

Article

Not peer-reviewed version

Enhancing Player Experience in an FPS with Dynamic Audio Cue Adjustment Based on Gaussian Progress Regression

[Xiaoxu Li](#) , [Yi Xia](#) , [Mustafa Can Gursesli](#) , [Xiao You](#) , [Siyuan Chen](#) , [Ruck Thawonmas](#) *

Posted Date: 1 October 2024

doi: 10.20944/preprints202409.2460.v1

Keywords: First-person shooter; Gaussian progress regression; Dynamic difficulty adjustment; Audio cue



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Enhancing Player Experience in an FPS with Dynamic Audio Cue Adjustment Based on Gaussian Progress Regression

Xiaoxu Li ¹, Yi Xia ¹, Mustafa Can Gursesli ², Xiao You ¹, Siyuan Chen ¹
and Ruck Thawonmas ^{3,*}

¹ Graduate School of Information Science and Engineering, Ritsumeikan University, 2-150 Iwakura-cho, Ibaraki, 567-8570, Osaka, Japan; gr0557hs@ed.ritsumei.ac.jp(X.L.); gr0666ih@ed.ritsumei.ac.jp(Y.X.); gr0609ie@er.ritsumei.ac.jp(X.Y.); gr0634hi@ed.ritsumei.ac.jp(S.C.)

² Department of Information Engineering, University of Florence, 50139 Florence, Italy; mustafacan.gursesli@unifi.it

³ College of Information Science and Engineering, Ritsumeikan University, 2-150 Iwakura-cho, Ibaraki, 567-8570, Osaka, Japan; ruck@is.ritsumei.ac.jp

* Correspondence: ruck@is.ritsumei.ac.jp

Abstract: This paper analyzes the experience of first-person shooter (FPS) players when the game difficulty is adjusted by personalizing their audio cue settings, considering the balance between player performance, modeled using Gaussian Process Regression (GPR), and prior data serving as designer preference. In addition, we investigate the reason why player experience changes according to in-game audio cues. Previous studies have proposed various dynamic difficulty adjustment (DDA) methods for FPS games. However, few studies have considered the role of audio cues in the player experience. This paper compares the player experience of personalized enemy audio cue volume setting (GPR-DDA) and that of predetermined setting in an FPS game. Two comprehensive experimental phases, involving 80 participants, are conducted to assess the efficacy of GPR-DDA. The experience of our players is measured using questions taken from the Game User Experience Satisfaction Scale (GUESS) questionnaire and a final survey asking their open-ended feedback. A large language model (LLM) is used to analyze the natural language expressions of the players according to their native languages. To effectively have an LLM assist a limited number of qualified human evaluators in the classification of player responses, we develop an original procedure for this task. The GUESS results show that GPR-DDA can improve the player experience, and, in addition, the high consistency in the classification results over multiple runs of the selected LLM, as well as the similarity between its results and those of our human evaluators, reflect the reliability of the proposed LLM-assisted procedure.

Keywords: first-person shooter; Gaussian progress regression; dynamic difficulty adjustment; audio cue

1. Introduction

Immersion and realism are crucial elements for games that aim to provide players with enjoyment [1]. The achievement of immersion and realism requires the collaboration of various in-game components, one of which is the integral audio system. Within the sound system of a game, in-game audio cues play a fundamental and influential role. Audio cues are sounds that provide information about the game state, such as the location, distance, and movement of enemies, allies, and objects [2]. Previous research has shown that they are effective in enhancing player abilities [2,3] and contribute to a heightened sense of immersion during gameplay [4,5]. This is particularly noticeable in first-person shooter (FPS) games, where audio cues are instrumental in helping players accurately identify potential threats [2].

Audio perception in games is a controversial topic. In 2018, a study on audio perception in virtual reality (VR) environments stated that the importance of user perception of audio in VR environments is not as important as visual perception [6]. In contrast, some more recent studies have shown that audio has very much importance in VR environments with reasonable design. For example, Eames [7] pointed out that the design of audio in VR environments can deliberately guide the player's attention to

achieve a better narrative effect. Lucas et al. [8] designed a VR audio game with a free-motion interface specifically for mobile phones, and their participants generally found the audio design of this game to be an enjoyable and challenging activity. As well, another study showed that designing or listening to music in the right environments can give users a better sense of immersion [9]. Considering to these results we can see that audio design can have different effects even in the same environments, with some presenting a positive effect of audio design, and others showing that audio is not particularly helpful.

One can see that in the aforementioned studies where audio perception was considered important, researchers tended to deliberately design the audio that corresponds to a certain purpose in order to achieve it. Therefore, a well-designed audio system can give or guide the player in some way to achieve a certain set of reasonable goals. In our previous study [10], we balanced player performance modeled by Gaussian Process Regression (GPR) and our prior data, obtained from test players and serving as designer preference, as our goal for setting the game difficulty.

In addition, it is worth noting that most of the studies on audio perception have pointed out that a reasonable design of audio for different purposes can have a certain effect on the user, sometimes positively, sometimes negatively. In a scoping review of audio in recent years [11], it was clearly stated that the experience given to players by audio in recent studies is not necessarily positive. This also means that in audio design, when we design an audio that achieves our goals, it cannot be taken for granted that the audio will give the user a positive experience.

For our previous research [10], we successfully used GPR as the player modeling tool together with the proposed goal recommendation algorithm to control the game difficulty through audio cue volume setting. However, the question of whether such an audio cue design gives the player a positive gaming experience has not been addressed. In the game domain, if the player's game experience is subpar, we cannot say that this approach is justified when applied to games in practice, though our method is successful in controlling the difficulty. In addition, a previous study mentioned that Dynamic Difficulty Adjustment (DDA) has high expectations but mixed results in enhancing the player's experience [12], while Almac also pointed out, "Different genres create different player experiences in terms of audiovisual aesthetics." [13] Therefore, the study of audio perception in VR games, as well as audio games or other game genres, may not be applicable to FPS games, so it is important to study the player's gaming experience with the algorithm proposed in our previous research to control the difficulty of the game.

In summary, this paper primarily addresses the following research questions (RQs):

RQ 1: *Can players have a positive gaming experience when a DDA mechanism is implemented through game volume?*

In this paper, we use questions taken from the Game User Experience Satisfaction Scale (GUESS) questionnaire [14] to investigate in detail whether players will have a more positive gaming experience with the DDA mechanism proposed in our previous study.

RQ 2: *How does the game volume affect the player's gaming experience?*

To better understand and analyze the player experience, we ask players to share with us their feelings and opinions in writing, using their native languages. We anticipate that these kinds of open-ended responses may consist of multiple languages, which means it may be difficult to find qualified human evaluators. Therefore, we use a state-of-the-art large language model (LLM) to help us on the language analysis task. Although LLMs have proven powerful in analyzing natural languages [15], they sometimes hallucinate [16]. We alleviate this issue by designing a procedure that enables an LLM to reliably help few human evaluators classify player responses.

