(6):57–63, 1995.

[92] Rahul Mohanani, Iflaah Salman, Burak Turhan, Pilar Rodríguez, and Paul Ralph. Cognitive biases in software engineering: A systematic mapping study. *IEEE Transactions on Software Engineering*, 46(12):1318–1339, 2018.

[93] Magne Jørgensen and Efi Papatheocharous. Believing is Seeing: Confirmation Bias Studies in Software Engineering. In *2015 41st Euromicro Conference on Software Engineering and Advanced Applications*, pages 92–95. IEEE, 2015.

[94] Gul Calikli and Ayse Bener. Empirical Analyses of the Factors Affecting Confirmation Bias and the Effects of Confirmation Bias on Software Developer/Tester Performance. In *Proceedings of the 6th International Conference on Predictive Models in Software Engineering*, pages 1–11, 2010.

[95] Laura M Leventhal, Barbee E Teasley, and Diane S Rohlman. Analyses of factors related to positive test bias in software testing. *International Journal of Human-Computer Studies*, 41(5):717–749, 1994.

[96] Iflaah Salman. Cognitive biases in software quality and testing. In *Proceedings of the 38th International Conference on Software Engineering Companion*, pages 823–826, 2016.

[97] Gül Çalıklı and Ayşe B Bener. Influence of confirmation biases of developers on software quality: an empirical study. *Software Quality Journal*, 21:377–416, 2013.

[98] Austen Rainer and Sarah Beecham. A follow-up empirical evaluation of evidence based software engineering by undergraduate students. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*. BCS Learning & Development, 2008.

[99] Magne Jørgensen and Erik Løhre. First Impressions in Software Development Effort Estimation: Easy to Create and Difficult to Neutralize. In *16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012)*, pages 216–222. IET, 2012.

[100] Peter A Kara, Werner Robitza, Alexander Raake, and Maria G Martini. The Label Knows Better: The Impact of Labeling Effects on Perceived Quality of HD and UHD Video Streaming. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[101] Peter A Kara, Werner Robitza, Nikolett Pinter, Maria G Martini, Alexander Raake, and Aniko Simon. Comparison of HD and UHD video quality with and without the influence of the labeling effect. *Quality and User Experience*, 4:1–29, 2019.

[102] Peter A Kara, Aron Cserkaszky, Maria G Martini, Laszlo Bokor, and Aniko Simon. The effect of labeling on the perceived quality of HDR video transmission. *Cognition, Technology & Work*, 22(3):585–601, 2020.

[103] Fanni Geyer, Vince A Szakal, Peter A Kara, and Aniko Simon. Cognitive-bias-induced differences in the perceived video quality of rugged and conventional smartphones. In *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 592–599. IEEE, 2022.

[104] Stéphane Bouchard, Stéphanie Dumoulin, Jeanne Talbot, André-Anne Ledoux, Jennifer Phillips, Johana Monthuy-Blanc, Geneviève Labonté-Chartrand, Geneviève Robillard, Matteo Cantamesse, and Patrice Renaud. Manipulating subjective realism and its impact on presence: Preliminary results on feasibility and neuroanatomical correlates. *Interacting with Computers*, 24(4):227–236, 2012.

[105] Unity Real-Time Development Platform.
`https://unity.com/`.
[Online; accessed 22-05-2025].

[106] Stuart J McKelvie. Graphic rating scales—How many categories? *British Journal of psychology*, 69(2):185–202, 1978.

[107] Kathleen Myers and Nancy C Winters. Ten-year review of rating scales. I: overview of scale functioning, psychometric properties, and selection. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41(2):114–122, 2002.

[108] Hans-Jürgen Möller. Standardised rating scales in psychiatry: methodological basis, their possibilities and limitations and descriptions of important rating scales. *The World Journal of Biological Psychiatry*, 10(1):6–26, 2009.

[109] 2024 Essential Facts About the U.S. Video Game Industry.
`https://www.theesa.com/resources/essential-facts-about-the-us-video-game-industry/2024-data/`.
[Online; accessed 22-05-2025].

[110] Video Games Europe: Key Facts Report 2023.
`https://www.videogameseurope.eu/wp-content/uploads/2024/09/Video-Games-Europe-2023-Key-Facts-Report_FINAL.pdf`.
[Online; accessed 22-05-2025].

[111] Tim Wulf, Diana Rieger, Anna S Kümpel, and Leonard Reinecke. Harder, better, faster, stronger? The relationship between cognitive task demands in video games and recovery experiences. *Media and Communication*, 7(4):166–175, 2019.

[112] Adam M Large, Benoit Bediou, Sezen Cekic, Yuval Hart, Daphne Bavelier, and Shawn C Green. Cognitive and behavioral correlates of achievement in a complex multi-player video game. *Media and Communication*, 7(4):198–212, 2019.

