

Article

Not peer-reviewed version

The Novice, The Expert, and The Algorithm: A Comparative Analysis of Human Expertise Transfer and AI Performance in Audio-Only Gaming Environments

[Ibrahim Khan](#)*, [Thai Van Nguyen](#), [Cvetković Tijan Juraj](#), [Ruck Thawonmas](#)*

Posted Date: 1 October 2025

doi: 10.20944/preprints202510.0077.v1

Keywords: audio-only gaming; game accessibility; reinforcement learning; human-AI comparison; expertise transfer; auditory display; player performance



Preprints.org is a free multidisciplinary 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 open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

The Novice, The Expert, and The Algorithm: A Comparative Analysis of Human Expertise Transfer and AI Performance in Audio-Only Gaming Environments

Ibrahim Khan ^{1,*}, Van Thai Nguyen ¹, Cvetković Tijan Juraj ² and Ruck Thawonmas ³

¹ Graduate School of Information Science and Engineering, Ritsumeikan University

² LABSS-ISTC, Consiglio Nazionale delle Ricerche Roma

³ College of Information Science and Engineering, Ritsumeikan University

* Correspondence: gr0556vx@rd.ritsumei.ac.jp

Abstract

This study provides a symmetrical, cross-genre comparison of human expertise transfer and "blind" artificial intelligence performance in audio-only gaming environments. Although previous research has focused on human performance in audio games and the feasibility of reinforcement learning agents trained on auditory inputs separately, a direct comparison of these two forms of expertise is missing from the literature. We fill this gap with a robust experimental design involving 37 human players (5 females, 31 males, 1 non-binary individual, aged 18-44), grouped by gaming experience, and specialized blind Artificial Intelligence (AI) agents across two different genres: a fighting game (Dare-FightingICE) and a first-person shooter (SonicDoom). Our findings show a complex, task-dependent relationship between performance. In DareFightingICE, expert humans significantly outperformed the AI, demonstrating effective cognitive transfer. Meanwhile, AI's performance matched the overall human average, establishing a key benchmark. Conversely, in SonicDoom, AI achieved superhuman speed in simple tasks but underperformed compared to expert humans in complex scenarios. This difference highlights that the AI's proficiency is specialized but fragile, whereas human expertise is more robust and adaptable. The results provide practical insights for designing audio-rich games and highlight the crucial need for AI models that extend beyond reactive policies.

Keywords: audio-only gaming; game accessibility; reinforcement learning; human-AI comparison; expertise transfer; auditory display; player performance

1. Introduction

The auditory dimension of interactive entertainment has evolved from a supplementary feature to a critical element of gameplay, immersion, and accessibility [1,2]. Most players find that sound improves their visual experience, but for a growing community of players with visual impairments and in the early stages of audio-only games, sound becomes the main way they perceive and navigate [3]. This reliance on non-visual modalities creates a unique problem space that not only drives innovation in accessible design [3] but also offers a powerful, constrained environment for investigating the fundamental principles of auditory cognition, skill acquisition, and intelligent decision-making in both humans and machines.

In this audio-centric paradigm, two distinct forms of expertise emerge. The first is **human cognitive transfer**, where experienced video game players leverage visually-honed skills, such as spatial awareness, timing, and strategic planning, and adapt them to an audio-only context. Research shows that expert players develop robust cross-modal representations of game mechanics, allowing them to retain effectiveness even when deprived of vision [4–6]. However, the robustness and

generalizability of this transfer remain open questions. It is unclear to what extent the cognitive advantages of gaming experience are universal or tightly coupled to the visual modality, a critical distinction for designing audio systems that are intuitive for players with pre-existing gaming skillsets.

The second form of expertise is that of **artificial intelligence**, specifically "blind" reinforcement learning (RL) agents trained exclusively on auditory input. While RL has achieved superhuman performance in visually complex games like Atari and StarCraft II [7,8], the development of agents operating on raw audio represents a specialized frontier [9]. These agents learn auditory policies natively, without prior visual experience, offering a fundamentally different model of skill acquisition. While such AIs are expected to excel at tasks requiring frame-perfect reactions and pattern recognition, their proficiency in tasks demanding abstract spatial reasoning and adaptive strategy is less certain [10,11].