Our study makes three notable contributions. Firstly, we investigate whether DDA through in-game audio cue volume setting—using the aforementioned goal recommendation algorithm and the player performance modeled by GPR—can give the player a positive gaming experience. This work reveals that in-game audio cue volume, if appropriately designed, can give the player a positive

gaming experience. Secondly, we identify what affects the player's gaming experience in the FPS game in use, equipped with the aforementioned DDA mechanism. Finally, the proposed procedure for having an LLM participate in the player-response classification task with only a few human evaluators is effective and can be used as a reference for other applications.

2. Related Work

In this section, we examine recent relevant research on audio cues and FPS games. We also discuss recent work on DDA and its impact on player experience. Finally, we explore research on LLMs, specifically the use of ChatGPT in games. The review of these studies will serve as theoretical support for this paper.

2.1. Audio Cues

When delving into the realm of audio cues, it is crucial to acknowledge their fundamental role within the in-game audio system. The in-game audio system works in tandem with background music, ambience, and sound effects [5,17]. Previous studies have supported the idea that in-game audio cues can enhance the player's gaming skills [4,8,18] and game immersion [7,19]. Some research has emphasized the impact of in-game audio on players' affective states [3,11,20], triggering positive or negative changes in their performance and experience [11,21]. In addition, our previous work found a correlation between the volume of enemy audio cues and player performance [10]. These findings highlight the potential of audio cues to influence player performance and affective states. However, there is a lack of studies analyzing the player experience in an FPS game with a DDA mechanism using audio cues. Our work addresses this research gap.

2.2. FPS Game

FPS games have garnered significant scholarly attention, with many studies dedicated to improving the gaming experience through the use of various gaming devices, such as game controllers [22,23] and virtual reality [22–24]. In a notable study by Holm et al. [25], a correlation was established between players' physiological data, such as heart rate and electrodermal activity, and their experience in FPS games. These investigations highlight the current interest in seeking ways to enhance players' gaming experiences.

Previous research focused on the use of audio cues to improve the player's FPS gaming abilities [2]. Furthermore, in our previous work, we endeavored to develop a DDA algorithm to enhance FPS gaming performance [10]. These studies demonstrated the roles of audio cues to FPS gaming performance and gaming skills. However, both studies have primarily concentrated on technical aspects while not considering the significance of user experience.

2.3. DDA

DDA is a pivotal game mechanism that has gained attention due to its potential to positively impact various facets of the gaming experience. Previous research has shown that DDA can influence player retention [26], confidence [27], overall experience [28–30], and gaming ability [31]. By adaptively modulating the difficulty level, DDA facilitates the provision of a tailored experience that caters to diverse player groups, aligning with their individual preferences. These research findings underscore the importance of DDA in game design.

From a different perspective, DDA proves beneficial for players in various ways. Novice players often prefer easier levels, as this fosters a smooth learning curve, whereas expert players tend to seek more challenging gameplay [31]. It is important to note that players may sometimes underestimate the intended difficulty level set by game designers [32]. Therefore, it is necessary to integrate diverse goals into DDA algorithms, accounting for distinct player characteristics. To address this issue, our previous work [10] proposed a new "goal" takes into account both designer preference and player performance,

resulting in a more personalized and effective DDA. However, our previous study did not investigate players' gaming experiences.

2.4. LLMs

LLMs have had a wide range of applications in recent years, with strong capabilities in solving various natural language processing tasks including question answering tasks, reading comprehension tasks, translation tasks as well as language modeling, cloze, and completion tasks [33]. Recent studies have also shown that using the right prompt, or the right prompt engineering technique, LLMs can help researchers achieve many goals [34]. ChatGPT, in particular, showed its potential [35] and has become a popular research target. For example, recent studies have used ChatGPT to categorize audience comments for a live game [36], assist game design [37], or generate game commentaries [38].

Although these studies exemplify that using ChatGPT to analyze open-ended player responses is effective, one should be aware of the possibility of hallucination in ChatGPT. In this paper, we provide a procedure for using ChatGPT in collaboration with only a few human experts to reliably analyze open-ended player responses. The provided procedure should give other researchers insights into how to use ChatGPT and the like reliably.

3. Preliminary Study

When designing a level, designers will decide whether the current design is on the difficult or easy side when difficulty setting is needed. However, it is important to understand that the player population has very discrete gaming skills. This means that even if the game designer decides that a certain value adjustment is easy, there are still players who think that it is difficult, and vice versa. Hence, balancing both sides—designer preference and player performance—is challenging.

In our previous research [10], GPR models players at each game level by mapping their game performance and audio cue volume. By choosing the enemy audio cue volume setting based on the goal recommendation algorithm (Eqn. 1) in each round, DDA is realized, making our players' performance close to their recommended goals. In the following, this DDA mechanism will be referred to as GPR-DDA.

$$Goal = w_3(w_1C_{Best} + (1 - w_1)C_{Worst}) + (1 - w_3)(w_2T_{Best} + (1 - w_2)T_{Worst}) \quad (1)$$

In this equation, C represents the respective performance of the current player, and T represents the respective performance of prior data derived beforehand for each game level objectively from test players or subjectively from game designers; the subscripts *Best* and *Worst* respectively stand for their best and worst performance among all levels, based on both the actual performance at each seen level and the predicted performance by GPR for each unseen level; the subscripts w_1 , w_2 , and w_3 are coefficients whose values range from 0 to 1. In this study, the term “goal” refers to the degree of challenge in a game. This algorithm computes the goal by considering the performance of both the player and prior data to avoid the game being overly easy, which can lead to player boredom, or excessively challenging, which can lead to player frustration [26].

Previous studies have used many machine learning algorithms for modeling players in DDA mechanisms, such as random forest [39], Monte Carlo tree search [40], and deep neural networks [41]. However, considering the response time in games, GPR, as a powerful and flexible non-parametric regression technique, has been utilized in recent studies. For example, due to being fast and accurate, GPR has been used to adapt game content [42] quickly and design games such as Flappy Bird and Spring Ninja that maximize user engagement [43]. Therefore, GPR was selected to model players in our previous study.

In summary, GPR-DDA bridges the gap in audio cue volume setting between the designer preference and player performance and achieves this with a relatively fast response time.

However, as a mechanism applied in an FPS game, we need to understand whether GPR-DDA is able to provide a positive overall gaming experience to the players. Therefore, analyzing the players'

experience of the game is important for our previously proposed method. In this paper, we follow the game design of our previous study, with a reasonable modification of the game map mentioned in Section 4.2. In addition, we set up two phases in the experiment, which is consistent with our previous study, to test whether GPR-DDA can provide a positive gaming experience.

4. Methodology

We divide our experiment into two distinct phases, each with its own unique group of players. In the first phase, players will be randomly divided into two groups; one group will play multiple levels with audio cues of increasing volume, and the other group vice versa. In contrast, in the second phase, players play multiple levels, at each of which GPR-DDA performs the audio cue setting. Our decision to have two unique groups of players for the two phases is based on the following reasons:

- First, we want to prevent players from having psychological expectations before the game. Players with experience in the first phase may correctly guess the objective of our experiment. They could adopt a learned playing strategy if asked to participate in the second phase.
- In addition, we need a set of prior data to compute the initial recommended level in the second phase. For this purpose, we use the players’ performance from the first phase as our prior data. As stated in the first reason, it would be inappropriate for our experiment to have the same players participate in both phases.

