Familiarity Vs Performance: the Link Between Gaming Experience and Audio Recognition

Ibrahim Khan[1,*], Mustafa Can Gursesli[2], Ruck Thawonmas[3], Antonio Lanata[4]
{[1]Graduate School, [3]College} of Information Science and Engineering, Ritsumeikan University, Japan
[2]Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
[4]Department of Information Engineering, University of Florence, Italy

2025 IEEE 14th Global Conference on Consumer Electronics (GCCE) | 979-8-3315-2416-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/GCCE65946.2025.11275562

***Abstract*—This paper investigates the relationship between prior gaming experience and performance of first-person shooter (FPS) in audio-reliant tasks within sound-only environments. We conducted a study with 22 participants (18 male, 3 female, 1 non-binary; age range = 18–35 years) using a modified audio-enhanced version of ViZDoom, testing two scenarios ("basic" and "deadly corridor") under vision and non-vision conditions. Participants were categorized into three familiarity groups based on self-reported FPS experience. Results indicate that higher familiarity correlated with faster task completion and improved target engagement in both visual and non-visual modes, with performance gaps widening in the more complex non-vision scenario. While limited by sample size, these findings suggest that prior gaming skills may reduce disorientation and enhance spatial audio recognition, particularly in challenging sound-only environments.**

***Index Terms*—First-Person Shooter, Audio-Only Environments, Spatial Audio Recognition, Gaming Experience.**

I. Introduction

The literature has indicated that experience in first-person shooter (FPS) gaming can enhance visual attention, spatial reasoning, and decision-making capabilities under time constraints [1]. Nonetheless, the degree to which these skills apply to audio-reliant tasks in visually deprived environments remains insufficiently explored. Existing studies regarding auditory spatial awareness frequently concentrate on non-gaming populations [2] or simplified tasks [3], thereby leaving gaps in our understanding of how domain-specific expertise influences performance in complex and dynamic scenarios. This study aims to address this gap by evaluating whether familiarity with FPS mechanics enhances navigation and target engagement in sound-only gaming environments.

We developed a controlled experiment utilizing a modified audio-enhanced version of ViZDoom [4] (Figure 1 shows the scenarios), a standardized FPS platform, to assess participants' performance in both visual and non-visual modes across two scenarios of varying difficulty. Through the analysis of completion times, survival metrics, and target engagement rates, we investigate the correlation between self-reported gaming familiarity and task success. Our research contributes to the expanding body of literature on skill transfer between gaming and sensory processing, with implications for the design of adaptive training systems for auditory-driven tasks. While our study is limited by sample size, these findings underscore the potential for skills acquired from FPS gaming to alleviate disorientation in sound-only contexts, thus providing a basis for future research on auditory training paradigms.

The primary contributions of this study are threefold: 1) Empirical evidence demonstrating the correlation between familiarity with FPS gaming and enhanced performance in audio-reliant tasks, particularly within challenging, vision-deprived environments. 2) A methodological framework for assessing audio-spatial navigation in dynamic, game-based scenarios, utilizing standardized FPS mechanics to isolate the effects of prior experience. 3) Insights into the varying impacts of gaming expertise across different levels of task complexity, indicating that performance disparities between groups based on familiarity become more pronounced under non-visual, high-stakes conditions.

II. Related Work

Sound effects are a key part of video games, influencing immersion, emotional tone, and player perception in ways that visuals alone cannot achieve [5]. They provide subtle yet powerful cues that guide attention, build atmosphere, and support gameplay interaction [6]. This role is particularly significant in FPS games, where spatial awareness and environmental cues are central to the experience. Carefully designed audio elements, such as footsteps, gunfire, and ambient noise, immerse players in the virtual environment, enabling them to intuitively interpret space and movement. In order to enhance this experience, a variety of sound-focused algorithms have been integrated into FPS scnerios. Broderick et al. highlight the use of the Head-Related Transfer Function, which provides precise spatial audio and helps players to locate sounds more easily in games such as *Overwatch* and *Counter-Strike: Global Offensive* [7]. Similarly, Semionov and McGregor et al. demonstrate that spatialised sound significantly improves players' ability to detect threats and enhances their situational perception in FPS settings [8]. Altogether, these findings emphasise the potential of sound design, particularly when supported by adaptive and spatial audio algorithms, to create immersive and perceptually rich FPS experiences.