A significant gap exists in the literature at the confluence of these two domains. While studies have explored human performance in audio games and separate research has demonstrated the feasibility of blind AI agents, a **direct, symmetrical comparison is conspicuously absent**. We lack a holistic understanding of how the transferred skills of human experts compare to the natively learned policies of specialized AIs, particularly across genres demanding different cognitive and reflexive abilities. This paper addresses this gap by positing a core question: *Do humans and AI excel at the same types of audio-based tasks, and where do the strengths of human cognitive transfer and machine-learned policies diverge?*

In this sense, we introduce a fully symmetrical, two-genre comparative framework. We evaluate the performance of human players, stratified into **Somewhat familiar, Familiar, and Very familiar** groups, and two specialized blind AI agents in both a fighting game (*DareFightingICE* [12]) and a first-person shooter (FPS) (*ViZDoom*[13]). These genres were purposefully selected to probe distinct skill sets: the fighting game emphasizes **precise timing and reaction** to discrete audio cues. At the same time, the FPS demands **continuous spatial reasoning and navigation** within a complex soundscape. This dual-genre design enables us to move beyond simply asking if an AI can play with audio to examining how its performance profile compares to that of humans and whether these profiles are genre-dependent.

This study makes the following primary contributions to the fields of game accessibility, human-computer interaction, and artificial intelligence:

1. We present the **first symmetrical, cross-genre experimental design** that directly compares the performance of blind AI agents against human players of varying skill levels in audio-only gaming environments.
2. We test the **generalizability of human gaming expertise** to the auditory domain, providing evidence on how high-level skills transfer across sensory modalities.
3. We offer insights into a set of **actionable design strategies** to improve the effectiveness and accessibility of audio-first games.

2. Related Work

2.1. Sound in Video Games

Video games are increasingly accepted as complex artistic creations that incorporate a variety of sensory and design elements [14]. Fundamentally, they operate as audio-visual systems that form interactive relationships with players [15,16]. Although visuals are usually the first element encountered, games rely on various mechanisms to maintain engagement, such as narrative, level design, and progression systems [17]. Notably, "sound" has emerged as an important element in sustaining engagement. The role of sound design in gaming extends far beyond simply providing background music [18]. It actively shapes the experience by reinforcing what's happening in the moment, hinting at what's to come, and guiding the player's focus. This dual role strengthens immersion and facilitates a personal connection between the player and the world of the game [19].

Game audio's role in creating a sense of being immersed in a virtual world is a widely recognized contribution. This is commonly defined as a state of deep mental involvement, where players lose

awareness of the real world and experience a heightened sense of presence in the fictional one [20]. Moreover, studies show that audio plays a crucial role in sensory immersion, whereby the combination of sound and visuals overwhelms real-world stimuli and directs players' focus towards the game [20,21]. When designed effectively, audio can make the virtual environment seem more realistic, giving players the feeling that they are genuinely "inside" the game.

Beyond immersion, sound has a considerable impact on the broader player experience, which includes enjoyment, engagement, flow, and social interaction [22]. Designers have significantly advanced the quality and complexity of game audio in recent years, recognizing that a richer auditory environment contributes directly to a more compelling and memorable gaming experience. A key way in which audio contributes to this experience is through its impact on enjoyment and engagement. Sound effects heighten excitement, satisfaction, and memorability during gameplay. In addition literature show that background sound effects, in particular, enhance immersion, elicit positive emotions [22,23]. Thus, an engaging soundtrack or precisely timed sound effects can transform a routine sequence into an emotionally resonant highlight, amplifying the immediacy and long-term satisfaction of play.