During each round, we record player data at one-second intervals, each consisting of their in-game coordinates, orientation, remaining health, and ammunition count. These data are essential for facilitating subsequent data processing and analysis, enabling us to replay and analyze each player’s trajectory. Figure 1 provides an overall workflow of our experiment.

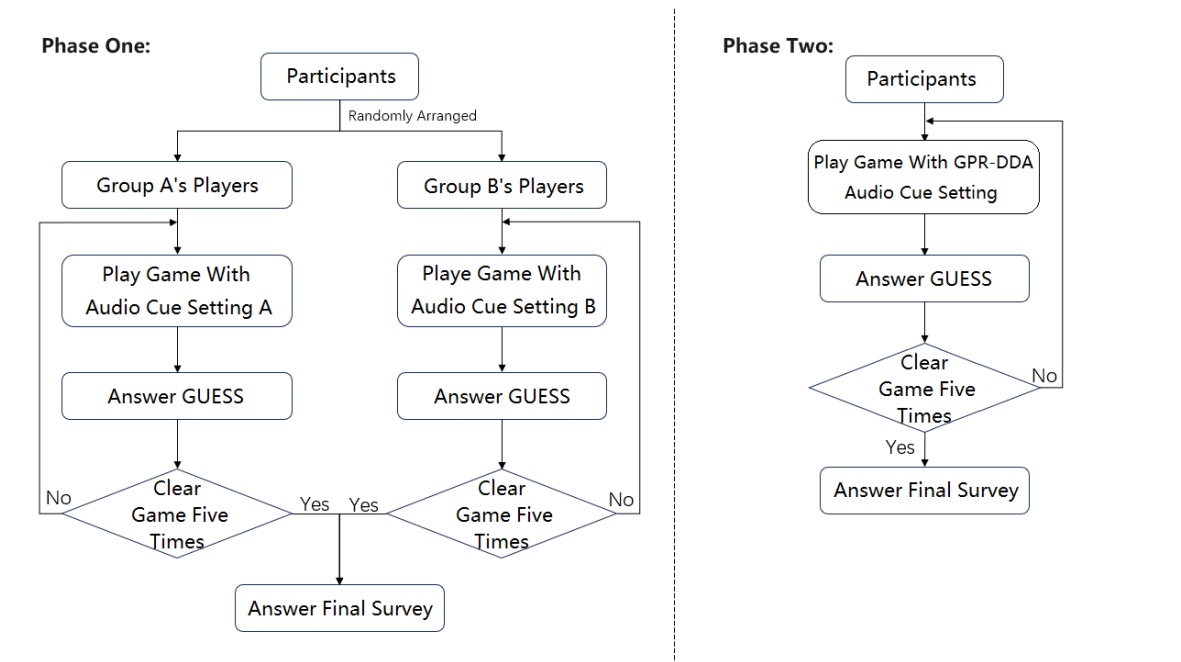


Figure 1. Experimental workflow.

Findings from the field of sports medicine and exercise science [44] suggested that an excessively prolonged experimental procedure might lead to player fatigue, potentially resulting in haphazard or arbitrary responses to a given questionnaire. Therefore, we limit the duration of our experiment to less than an hour per participant. One round consists of playing a game level and answering a questionnaire and takes about 8 to 10 minutes. Hence, it is reasonable for each player to participate in five rounds (about 40 to 50 minutes) and a final survey asking for general information (about 10 minutes).

4.1. Procedure

4.1.1. Phase One

In phase one, which serves as our baseline, players participated in an FPS game with a difficulty curve that we set manually. We establish two distinct sets of manually implemented difficulty. In the first set, called Group A, the enemy volume progresses from low to high, ranging from 50% to 100%, across a total of five levels, with an equal increment, i.e., 0.5, 0.625, 0.75, 0.875, and 1. Conversely, the second set, called Group B, involves a reverse progression in which the enemy volume decreases, i.e., 0.5, 0.375, 0.25, 0.125, and 0. The primary rationale for having two groups is based on the following considerations:

- Firstly, due to our restriction on players participating in only five rounds of gameplay, it is deemed impractical to have a linear progression from 0% to 100% across five levels with a uniform increment. Such a dramatic variation in enemy volume between two consecutive levels would likely hinder the understanding of our collected data.
- Secondly, dividing into two groups allows us to examine player experiences when encountering scenarios where the enemy volume increases from low to high and the opposite. Although this is not the primary focus of our experiment, it has the potential to uncover discrepancy in player experiences under different volume trends. Therefore, this may yield additional insights, enhancing the reliability of the audio cue volume design.

During phase one, players will be randomly assigned to one of the groups and will not be told which group they will be assigned to in advance. Players are expected to discern changes in audio-cue volume on their own. This approach is implemented to avoid creating any psychological expectations in players prior to gameplay, as we consider that such expectations could compromise the reliability of the experimental data.

4.1.2. Phase Two

In phase two, we use GPR-DDA to control the in-game audio cue volume. In our previous research, the goal recommendation algorithm was introduced to derive the player's target performance (Eqn. 1) for the DDA mechanism, where w_1 , w_2 , and w_3 were set to 0.5 [10]. In this work, we follow this recipe.

GPR-DDA recommends to the player a level with an audio cue volume setting at which the player's actual or predicted performance is closest to the goal [10]; however, if the player has already played such a level, a random level will be chosen [10]. This mechanism not only ensures that the player will not encounter the same level repeatedly but also expedites the discovery of the global optimum of audio cue volume setting rather than local ones.

4.2. FPS Games

Our experimental game is an FPS that requires a player to eliminate enemies on the map within a specific ammunition limit or locate the escape point, both while having limited health. This game simulates a scenario in which the player must escape an area under heavy siege. They can choose to either eliminate all enemies to complete the level or find the escape point, regardless of the number of enemies they have eliminated. In this game, the player needs to listen attentively to the enemies' audio cues to determine their positions for evasion or elimination. Figure 2 illustrates a typical scene of gameplay in our game.

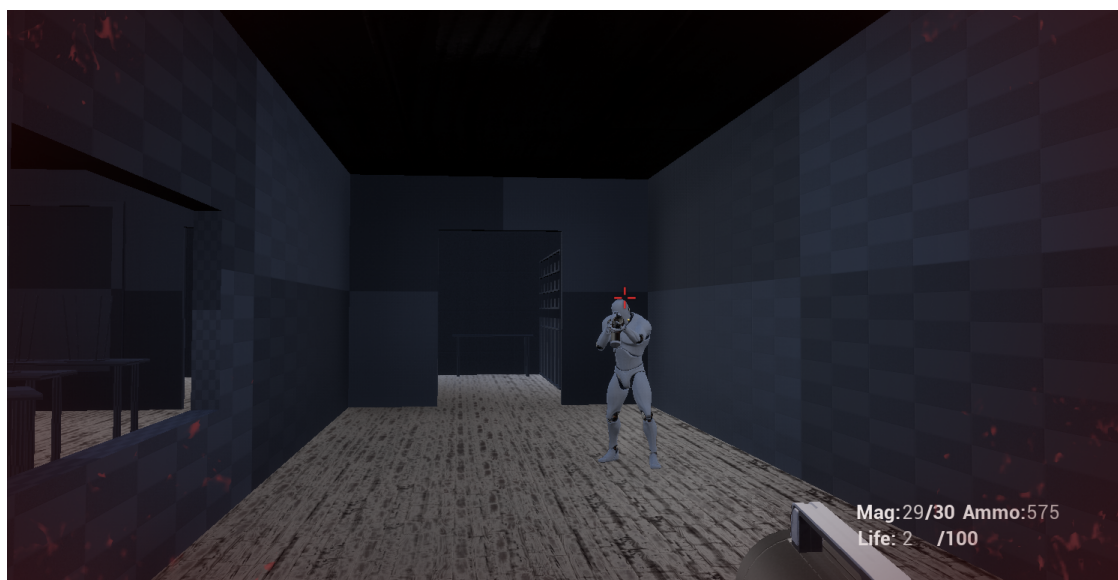


Figure 2. Game Scene.