The performance of players in FPS games is strongly influenced by cognitive and perceptual demands, with real-time decision-making, reaction speed, and environmental awareness playing crucial roles [9]. Such games typically demand rapid responses and strategic movement under pressure, and even minor enhancements in perceptual clarity, such as improved sound and visual feedback, can have a significant impact on

* Corresponding author: gr0556vx@ed.ritsumei.ac.jp

gameplay outcomes [10]. Considering the classic sports (as mechanics, they are similar to video games, reaction times, etc.), athletes also show the same pattern. For instance, studies show that high-performers rely on a combination of sensory cues and muscle memory to improve precision and reaction time [11]. Furthermore, studies by Sadowska et al suggest that physiological stress responses, such as an elevated heart rate and increased cortisol levels, are linked to performance fluctuations during competitive FPS sessions [12]. These insights suggest that performance is shaped not only by skill, but also by perceptual support and emotional regulation. As technology advances, integrating adaptive performance and biofeedback systems may offer new ways to enhance player capabilities in challenging FPS environments. However, existing literature on this topic is limited, and more research is needed to explore the impact of different sound strategies.

III. Methodology

A. Participants and Design

Our study utilizes the ViZDoom FPS platform [4]. ViZDoom is an AI research platform based on the classic first-person shooter game Doom. It is particularly suitable for research in various research domains.

Twenty-six healthy adults (21 male, 4 female, 1 non-binary; age range = 18–35 years) (10 "very familiar," 8 "familiar," 6 "somewhat familiar" with FPS games) participated in a within-subjects study. Familiarity categories were derived from post-questionnaire self-reports, with groupings consolidated due to limited sample size. Participants completed two ViZDoom scenarios ("Basic" and "Deadly Corridor") across six rounds (three vision, three non-vision), counterbalanced to mitigate order effects.

The study was conducted in full accordance with the Declaration of Helsinki and the ethical guidelines of the American Psychological Association. Informed consent was secured from every participant before the research commenced. While approval from an ethics committee was not obtained, this was in compliance with Japanese national law, specifically the Act on the Protection of Personal Information (APPI), which does not mandate such a review.

a) Audio-enhanced ViZDoom: We chose the ViZDoom platform for this study because it supports evaluation with human participants. ViZDoom utilizes assets from Freedoom [13], an open-source project that offers free replacements for the original Doom resources. For our experiments, we focused on two scenarios, Basic and Deadly Corridor, as they capture a range of challenges typical in FPS games. Limiting the study to these two scenarios allowed us to maintain focus while still addressing key gameplay variations. For the audio-enhanced version, we incorporated sound effects informed by prior research and practices in accessible FPS game design. Specifically, we added footstep sounds and wall collision cues to support player navigation and spatial orientation, as these have been shown to be effective aids [14] and are common across game genres. In object-seeking tasks, we introduced proximity-based audio cues to help players locate targets more efficiently, consistent with findings from related studies [8], [15]. Additionally, we implemented enemy presence indicators in the Basic scenario, which originally lacked such feedback.

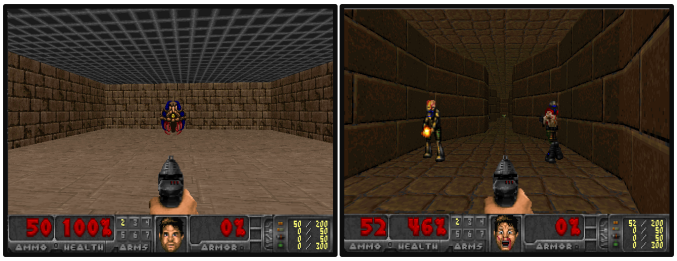

Fig. 1. The two scenarios used in this study. Basic and Deadly Corridor from left to right.

b) Basic Scenario: This initial task takes place in a rectangular environment. The player is confined to horizontal movement and a shooting action. The objective is to eliminate a single enemy, which spawns randomly on the wall opposite the player and is defeated by one hit. The trial concludes successfully when the monster is killed or unsuccessfully if the 8.5-second time limit expires.

c) Deadly Corridor Scenario: This scenario is designed to teach goal-oriented navigation under threat. The player must traverse a corridor to acquire a target item (a green vest) at the far end. Survival is complicated by six hostile enemies positioned along the corridor's sides who shoot at the player. The player has 60 seconds to complete the objective.