2.2. Sound and Performance in First Person Shooter and Fighting Games

In high-intensity genres such as first-person shooters (FPS) and fighting games, sound serves to enhance performance. In FPS games, for example, sound such as footsteps, weapon fire, and ambient noise are central to spatial awareness and situational judgement. Spatial audio techniques, including the Head-Related Transfer Function (HRTF), have been shown to enhance the ability to localise sounds in competitive games such as Overwatch and Counter-Strike: Global Offensive [24]. Further research highlights that spatialised sound enhances players' ability to detect threats and improves situational perception [25]. As FPS performance relies on rapid decision-making, reaction time, and stress management, even subtle improvements in perceptual clarity can impact outcomes. Studies also show that physiological responses such as elevated heart rate and increased cortisol levels affect performance during competitive play [26], suggesting that future integration of adaptive audio and biofeedback systems could optimize skill performance. In addition, the performance of the players in the FPS games is closely related to the cognitive and perceptual demands. Real-time decision-making, rapid reaction speed, and environmental monitoring are crucial to success [27]. Even minor improvements in perceptual clarity through enhanced audio and visual feedback can influence gameplay outcomes [28]. This parallels findings in traditional sports, where high performers rely on multisensory integration and practiced responses to optimize performance under pressure [29].

Fighting games also make extensive use of sound to reinforce gameplay dynamics and enhance player performance. Sounds are often used to distinguish between characters, highlight attack sequences, or indicate successful hits. Games such as Tekken 7 [30], Mortal Kombat 11 [31], and Killer Instinct [32] incorporate a wide range of sound effects to enhance player responsiveness, including distinct character-based audio features and stage completion indicators. More recently, Street Fighter 6 has introduced advanced audio design elements, such as distance indicators and sound variations that correspond to attack strength. This feature helps players to gauge combat situations more precisely. These auditory layers improve clarity and add depth to the competitive experience by reinforcing timing, spacing, and tactical decision-making.

Beyond static sound effects, recent studies have explored adaptive approaches to sound in fighting games. For example, Khan et al. [33] investigated the use of adaptive background music (BGM) as a dynamic gameplay element, demonstrating that a deep learning agent performed significantly better when supported by adaptive BGM than when relying on static sound design. Such findings highlight the potential of responsive, context-sensitive audio systems to influence performance outcomes and facilitate deeper engagement.

Together, FPS and fighting games demonstrate how sound functions as one of the main gameplay mechanics rather than as a supplementary enhancement. Although FPS research emphasizes situational awareness and perceptual clarity, fighting game design showcases the role of sound in timing, spacing, and feedback.

2.3. Accessibility in Video Games

The implementation of sound as a crucial component in video games has long been seen as a way of making them more accessible, particularly for players with visual impairments [34]. As well as improving immersion, sound can be used for navigation and providing feedback, enabling players to engage with game environments without relying purely on visual cues. For many visually impaired players (VIPs), audio is the main way information about the spatial layout, how to interact with objects, and the mechanics of the game is conveyed. This effectively transforms sound from a supportive element into a fundamental part of how the game is played (REF). Early accessible games such as *Shades of Doom* [35], *AudioQuake* [36], and *Terraformers* [37] led the way in demonstrating how auditory design could effectively communicate spatial orientation, object distance, and environmental details. They relied on techniques such as stereo panning, directional tones, sonar-like pulses and musical earcons to translate three-dimensional virtual spaces into auditory experiences. These experiments demonstrated that gameplay could be fully realized through audio alone, challenging the long-standing assumption that video games are primarily visual media.

As accessible design practices matured, researchers and developers introduced more advanced systems to refine these auditory strategies. Tools such as *NavStick* [38] and *Surveyor* [39] built on earlier methods by providing real-time spatial feedback and line-of-sight guidance to help players identify unexplored areas within virtual environments. These innovations reduced cognitive load for VIPs and supported more efficient navigation for sighted players, demonstrating that inclusive design can enhance the overall gaming experience [40]. Similarly, the integration of voice-over narration and layered audio cues in mainstream titles has made menus, tutorials and complex environments more navigable for a wider audience.