In this study, we have retained the FPS game mechanisms and the opponent AI algorithm used in our previous work [10] (as detailed in Appendix-A), while introducing modifications to the game's map design. The original game featured five small maps, each measuring 500 units by 500 units in size, in Unreal Engine. These maps were characterized by their simplicity, with a single linear path for progression and the absence of multi-floor design. We consider that such a map design is ill-suited for contemporary FPS games. In our previous study [10], the small map size should restrict player exploration and lead to a lack of engagement, potentially impacting the accuracy of the player experience assessment.

To address these concerns, we have developed a new game map with dimensions of 8000 units by 8000 units in Unreal Engine. This map is composed of two floors (Figure 3), offering a significant increase in size, being 512 times larger than each map in the previous work. Furthermore, to align the map design more closely with commercial FPS games, we have incorporated principles and concepts reminiscent of those found in commercially available games like Rainbow Six Siege¹.

It is essential to note that in the design of our previous work [10], those five maps were crafted to prevent players from easily memorizing map layouts, thereby safeguarding the accuracy of the experiment. We endorse this perspective here. Consequently, in our newly designed larger map, we have incorporated eight entirely distinct and randomized spawn points. In each game round, the player is randomly assigned to one of these spawn points. We are aware that random allocation might occasionally result in having the player respawned at the same point. However, this approach aligns closely with the logic found in commercial games and complements our map design philosophy, resembling key elements of commercially available games.

¹ Tom Clancy's Rainbow Six Siege, Ubisoft, 2015

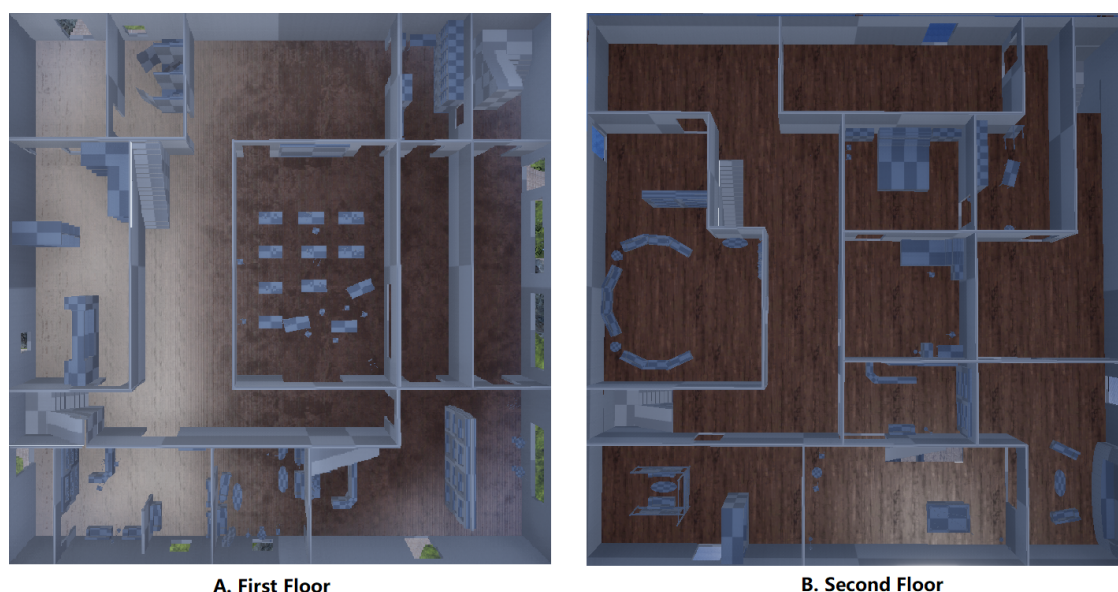


Figure 3. Game Map Design.

4.3. Participant Recruitment

In this paper, we target the predominant gaming demographic interested in FPS games. To ensure a sufficient number of respondents for both phases, we recruited students from two different classes at our university as the main participants for our experiment. Furthermore, in order to increase the diversity of the respondents, we also conducted phase one by using Amazon Mechanical Turk (AMT). In addition, we randomly distributed the experiment to volunteer members of the general public in phase two, ensuring that these participants had not taken part in our phase one on AMT. Participation in our experiment was voluntary, except for those from AMT who received a payment of 0.01 dollar, in line with the minimum payment policy of the AMT platform. All participants in the experiment gave their informed consent beforehand. In addition, all participants were different from our previous study.

4.4. Questionnaire

At the end of each round, our participants were required to complete a modified GUESS questionnaire in their phase. This questionnaire is crucial for capturing the gaming experience aspects of each round for every player, which allows us to construct individual experience curves over the course of five rounds of gameplay. Furthermore, it will enable us to compute the average gaming experience value within specific difficulty intervals, defined as audio cue volume ranges.

In Table 1, we present our GUESS questionnaire, which employs a seven-point Likert scale scoring system (1 = Strongly Disagree, 7 = Strong Agree). It comprises 11 questions, including three questions for detecting whether the participants are providing fraudulent responses. These three questions incorporate two fraudulent respondent detection mechanisms and are not used in our gaming experience analysis.

- The first question in the fraudulent respondent detection category in Table 1 uses a consistency check. This means that each participant must consistently answer this question in all five rounds of the game. Inconsistent responses to this question in the five questionnaires make all the responses provided by the participant unreliable, and their data will be discarded.
- The second and third questions in the fraudulent respondent detection category in Table 1 use a logical inconsistency check. We obtain these two questions by negating two selected questions from the eight game experience questions. For each pair of original and negated questions, we consider their responses reliable when their sum ranges between six and ten and those with a

sum value outside of this range unreliable. If a participant gives unreliable responses for at least one such pair in a given round, the participant’s response data for that round will be discarded.

Table 1. Modified GUESS Questionnaire.

| Factor | Item |
|---|--|
| Usability / Playability (Cronbach’s $\alpha = 0.84$) | I think it is easy to learn how to play the game. |
| | I find the controls of the game to be straightforward. |
| Enjoyment (Cronbach’s $\alpha = 0.81$) | I think the game is fun. |
| | I enjoy playing the game. |
| Play Engrossment (Cronbach’s $\alpha = 0.84$) | I feel detached from the outside world while playing the game. |
| | I do not pay attention to events happening in the real world during the game. |
| Audio Aesthetics (Cronbach’s $\alpha = 0.88$) | I enjoy the sound effects in the game. |
| | I feel the game’s audio (e.g., sound effects, music) enhances my gaming experience. |
| Fraudulent respondent detection mechanisms | If you are female, please click button 2; if you are male, please click button 3; if you do not fall into either of the above categories, please click button 4. |
| | I pay attention to events happening in the real world during the game. |
| | I do not think the game is fun. |

It is worth noting that the order of the 11 questions in our GUESS questionnaire is randomized. This randomization aims to enhance the assessment of potential fraudulent responses from players. Following this data cleaning process, we obtain valid GUESS questionnaire responses for subsequent statistical testing.