B. Procedure

In the "basic" scenario, players eliminated a non-responsive monster in a single room, with rounds concluding upon success or time expiration. The "deadly corridor" necessitated navigating a hostile hallway to retrieve a vest while avoiding adversaries. Non-vision modes eliminated visual feedback, requiring reliance on spatial audio cues (e.g., enemy footsteps, weapon sounds). To balance difficulty, non-vision "deadly corridor" rounds provided additional health points. Performance metrics included time-to-completion, monster eliminations, and health retention.

After completing both the non-vision and vision modes across the two scenarios, participants were asked to fill out a questionnaire based on the four factors from the Game User Experience Satisfaction Scale (GUESS) [16]. Each factor, audio aesthetics, usability/playability, play engrossment, and enjoyment, was evaluated through three questions. Responses were recorded using a 7-point Likert scale, where 1 indicated strong disagreement, 4 was neutral, and 7 indicated strong agreement. The study protocol followed is illustrated in Figure 2.

C. Analysis

Data were aggregated across rounds for each condition. Descriptive statistics and inter-group comparisons were prioritized due to sample size constraints, with trends evaluated through relative differences in means across familiarity tiers.

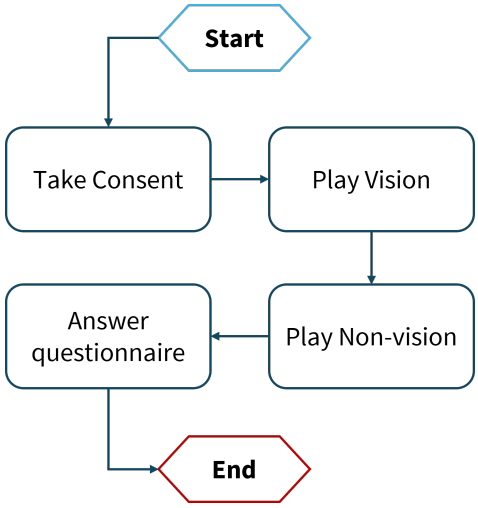

Fig. 2. Protocol of the user study.

IV. Evaluation and Results

A. Basic Scenario

In both vision and non-vision modes, participants classified as "very familiar" completed tasks in the shortest duration (vision: M = 1.656s; non-vision: M = 5.834s), followed next by those identified as "familiar" (vision: M = 2.037s; non-vision: M = 6.334s) and the "somewhat familiar" groups (vision: M = 5.584s; non-vision: M = 7.731s). The monster kill rates in vision mode were identical for the "very familiar" and "familiar" groups, whereas the rates for the "somewhat familiar" participants were lower. On the other hand in non-vision mode the kill rate was highest in "very familiar" and lowest in "somewhat familiar" group. Table I shows the details of the *Basic Scenario* evaluation.

TABLE I
Performance metrics for the Basic scenario in both vision and no-vision mode of the user study.

Setting	Familiarity	Average Time	Average Kills
Vision	Very familiar	1.656	1.000
No Vision	Very familiar	5.834	0.666
Vision	Familiar	2.037	1.000
No Vision	Familiar	6.334	0.593
Vision	Somewhat familiar	5.584	0.833
No Vision	Somewhat familiar	7.731	0.415

B. Deadly Corridor

Performance disparities manifested more distinctly in non-vision mode. "Very familiar" players maintained the highest health levels and completed tasks most rapidly. It is noteworthy that this trend was the same in non-vision mode, with the "familiar" group retained more health compared to the "somewhat familiar" participants, but less when compared to "Very familiar" players. Table VI shows the details of the *Deadly Corridor* evaluation.

TABLE II
GUESS factors from all groups.

Factor	Mean	STD
Audio Aesthetics	17.07	1.91
Usability/Playability	15.31	3.01
Play Engrossment	14.21	3.83
Enjoyment	14.55	3.43

TABLE III
GUESS factors from the very familiar group.