2.4. Audio and Player Performance

Literature has shown that sound design extends beyond aesthetic immersion to directly impact player performance. Audio cues can take the role of reaction triggers, navigational aids, or timing mechanisms that enable faster and more accurate responses to in-game events. In rhythm-action games, for instance, players rely heavily on musical synchronization to optimize timing and performance [41]. In fast-paced genres such as fighting games and first-person shooters, auditory feedback plays a similar role by helping players anticipate attacks, locate opponents, and make split-second tactical decisions [42]. The effectiveness of audio in performance contexts is further supported by work on training and expertise. Experienced players demonstrate improved auditory recognition and response times compared to novices, suggesting that auditory expertise develops through repeated exposure and practice [43]. Moreover, studies have found that variations in audio presentation, such as distinctiveness of the cue or adaptivity, can influence both reaction speed and decision accuracy [44].

These findings highlight that audio design is not only an accessibility consideration but also a performance-critical factor. However, although previous studies have demonstrated the importance of sound for human players, the potential of auditory channels for artificial agents has received significantly less attention. Our study addresses this gap by comparing human and AI performance when constrained to auditory information, revealing the role of auditory expertise in shaping effectiveness and immersion across different game genres.

3. Methodology

In this study, we investigate the performance of human players and AI agents in audio-only gaming environments. The experimental setup followed a **2 (Player Type: Human, AI) x 2 (Genre: Fighting, FPS)** comparative design. For the human component, a within-subjects factor of **2 (Modality: Vision, Audio-Only)** was used to directly measure the effect of visual deprivation. This design facilitates a rigorous, symmetrical comparison of natively learned AI policies against transferred human skills across two distinct gaming contexts.

3.1. Experimental Platforms & Tasks

In order to probe a diverse range of auditory skills, we selected two research platforms from genres that emphasize fundamentally different abilities: an FPS that demands **spatial navigation and reasoning**, and a fighting game that prioritizes **timing and reaction**.

3.1.1. FPS Testbed: SonicDoom

For the FPS genre, we employed the **SonicDoom** [45] framework, an auditory enhanced version of the research platform VizDoom [13]. SonicDoom features sound source localization, distinct audio cues for enemies and items, footsteps sounds, and wall collision sounds. The specific tasks for this study were the "Basic" and "Deadly Corridor" scenarios from the VizDoom environment.

The Basic scenario involves a straightforward setup wherein the agent or player is situated within a square room, facing a monster positioned on the opposite side. The player's objective is to eliminate the monster before the allotted time expires. The Deadly Corridor scenario requires the agent or player to navigate a narrow, linear hallway while avoiding or neutralizing enemies positioned on both sides of the corridor to retrieve a green vest at the end. A round of the Basic scenario lasted 8 seconds, while a round of Deadly Corridor lasted 60 seconds, which is the default length of these scenarios. Figure 1 shows an example of both scenarios.