Additionally, our participants are required to complete a final survey at the end. Each will receive a unique ID at the end of the game. This ID will be used in the final survey to identify fraudulent respondents who, in this context, have not completed the game but the survey. One purpose of the final survey is to gather some general information from them, including their age range, gender, and proficiency in playing FPS games. General information will facilitate our data analysis by allowing us to investigate potential biases related to age, gender, or skill in playing FPS games, ensuring that we can discern whether certain findings are influenced by these factors. Concurrently, we also request players to provide feedback on their experiences while playing our game and any suggestions through the final survey. This feedback is intended to reveal additional aspects that may not have been adequately captured by GUESS regarding issues arising from changes in the volume of the audio cues of the game enemies.

5. Experiment

5.1. Participants and Their Responses

Players agreeing to participate in the experiment will be given anonymous access to our experimental game. Their performance data will be uploaded to our database as they play, and at the end of each round, they will be required to complete our GUESS questionnaire built into the game, the answers of which will also be uploaded to our database. Internet access is required to play the game.

Our experiment consists of two phases. In both phases, the participants are entirely different. A total of 80 people participated in the experiment as a player, with 40 in each phase. However, some players completed part of the game, and some did not fill out the final survey. After counting, with 40 players in phase one, we can identify 17 players with 70 GUESS responses from Group A, 16 players with 72 GUESS responses from Group B, and seven players who only completed their first assigned level, with an enemy volume of 0.5, and our GUESS questionnaire at the end of the level. For phase two, all 40 players are joined and we collected 136 GUESS responses. Regarding the final survey, we received 38 responses in phase one and 34 responses in phase two. In addition, the final survey revealed that our participants are from various backgrounds, i.e., 48 university students, 22

non-students, and two users from AMT (62.5% male, 23.75% female, 1.25% others, 12.5% prefer not to say or unknown; age ranged from 18 to 59); they come from six different countries: America, China, Italy, Japan, Turkey, and Venezuela (listed in alphabetical order).

Table 2 presents the responses and provides detailed insights into the distribution characteristics of the players in both phases. Based on the general information provided by the final survey, the *p-value* of the Shapiro-Wilk test, (Phase One: 5.201×10^{-6} ; Phase Two: 8.693×10^{-6}) shows that the player distributions in both phases with respect to the FPS ability do not follow a normal distribution. We used 0 (unfamiliar), 1 (somewhat familiar), and 2 (very familiar) as the scale to represent their FPS ability. As the respondents from the two phases involve different groups of people, we used the Mann-Whitney U test to assess the significance in difference between the respondents from the two phases. The *p-value* of the Mann-Whitney U test is 0.67, i.e., not statistically significant. This situation demonstrates that, despite consisting of different individuals, their abilities are similar. Therefore, comparing the GUESS values between phase one and phase two is dependable.

Table 2. Players’ general information for both phases.

| | FPS ability | |
|---|--|----------------------|
| | Phase one | Phase two |
| Mean value | 1.18 | 1.24 |
| Standard deviation | 0.68 | 0.73 |
| Shapiro-Wilk Test (<i>p-value</i>) ¹ | 5.2×10^{-6} | 8.7×10^{-6} |
| | Statistical significance test of FPS ability | |
| | Phase one | Phase two |
| Mann-Whitney U test (<i>p-value</i>) ¹ | 0.67 | |

¹ We use two-tailed testing, $\alpha = 0.05$.

In addition, in phase one, 94.7% are aged 18 to 29, 2.6% are aged 30 to 39, and 2.6% are over 40. In phase two, 90.9% are aged 18 to 29, 9.1% are aged 30 to 39, and none are over 40. Looking at the FPS ability and the age distribution of the players in both phases, we can also assume that there is no significant difference in the perception of audio in FPS games between players in both two phases.

After the data cleaning process, using the fraudulent respondent detection mechanisms for our GUESS questionnaire described in 4.4, we were left with valid data. In phase one, we received 100 out of 142 valid GUESS questionnaire responses, each a set of answers, from 30 players; and in phase two, we received 111 out of 136 such responses from 29 players. For the final survey, we identified one duplicate response from phase one and five fraudulent respondents from phase two, which we removed. Therefore, the final survey had 66 valid respondents (37 from phase one and 29 from phase two). Our data are accessible on the supplementary page.²

5.2. Metrics for Questionnaire Data

To examine whether the difference in the GUESS values for the players in phase one and phase two is statistically significant, we first calculate the average GUESS value—with the lowest score of 8 points and the highest of 56 points—for each player over all the rounds played by the player. After that, we use the Shapiro-Wilk test to check for a normal distribution of the data. If the GUESS values of players in both phases follow a normal distribution, we use the unpaired t-test; otherwise, we use the Mann-Whitney U test. These tests are used to determine which phase offers a better gaming experience for our participants.

In addition, we examine whether the GUESS value of each phase is statistically and significantly different from the GUESS mid-value of 32. Depending on the results of the Shapiro-Wilk test for each

² <https://github.com/Lxx007/Anonymous>

phase’s GUESS value, we use either the one-sample t-test or the Wilcoxon signed-rank test. These tests are used to examine whether the gaming experience of players in each phase significantly surpasses the mid-value of 32.

5.3. Metrics for Open-Ended Player Responses

Our final survey results consist of both English and Chinese languages due to the diverse origins of our participants, spanning six different countries, including non-English native speakers. As a result, we task three suitable human evaluators who are proficient in both English and Chinese with the analysis. With limited qualified human evaluators, we ask an LLM, ChatGPT, to help us with our classification task, described below, for open-ended player responses. In previous research [36], an endeavor was made to employ ChatGPT to classify bullet comment types in games, achieving success. This shows that ChatGPT can be used for classification tasks. In addition, ChatGPT, as a powerful LLM, has the ability on translation tasks and reading comprehension tasks, which can be used in multilingual semantic analysis for our classification task. Therefore, ChatGPT and human evaluators are employed to classify all player responses from the final survey. The objective of this analysis is to uncover the reasons behind the objective results.

However, the same previous research also mentioned that both human evaluators and ChatGPT will have inconsistent answers in some complex sentences. To mitigate this issue, as well as the effect of hallucinations present in ChatGPT, we have designed the following procedure:

- For ChatGPT, we use it with an originally crafted prompt for classification, running it three times for each classification instance. We consider a result of interest reliable if and only if it is consistent across all the three runs; otherwise, it is deemed unreliable and labeled as 0.
- For our human evaluators, we compare their responses. For each classification instance, if they all give the same answer, we consider the answer reliable; otherwise, it is deemed unreliable and labeled as 0.