Factor	Mean	STD
Audio Aesthetics	17.45	2.16
Usability/Playability	15.73	2.97
Play Engrossment	13.73	4.54
Enjoyment	15.27	3.29

TABLE IV
GUESS factors from the familiar group.

Factor	Mean	STD
Audio Aesthetics	16.82	1.54
Usability/Playability	15.73	2.69
Play Engrossment	13.91	2.21
Enjoyment	13.73	3.95

TABLE V
GUESS factors from the somewhat familiar group.

Factor	Mean	STD
Audio Aesthetics	16.86	2.19
Usability/Playability	14.00	3.61
Play Engrossment	15.43	4.86
Enjoyment	14.71	2.93

C. Guess Factors

Table II summarizes the overall results of the user study using the GUESS framework, which evaluates four core dimensions of game user experience: audio aesthetics, usability/playability, play engrossment, and enjoyment. Each dimension was assessed through three items on a 7-point Likert scale, giving a maximum possible score of 21 per factor. Across all participants, the results show above-average ratings in all four dimensions, indicating a generally positive reception of the audio-enhanced ViZDoom experience. audio

TABLE VI
PERFORMANCE METRICS FOR THE DEADLY CORRIDOR SCENARIO IN BOTH VISION AND NO-VISION MODE OF THE USER STUDY.

Setting	Familiarity	Average Health	Average Kills	Total Average Deaths	Average Time
Vision	Very familiar	12.727	4.303	1.800	8.587
No Vision	Very familiar	6.000	3.274	2.800	28.795
Vision	Familiar	0.222	3.186	2.650	7.347
No Vision	Familiar	-10.667	2.037	2.875	26.090
Vision	Somewhat familiar	-8.800	1.635	3.000	5.177
No Vision	Somewhat familiar	-14.500	1.305	3.550	15.743

aesthetics received the highest mean score (M = 17.07, SD = 1.91), highlighting strong appreciation for the sound design. This was followed by usability/playability (M = 15.31, SD = 3.01), suggesting that the added auditory cues improved gameplay accessibility and control. Enjoyment (M = 14.55, SD = 3.43) and Play Engrossment (M = 14.21, SD = 3.83) also received favorable scores, reflecting participants' immersion and satisfaction.

To further analyze how prior experience with similar games might influence perception, Tables III, IV, and V break down the results by participants' familiarity level with FPS games. The very familiar group rated audio aesthetics the highest (M = 17.45), along with high usability/playability (M = 15.73), suggesting that experienced players still found value in the audio enhancements. The Familiar group showed similar trends with audio aesthetics (M = 16.82) and usability/playability (M = 15.73), but slightly lower scores in enjoyment and play engrossment. Interestingly, the Somewhat Familiar group reported the highest play engrossment (M = 15.43) among the three groups, implying that players with moderate experience may benefit most from enhanced audio cues in terms of immersion. Overall, the findings across all familiarity levels reinforce the effectiveness of accessible sound design in enhancing usability, enjoyment, and engagement in FPS gameplay.

D. Discussion of Trends

Despite being constrained by small subgroup sizes, the results indicate that familiarity with FPS games correlates with a reduction in disorientation during audio-only tasks, particularly within complex scenarios. The widening performance gap in the non-vision conditions of the "deadly corridor" suggests that prior gaming experience may enhance auditory spatial mapping under pressure. These preliminary findings underscore the potential utility of skills derived from FPS games in audio-dependent applications, although further research is required to delineate specific cognitive mechanisms involved.

V. DISCUSSION

This study's findings demonstrate a significant correlation between a player's self-reported familiarity with FPS games and their performance in audio-reliant tasks, particularly when visual information is absent. The results suggest that skills honed in visually rich FPS environments are transferable to sound-only contexts, challenging the notion that gaming expertise is purely visual. Participants with higher familiarity consistently outperformed their less experienced counterparts in task completion times and target engagement across both scenarios.