[113] Andrew JA Seyderhelm and Karen L Blackmore. How hard is it really? Assessing game-task difficulty through real-time measures of performance and cognitive load. *Simulation & Gaming*, 54(3):294–321, 2023.

[114] Hugo Mitre-Hernandez, Roberto C Carrillo, Carlos Lara-Alvarez, et al. Pupillary responses for cognitive load measurement to classify difficulty levels in an educational video game: Empirical study. *JMIR Serious Games*, 9(1):e21620, 2021.

[115] Xiaofang Zhong and Jinjie Xu. Measuring the effect of game updates on player engagement: A cue from DOTA2. *Entertainment Computing*, 43:100506, 2022.

[116] Kaiyue Liu and Samaneh Samiee. Too much Patching? Protocological Control and Esports Player Counts. 2025.

[117] Mark Claypool, Artian Kica, Andrew L Manna, Lindsay O'Donnell, and Tom Paolillo. On the impact of software patching on gameplay for the league of legends computer game. *The Computer Games Journal*, 6:33–61, 2017.

[118] Kenneth Anderson. Software patches and their impacts on online gaming communities, 2019.

[119] Del Gallo, R. The Politics of a Game Patch: Patch Note Documents and the Patching Processes in League of Legends.
`https://utd-ir.tdl.org/server/api/core/bitstreams/10cc834a-ff20-4187-86f0-f3ef71a3c196/content`, 2023.
[Online; accessed 22-05-2025].

[120] Ashish Arora, Ramayya Krishnan, Rahul Telang, and Yubao Yang. An Empirical Analysis of Software Vendors' Patch Release Behavior: Impact of Vulnerability Disclosure. *Information Systems Research*, 21(1):115–132, 2010.

[121] Zhiqiang Lin, Xuxian Jiang, Dongyan Xu, Bing Mao, and Li Xie. AutoPaG: Towards Automated Software Patch Generation with Source Code Root Cause Identification and Repair. In *Proceedings of the 2nd ACM symposium on Information, computer and communications security*, pages 329–340, 2007.

[122] Steven D Webb and Sieteng Soh. Cheating in networked computer games: a review. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, pages 105–112, 2007.

[123] Jussila, A. Conceptualizing video game updates.
`https://jyx.jyu.fi/bitstreams/b64f85e3-94c3-4544-99a8-e41c8c8f8991/download`, 2022.
[Online; accessed 22-05-2025].

[124] Andrew Truelove, Eduardo SD Almeida, and Iftekhar Ahmed. We'll fix it in post: What do bug fixes in video game update notes tell us? In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 736–747. IEEE, 2021.

[125] Jacob Mertens. Broken Games and the Perpetual Update Culture: Revising Failure With Ubisoft's Assassin's Creed Unity. *Games and Culture*, 17(1):70–88, 2022.

[126] Peter A Kara, Laszlo Bokor, Andreas Sackl, and Mariana Mourão. What your phone makes you see: Investigation of the effect of end-user devices on the assessment of perceived multimedia quality. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2015.

[127] Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, 61(3):520–531, 2015.

[128] Ruigang Fang, Dapeng Wu, and Liquan Shen. Evaluation of image quality of experience in consideration of viewing distance. In *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 653–657. IEEE, 2015.

[129] Hadi Amirpour, Raimund Schatz, Christian Timmerer, and Mohammad Ghanbari. On the impact of viewing distance on perceived video quality. In *2021 international conference on visual communications and image processing (VCIP)*, pages 1–5. IEEE, 2021.

[130] László Harmat, Örjan D Manzano, Töres Theorell, Lennart Högman, Håkan Fischer, and Fredrik Ullén. Physiological correlates of the flow experience during computer game playing. *International Journal of Psychophysiology*, 97(1):1–7, 2015.

[131] Mina Yeo, Seokbeen Lim, and Gilwon Yoon. Analysis of biosignals during immersion in computer games. *Journal of Medical Systems*, 42:1–9, 2018.

[132] Steven Schmidt, Stefan Uhrig, and Domenic Reuschel. Investigating the relationship of mental immersion and physiological measures during cloud gaming. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2020.

[133] Brenda K Wiederhold, Renee Davis, and Mark D Wiederhold. The effects of immersiveness on physiology. In *Virtual environments in clinical psychology and neuroscience*, pages 52–60. IOS Press, 1998.

[134] Brenda K Wiederhold, Dong P Jang, Mayumi Kaneda, Irene Cabral, Yair Lurie, Todd May, In Y Kim, Mark D Wiederhold, and Sun I Kim. An investigation into physiological responses in virtual environments: an objective measurement of presence. *Towards cyberpsychology: Mind, cognitions and society in the internet age*, 2:175–183, 2001.

[135] Andreas Halbig and Marc E Latoschik. A systematic review of physiological measurements, factors, methods, and applications in virtual reality. *Frontiers in Virtual Reality*, 2:694567, 2021.