Afterward, we combine the results of ChatGPT and the human evaluators as follows:

- If the result from the human evaluators is labeled as 0 but that from ChatGPT is not, we select ChatGPT’s result.
- If both results from ChatGPT and the human evaluators are labeled as 0, the result is categorized as 0.
- Otherwise, we select the result from the human evaluators. In this case where both results from ChatGPT and the human evaluators are not labeled as 0, we prioritise the response of the human evaluators because we consider them to be more reliable than LLMs.

To gain clarity on our players’ feelings about the audio cue design in each phase, we categorize all open-ended player responses from the final survey into three categories:

1. Not mentioned
2. The paragraph directly mentions audio cues in the game and complains about it
3. The paragraph directly mentions audio cues in the game and praises it

Prompt 1 shows the prompt. It is used by ChatGPT (via OpenAI API) and our human evaluators to analyze the open-ended player responses, individually shown under @paragraphs. This ensures that both ChatGPT and the human evaluators performed this classification task based on the same instruction.

Prompt 1: Prompt for analyzing open-ended responses from the participants

Please classify the following @paragraphs into three categories based on their content related to audio cues in the game. If the paragraph doesn’t clearly and directly express a positive or negative sentiment towards the game’s audio cues, please choose A.

If the paragraph clearly and directly expresses a negative sentiment towards the game’s audio cues, please choose B.

If the paragraph clearly and directly expresses a positive sentiment towards the game’s audio cues, please choose C.

If the text doesn’t mention audio cues, it means there are no complaints or compliments about audio cues. Please avoid overinterpreting.

A: Not mentioned

B: The paragraph directly mentions audio cues in the game and complains about it

C: The paragraph directly mentions audio cues in the game and praises it

Please directly tell me the classification; there is no need to explain the reasons within.

@paragraphs

It was fun at first. However, FPS games sometimes make me sick. So it’s no longer comfortable to play.

6. Results and Discussions

6.1. Results and Discussions of Questionnaires Data

The average GUESS value for phase one, which consists of 100 responses, is 32.85. For phase two, which has 111 responses, the average GUESS value is 38.58. The average GUESS value in phase two is higher than that in phase one, indicating that, on average, the players in phase two prefer the level recommendation mechanism. In addition to the average value, we statistically examined which phase can provide players with a superior gaming experience. Table 3 shows that the average GUESS values of the players in phase two follow a normal distribution, while those in phase one do not. Therefore, we used a Mann-Whitney U test to assess the significance in difference of the GUESS result between the two phases. The *p-value* of the Mann-Whitney U test shows a significant difference between phase one and phase two, with the latter outperforming the former.

Table 3. Comparison of results between phase one and phase two, as well as each phase and the midpoint.

| | GUESS | | |
|---|-------------------------------|-----------------------|-----------------------|
| | Phase one | Phase two | |
| Mean value | 32.85 | 38.58 | |
| Standard deviation | 6.84 | 6.59 | |
| Shapiro-Wilk test p -value ¹ | 3.49×10^{-3} | 0.69 | |
| | Statistical significance test | | |
| | Phase one vs Phase two | Phase one vs Midpoint | Phase two vs Midpoint |
| Mann-Whitney U test p -value ¹ | 1.24×10^{-3} | -- | -- |
| Wilcoxon signed-rank test p -value ¹ | -- | 0.9 | -- |
| One sample t-test p -value ¹ | -- | -- | 1.3×10^{-5} |

¹ We use two-tailed testing, $\alpha = 0.05$.

Moreover, we compared each group with the midpoint value of 32 for GUESS. The conducted Wilcoxon signed-rank test reveals that the GUESS value for the players in phase one does not significantly differ from the midpoint value. However, the average GUESS value for the players in phase two significantly exceeds the midpoint value of 32, when tested using a one-sample t-test.

In summary, the average GUESS value in phase two was higher than that in phase one, indicating that the players in phase two had a better gaming experience than those in phase one. In addition, it was significantly higher than the mid-value of 32. Consequently, these results indicate that players will have a better gaming experience in an FPS game with GPR-DDA.

Our results support the idea that GPR-DDA has the ability to give the player a better gaming experience than the linearly changing enemy audio cue volume. These results answer our **RQ 1:** (*Can players have a positive gaming experience when a DDA mechanism is implemented through game volume?*). Therefore, when considering the goals of a DDA mechanism, balancing both designer preference

and player performance is a better choice. As done in our previous research, three coefficients have been introduced to allow designers to adjust the weights of DDA on designer preference and player performance to fit their games. Our results support that players will have a positive gaming experience when w_1 , w_2 , and w_3 are set to 0.5. In addition, we posit that it will be difficult to generalize the positive gaming experience to all players if the coefficients are set to extremes, e.g. w_3 is set to 0, by which designer preference is only taken into account. In summary, setting all the coefficients to 0.5 is reasonable, but extreme weighting is not a good choice, while we highly recommend that the coefficients be adjusted accordingly for each game project.

6.2. Results and Discussions of Open-Ended Player Responses

In our procedure outlined in 5.3, a given instance will be labeled as A, B, or C when multiple human evaluators unanimously select a label of interest or when this does not happen but multiple rounds of ChatGPT consistently select it.

This procedure ensures that only player feedback with such a label is considered in the following analysis (Table 4), increasing the reliability of relevant findings. The consistency among the responses from our three human evaluators is 81.25%, while that among the three rounds of responses from ChatGPT is 90.28%. Furthermore, ChatGPT’s responses have shown a high similarity of 75% to those of its human counterpart. These results demonstrate the potential of ChatGPT in providing consistent and human-like answers. As a result, we recommend applying the procedure in 5.3 when using LLMs such as ChatGPT to help on classification tasks when human evaluators fail to reach a consensus.

Table 4. Feedback on audio cue design in both phases.

| | Complaint | | Praise | | Not Mentioned | | Label 0 | |
|--------------------------------|---------------------------------|----------------|----------------------------------|-----------|---------------|-----------|-----------|-----------|
| | Phase One | Phase Two | Phase One | Phase Two | Phase One | Phase Two | Phase One | Phase Two |
| Number of players | 9 | 7 | 0 | 1 | 25 | 29 | 3 | 2 |
| | Reasons for complaints on audio | | | | | | | |
| | Volume too high | Volume too low | Not realistic & Need improvement | | | | | |
| Number of players in phase one | 4 | 2 | 3 | | | | | |
| Number of players in phase two | 2 | -- | 5 | | | | | |

Table 4 shows the results of our analysis, which reveal interesting trends. During phase one, 9 out of 37 final survey respondents reported complaints, accounting for 24.32%. In phase two, 7 out of 29 final survey respondents complained, which accounted for 24.13%, slightly less than in phase one. However, during phase two, one player explicitly praised the in-game audio cue, which did not occur during phase one. The results indicate that in phase two, players had fewer complaints about the enemy audio cue design and expressed more positive feedback. We argue that this is a significant factor in the overall improved player experience, according to the results from the modified GUESS questionnaire, in phase two compared to phase one.