The most compelling evidence for this skill transfer emerged from the "Deadly Corridor" scenario. The performance gap between familiarity groups widened considerably in the non-vision mode, indicating that prior gaming experience provides a distinct advantage when tasks are complex and pressure is high. Experienced players likely possess a more developed cognitive framework for spatial audio mapping, allowing them to interpret auditory cues not just as sounds but as actionable data points representing threats and objectives in a three-dimensional space. This enhanced auditory-spatial reasoning appears to mitigate the disorientation that typically accompanies the loss of vision, a conclusion supported by the superior completion times and health retention of the "very familiar" group in the non-visual corridor task. This aligns with existing literature [17], [18] suggesting FPS gaming enhances cognitive functions like spatial reasoning and decision-making under pressure, and this study extends that understanding into the auditory domain.

This study hints at the importance of stress adaptation, a skill honed in the high-stakes environments of FPS games. The "Deadly Corridor" was designed to teach goal-oriented navigation under threat from hostile enemies. Experienced players may have a more regulated physiological response to such threats, as suggested by literature [18], allowing them to maintain cognitive focus and execute tasks more effectively. In the non-vision mode, where threats are perceived only through sound, this ability to manage stress and make decisions based on auditory information becomes paramount.

Furthermore, the overall positive user experience reflected in the GUESS ratings reinforces the effectiveness of the implemented audio design in the audio-enhanced ViZDoom. The highest-rated dimension across all participants was audio aesthetics (M = 17.07, SD = 1.91), followed by usability/playability (M = 15.31, SD = 3.01). This indicates that the auditory elements were not only functional but also perceived as high-quality and integral to the gameplay experience.

A deeper look into subgroup results across varying familiarity levels further supports this observation. Participants in the very familiar group had the highest score for audio aesthetics (M = 17.45, SD = 2.16), showing that even seasoned players appreciated the sophistication and clarity of the sound design. This group also rated usability/playability highly (M = 15.73), suggesting that the added auditory feedback contributed positively to their interaction and control within the game. Interestingly, the Somewhat Familiar group reported the highest levels of Play Engrossment (M = 15.43), implying that players with moderate experience may have benefited most from the audio cues in terms of immersion. The auditory enhancements likely bridged gaps in familiarity, helping these users become more absorbed and comfortable in the non-visual gameplay.

These consistently favorable ratings across groups can be

attributed to the thoughtful integration of specific audio cues, such as footstep sounds, wall collision alerts, proximity indicators, and enemy presence signals. These cues provided crucial spatial and gameplay information, especially in the absence of visual feedback. The fact that players across all experience levels responded positively suggests that the accessible sound design was not only intuitive and helpful but also enhanced the emotional and aesthetic experience of gameplay. Together, these findings underscore the importance of audio-first design strategies in developing accessible and enjoyable gaming experiences. Incorporating functional yet engaging audio elements can significantly enhance both usability and player satisfaction, making games more inclusive for players with visual impairments and more immersive for the general player base.

VI. Limitation

While this study provides valuable insights, its findings should be considered in light of several methodological limitations. The research was conducted with a small number of participants, a factor that constrained the statistical analysis and required the consolidation of participant familiarity groups. The sample was also demographically specific, composed primarily of male participants within an 18-35 age range, which restricts the generalizability of the findings to a broader population. Furthermore, the primary variable of player experience was determined by self-reported familiarity categories collected after the experiment, an inherently subjective method.

The experimental design and data reporting also present limitations. To balance difficulty, the methodology granted participants additional health points specifically for the non-vision rounds of the "deadly corridor" scenario, introducing a confounding variable that complicates direct comparisons of survival between vision and non-vision modes. The study's scope was confined to two specific scenarios, "Basic" and "Deadly Corridor," which may not represent the full spectrum of challenges in FPS games, potentially limiting the applicability of the results to other gameplay types.

VII. Conclusion

This study demonstrates that prior FPS gaming experience significantly enhances performance in audio-reliant tasks, particularly in vision-deprived, high-pressure environments. Participants with higher familiarity exhibited faster task completion and improved target engagement. The widening performance gap in the non-visual "deadly corridor" scenario underscores the transferability of gaming-acquired skills, such as spatial audio mapping and stress adaptation, to sound-only contexts. The larger differences in health retention trends between familiarity groups under non-visual conditions as compared to visual conditions suggest nuanced interactions between gaming expertise and sensory reliance, warranting deeper investigation. Future studies should expand to larger cohorts, employ objective familiarity measures, and explore cognitive mechanisms (e.g., auditory working memory) to refine auditory learning frameworks. This study bridges gaming expertise and sensory task performance, offering foundational insights for both game design and applications requiring sound-driven spatial navigation.

References

[1] I. Spence and J. Feng, "Video games and spatial cognition," *Review of general psychology*, vol. 14, no. 2, pp. 92–104, 2010.

[2] P. Voss, "Auditory spatial perception without vision," *Frontiers in psychology*, vol. 7, p. 1960, 2016.

[3] V. Nair, S.-e. Ma, R. E. Gonzalez Penuela, Y. He, K. Lin, M. Hayes, H. Huddleston, M. Donnelly, and B. A. Smith, "Uncovering visually impaired gamers' preferences for spatial awareness tools within video games," in *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*, 2022, pp. 1–16.

[4] M. Wydmuch, M. Kempka, and W. Jaskowski, "ViZDoom Competitions: Playing Doom from Pixels," *IEEE Transactions on Games*, vol. 11, no. 3, pp. 248–259, 2019, the 2022 IEEE Transactions on Games Outstanding Paper Award.

[5] K. Collins, *Game sound: an introduction to the history, theory, and practice of video game music and sound design*. Mit Press, 2008.

[6] M. Bélteky, "Emotion in video game audio: Can sound effects be perceived to contain positive emotional content?" 2019.

[7] J. Broderick, J. Duggan, and S. Redfern, "The importance of spatial audio in modern games and virtual environments," in *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 2018, pp. 1–9.

[8] K. Semionov and I. McGregor, "Effect of various spatial auditory cues on the perception of threat in a first-person shooter video game," in *Proceedings of the 15th International Audio Mostly Conference*, 2020, pp. 22–29.

[9] P. Ng, K. Nesbitt, and K. Blackmore, "Sound improves player performance in a multiplayer online battle arena game," in *Australasian Conference on Artificial Life and Computational Intelligence*. Springer, 2015, pp. 166–174.

[10] P. Ng and K. Nesbitt, "Informative sound design in video games," in *Proceedings of the 9th Australasian conference on interactive entertainment: matters of life and death*, 2013, pp. 1–9.

[11] N. Schaffert, K. Mattes, and A. O. Effenberg, "A sound design for acoustic feedback in elite sports," in *International Symposium on Computer Music Modeling and Retrieval*. Springer, 2009, pp. 143–165.

[12] D. Sadowska, T. Sacewicz, K. Rebiś, T. Kowalski, and J. Krzepota, "Examining physiological changes during counter-strike: Global offensive (cs: Go) performance in recreational male esports players," *Applied Sciences*, vol. 13, no. 20, p. 11526, 2023.

[13] F. Project, "Freedoom," accessed: 2025-03-14. [Online]. Available: https://freedoom.github.io/

[14] C. Johanson and R. L. Mandryk, "Scaffolding player location awareness through audio cues in first-person shooters," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 3450–3461.

[15] C. Chen, U. Jain, C. Schissler, S. V. A. Gari, Z. Al-Halah, V. K. Ithapu, P. Robinson, and K. Grauman, "Soundspaces: Audio-visual navigation in 3d environments," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 2020, pp. 17–36.

[16] M. H. Phan, J. R. Keebler, and B. S. Chaparro, "The development and validation of the game user experience satisfaction scale (guess)," *Human factors*, vol. 58, no. 8, pp. 1217–1247, 2016.

[17] C. Reynaldo, R. Christian, H. Hosea, and A. A. Gunawan, "Using video games to improve capabilities in decision making and cognitive skill: a literature review," *Procedia Computer Science*, vol. 179, pp. 211–221, 2021.

[18] F. Neri, C. L. Smeralda, D. Momi, G. Sprugnoli, A. Menardi, S. Ferrone, S. Rossi, A. Rossi, G. Di Lorenzo, and E. Santarnecchi, "Personalized adaptive training improves performance at a professional first-person shooter action videogame," *Frontiers in Psychology*, vol. 12, p. 598410, 2021.