To answer our **RQ 2** (*How does the game volume affect the player’s gaming experience?*) in detail, we have taken the additional step of conducting a manual review of the primary reasons for player complaints. Out of the nine complaining players in phase one, two players found the audio cue volume too low, four players found it too high, and three players thought that the sound effects needed improvement. In contrast, out of the seven complaining players in phase two, five requested more realistic sound effects, particularly for footsteps, which is not the main focus of this study, while only two players felt that the audio cue volume was too high.

As a result, we argue that having more varied complaints in a DDA setting tends to indicate the DDA setting provides a worse gaming experience. Regarding feedback about unrealistic sound, only one player in phase one specifically mentioned poor footstep simulation, limiting further discussion. Finally, more complaints about loud audio cues in both phases imply that lower volumes might enhance the gaming experience in FPS games.

7. Conclusion

This paper explored whether GPR-DDA can enhance the gaming experience of players, and how audio cue volume effects it. A comparison of the GUESS values of the players in phases one and two revealed that GPR-DDA, used in phase two, significantly improves the player experience. Moreover, based on player feedback from the final survey, we identified why the GPR-DDA is superior to the baseline, used in phase one. Based on these findings, we draw the following conclusions.

Firstly, because the relationship between gaming experience and audio cue volume is highly personalized, it is challenging to manually adjust the game audio cue volume for each player to achieve not only the player's performance near a given goal but also a better gaming experience. However, our results suggest that choosing a lower volume of audio cues might give a better experience. More importantly, GPR-DDA can improve players' individual gaming experience in an FPS game.

Secondly, we extend findings from those in our previous work [10]. Our previous study suggested that a moderate audio cue volume (around 0.5) leads to a higher player performance. In the current study, our findings indicate that players prefer a lower audio cue volume, which results in a better gaming experience. Therefore, we infer that using audio cue volume to control game difficulty does not necessarily result in a better gaming experience for each individual, even if it improves their performance. As a result, for an FPS game, we recommend initializing the audio cue volume at a low value (less than 0.5) for players who prioritize gaming experience and around 0.5 for players who prioritize their performance. In addition, the coefficients in Eqn. 1 should be adjusted accordingly for GPR-DDA to improve both player performance and gaming experience.

Thirdly, our analysis of the players' feedback from the final survey suggests that the degree of variety in complaints can serve as a metric to evaluate player experience. Additionally, the proposed procedure in 5.3 allows language models such as ChatGPT to reliably assist humans in performing classification tasks for natural language analysis. Finally, since the game in use in this study aligns the design philosophy and game mechanisms with those of commercial 3D FPS games, our study has potential applications for such games on the market.

8. Limitation and Future Work

FPS games demand players to focus more on audio cues, which provide strategic information to gameplay rather than merely serving as ambiance or decoration. This study acknowledges its potential limitations beyond the FPS genre, such as role-playing, racing, or fighting games, where the significance of audio cues might be less pronounced. In addition, our game is a 3D game, the perception of audio in 3D and 2D games may be different. As a result, more detailed studies are required for other game genres and 2D games.

Furthermore, this study focuses solely on controlling the volume of audio cues. However, audio cues in gaming possess various characteristics, including reflection and attenuation. For research requiring a detailed examination of audio cue properties, this study lays a robust groundwork for implementing diverse alterations to audio cues.

Our future research will focus on extending this study to various game genres and 2D games to emphasize the ongoing exploration of audio cue significance. This should contribute to the advancement of audio cue studies, whether aimed at enhancing gaming enjoyment, educational significance, or aiding visually impaired individuals.

Author Contributions: Conceptualization, X.L. and R.T.; methodology, X.L., Y.X. and R.T.; software, X.L.; validation, Y.X.; formal analysis, X.L.; investigation, X.L.; resources, R.T.; data curation, X.L., Y.X., M.C.G., X.Y. and S.C.; writing—original draft preparation, X.L.; writing—review and editing, Y.X., M.C.G., S.C. and R.T.; visualization, X.L.; supervision, R.T.; project administration, X.L.; funding acquisition, R.T.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in GitHub at <https://github.com/Lxx007/Anonymous>.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|---|
| FPS | First-Person Shooter |
| GPR | Gaussian Process Regression |
| DDA | Dynamic Difficulty Adjustment |
| GUESS | Game User Experience Satisfaction Scale |
| LLM | Large Language Model |
| VR | Virtual Reality |
| RQ | Research Question |

Appendix A. Opponent AI Design

Below are eight rules to govern the AI logic of our enemy characters.

- They rely on their line of sight to detect the player's presence.
- The player's audio cue does not trigger their detection of the player, making the game more user-friendly for less experienced players.
- They exhibit roaming behavior within the game arena, ensuring a unique experience for each playthrough.
- When they spot the player or sustain damage from the player, they immediately initiate attacks.
- After engaging in combat, if the player moves out of their line of sight, they stop shooting and adopt a slower pace to covertly track the player, hiding their audio cues. This design element adds an intriguing dynamic to the game.
- If the player reenters their line of sight during the tracking phase, they launch an instant attack.
- If they fail to locate the player within 60 seconds, they will return to their roaming behavior.
- The volume of audio cues emitted by them follows the method described in this paper.

The implementation of these rules is expected to enhance the overall engagement and immersion of players in our FPS game.

References

1. Christou, G. The interplay between immersion and appeal in video games. *Computers in human behavior* **2014**, *32*, 92–100.
2. Johanson, C.; Mandryk, R.L. Scaffolding player location awareness through audio cues in first-person shooters. In Proceedings of the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 3450–3461.
3. Grimshaw, M.; Tan, S.L.; Lipscomb, S.D. Playing with sound: The role of music and sound effects in gaming. **2013**.
4. Ribeiro, G.; Rogers, K.; Altmeyer, M.; Terkildsen, T.; Nacke, L.E. Game atmosphere: effects of audiovisual thematic cohesion on player experience and psychophysiology. In Proceedings of the Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 2020, pp. 107–119.
5. Andersen, F.; King, C.L.; Gunawan, A.A.; et al. Audio influence on game atmosphere during various game events. *Procedia Computer Science* **2021**, *179*, 222–231.
6. Rogers, K.; Ribeiro, G.; Wehbe, R.R.; Weber, M.; Nacke, L.E. Vanishing importance: studying immersive effects of game audio perception on player experiences in virtual reality. In Proceedings of the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–13.
7. Eames, A. Beyond Reality. In Proceedings of the Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry, New York, NY, USA, 2019; VRCAI '19. <https://doi.org/10.1145/3359997.3365736>.

8. Gilberto, L.G.; Bermejo, F.; Tommasini, F.C.; García Bauza, C. Virtual Reality Audio Game for Entertainment & Sound Localization Training. *ACM Transactions on Applied Perception* **2023**.
9. Turchet, L.; Carraro, M.; Tomasetti, M. FreesoundVR: soundscape composition in virtual reality using online sound repositories. *Virtual Reality* **2023**, *27*, 903–915.
10. Li, X.; Wira, M.; Thawonmas, R. Toward Dynamic Difficulty Adjustment with Audio Cues by Gaussian Process Regression in a First-Person Shooter. In Proceedings of the International Conference on Entertainment Computing. Springer, 2022, pp. 154–161.
11. Bosman, I.d.V.; Buruk, O.O.; Jørgensen, K.; Hamari, J. The effect of audio on the experience in virtual reality: a scoping review. *Behaviour & Information Technology* **2024**, *43*, 165–199.
12. Guo, Z.; Thawonmas, R.; Ren, X. Rethinking dynamic difficulty adjustment for video game design. *Entertainment Computing* **2024**, p. 100663.
13. Almaç, N. Effect of Audio-Visual Appeal on Game Enjoyment: Sample from Turkey. *Acta Ludologica* **2023**, *6*, 42–61.
14. Phan, M.H.; Keebler, J.R.; Chaparro, B.S. The development and validation of the game user experience satisfaction scale (GUESS). *Human factors* **2016**, *58*, 1217–1247.
15. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; et al. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology* **2024**, *15*, 1–45.
16. Alkaissi, H.; McFarlane, S.I. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus* **2023**, *15*.
17. Ekman, I. Meaningful noise: Understanding sound effects in computer games. *Proc. Digital Arts and Cultures* **2005**, *17*.
18. Cowan, B.; Kapralos, B.; Collins, K. Does improved sound rendering increase player performance? a graph-based spatial audio framework. *IEEE Transactions on Games* **2020**, *13*, 263–274.
19. Cassidy, G.; MacDonald, R. The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae* **2009**, *13*, 357–386.
20. Li, T.; Ogihara, M. Detecting emotion in music **2003**.
21. Hébert, S.; Béland, R.; Dionne-Fournelle, O.; Crête, M.; Lupien, S.J. Physiological stress response to video-game playing: the contribution of built-in music. *Life sciences* **2005**, *76*, 2371–2380.
22. Monteiro, D.; Liang, H.N.; Wang, J.; Chen, H.; Baghaei, N. An In-Depth Exploration of the Effect of 2D/3D Views and Controller Types on First Person Shooter Games in Virtual Reality. In Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2020, pp. 713–724. <https://doi.org/10.1109/ISMAR50242.2020.00102>.
23. Krompiec, P.; Park, K. Enhanced Player Interaction Using Motion Controllers for First-Person Shooting Games in Virtual Reality. *IEEE Access* **2019**, *7*, 124548–124557. <https://doi.org/10.1109/ACCESS.2019.2937937>.
24. Monteiro, D.; Chen, H.; Liang, H.N.; Tu, H.; Dub, H. Evaluating Performance and Gameplay of Virtual Reality Sickness Techniques in a First-Person Shooter Game. In Proceedings of the 2021 IEEE Conference on Games (CoG), 2021, pp. 1–8. <https://doi.org/10.1109/CoG52621.2021.9619145>.
25. Holm, S.K.; Kaakinen, J.K.; Forsström, S.; Surakka, V. The role of game preferences on arousal state when playing first-person shooters **2020**.
26. Chanel, G.; Rebetez, C.; Bétrancourt, M.; Pun, T. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In Proceedings of the Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era, 2008, pp. 13–17.
27. Constant, T.; Levieux, G. Dynamic difficulty adjustment impact on players' confidence. In Proceedings of the Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–12.
28. Bostan, B.; Ögüt, S. In pursuit of optimal gaming experience: challenges and difficulty levels. In Proceedings of the Entertainment= Emotion, ed PVMT Soto. Communication présentée à l'Entertainment= Emotion Conference (Benasque: Centro de Ciencias de Benasque Pedro Pascual (CCBPP)). Citeseer, 2009.
29. Alexander, J.T.; Sear, J.; Oikonomou, A. An investigation of the effects of game difficulty on player enjoyment. *Entertainment computing* **2013**, *4*, 53–62.
30. Darzi, A.; McCrea, S.M.; Novak, D.; et al. User experience with dynamic difficulty adjustment methods for an affective exergame: Comparative laboratory-based study. *JMIR Serious Games* **2021**, *9*, e25771.

31. Vassileva, D.; Bontchev, B. Dynamic game adaptation based on detection of behavioral patterns in the player learning curve. In Proceedings of the EDULearn18: 10th annual International Conference on Education and New Learning Technologies. IATED Academy, 2018, pp. 8182–8191.
32. Baldwin, A.; Johnson, D.; Wyeth, P. Crowd-pleaser: Player perspectives of multiplayer dynamic difficulty adjustment in video games. In Proceedings of the Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, 2016, pp. 326–337.
33. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In Proceedings of the Advances in Neural Information Processing Systems; Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; Lin, H., Eds. Curran Associates, Inc., 2020, Vol. 33, pp. 1877–1901.
34. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A survey of large language models. *arXiv preprint arXiv:2303.18223* **2023**.
35. Wu, T.; He, S.; Liu, J.; Sun, S.; Liu, K.; Han, Q.L.; Tang, Y. A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica* **2023**, *10*, 1122–1136.
36. Li, X.; You, X.; Chen, S.; Taveekitworachai, P.; Thawonmas, R. Analyzing Audience Comments: Improving Interactive Narrative with ChatGPT. In Proceedings of the International Conference on Interactive Digital Storytelling. Springer, 2023, pp. 220–228.
37. Lanzi, P.L.; Loiacono, D. Chatgpt and other large language models as evolutionary engines for online interactive collaborative game design. In Proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference, 2023, pp. 1383–1390.
38. Nimpattanavong, C.; Taveekitworachai, P.; Khan, I.; Nguyen, T.V.; Thawonmas, R.; Choensawat, W.; Sookhanaphibarn, K. Am I fighting well? fighting game commentary generation with ChatGPT. In Proceedings of the Proceedings of the 13th International Conference on Advances in Information Technology, 2023, pp. 1–7.
39. Blom, P.M.; Bakkes, S.; Spronck, P. Modeling and adjusting in-game difficulty based on facial expression analysis. *Entertainment Computing* **2019**, *31*, 100307.
40. Moon, J.; Choi, Y.; Park, T.; Choi, J.; Hong, J.H.; Kim, K.J. Diversifying dynamic difficulty adjustment agent by integrating player state models into Monte-Carlo tree search. *Expert Systems with Applications* **2022**, *205*, 117677.
41. Pfau, J.; Liapis, A.; Volkmar, G.; Yannakakis, G.N.; Malaka, R. Dungeons & replicants: automated game balancing via deep player behavior modeling. In Proceedings of the 2020 IEEE Conference on Games (CoG). IEEE, 2020, pp. 431–438.
42. Khajah, M.M.; Roads, B.D.; Lindsey, R.V.; Liu, Y.E.; Mozer, M.C. Designing engaging games using Bayesian optimization. In Proceedings of the Proceedings of the 2016 CHI conference on human factors in computing systems, 2016, pp. 5571–5582.
43. Gonzalez-Duque, M.; Palm, R.B.; Risi, S. Fast game content adaptation through Bayesian-based player modelling. In Proceedings of the 2021 IEEE Conference on Games (CoG). IEEE, 2021, pp. 01–08.
44. Mohr, M.; Ermidis, G.; Jamurtas, A.Z.; Vigh-Larsen, J.F.; Poullos, A.; Draganidis, D.; Papanikolaou, K.; Tsimeas, P.; Batsilas, D.; Loules, G.; et al. Extended match time exacerbates fatigue and impacts physiological responses in male soccer players. *Medicine and Science in Sports and Exercise* **2023**, *55*, 80.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.