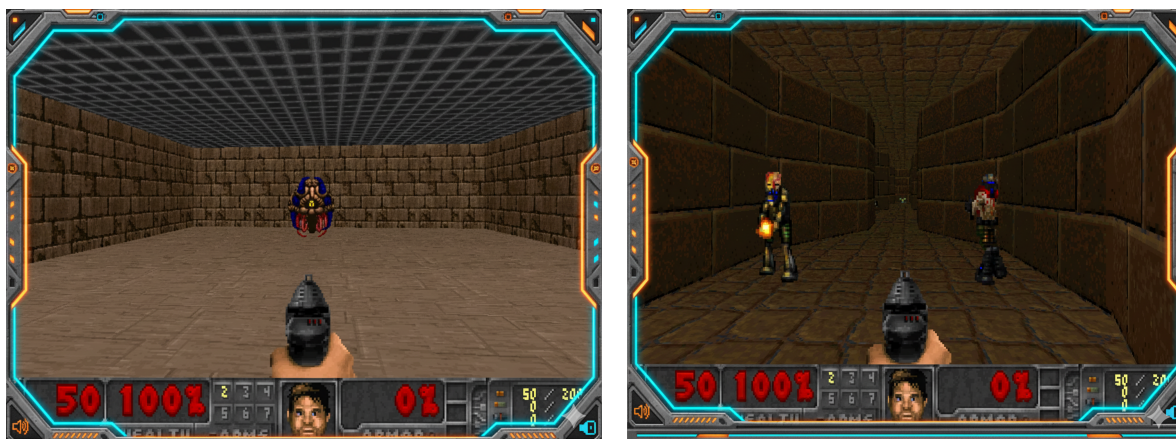


Figure 1. Basic and Deadly Corridor scenarios in VizDoom.

3.1.2. Fighting Game Testbed: DareFightingICE

Our fighting game platform is **DareFightingICE** [12], a Java-based framework with a sound design focusing on the VIPs. The platform is open-source and provides a controlled environment for AI and human-computer interaction research. The task for all participants and agents was to compete in a best-of-three match against a standardized MCTS (Monte Carlo Tree Search) opponent []. Figure 2 shows an example of the DareFightingICE platform.



Figure 2. An example of the DareFightingICE platform.

The audio system was a critical component of this testbed. We used the winning sound design from the 2022 DareFightingICE Sound Design and AI competition [46]. This sound design features **52 distinct, high-clarity sound effects** that encode gameplay information, including character movement (walking, jumping), attack types (light, heavy, special), character state (damage taken, stun), and relative horizontal distance to the opponent, which is conveyed through stereo panning. The sound design also has three special sound effects designed to help the non-visual play.

1. Heartbeat: This sound plays when the player's health drops below 50. For player one, the sound plays through the left speaker, and for player two, through the right.
2. Energy Increase: This sound plays when the player's energy rises by 50 from the previous value. For player one, the sound is heard on the left speaker, and for player two, on the right.
3. Border Alert: This sound plays when a player reaches the end of the stage on either side. The sound plays on the left side when a player reaches the left end and on the right side when they reach the right end.

3.2. AI Agent Implementation

A unified reinforcement learning (RL) approach was used to develop blind AI agents for both platforms, ensuring that performance differences could be attributed to the demands of the genre rather than the agent's core architecture.

Agents were trained using the **Proximal Policy Optimization (PPO)** algorithm [47], a widely adopted policy gradient method in recent years. PPO extends previous approaches by incorporating a trust region-like objective that stabilizes policy updates. It remains one of the state-of-the-art reinforcement learning algorithms.

3.2.1. VizDoom

For VizDoom, we used a deep reinforcement learning blind AI agent introduced in [45] and trained it using only sound as input. For the sake of a self-contained presentation, an overview is given below, summarized from previous work.

Audio observation: VizDoom operates at 35 frames per second, so each frame spans approximately 29 ms in real time. The audio signal is sampled at 22,050 Hz, and we aggregate four consecutive frames, resulting in 2,520 samples per channel. Consequently, at each timestep the sound waveform provided by VizDoom can be represented as a vector $s \in \mathbb{R}^{2 \times 2520}$, where each element s_i lies within the range $[-1, 1]$. A Fast Fourier Transform (FFT) encoder, similar to the approach in [45], is employed. This encoder comprises an FFT block, followed by a max-pooling layer and two fully connected layers. Both channels are processed separately before being concatenated into a single vector, which is then fed into a policy network, consisting of a fully connected layer network and a Gated Recurrent Unit. The overall architecture of the agent and the audio encoder is shown in Figure 3.

Training: We trained AI agents using Sample Factory[48], a reinforcement learning framework, used in a previous work[45]. We used the same hyperparameter set as in [48]. All experiments were run on a single server, which has a 116-core CPU with an Nvidia A100 GPU. The scenarios in use were Basic and Deadly Corridor.

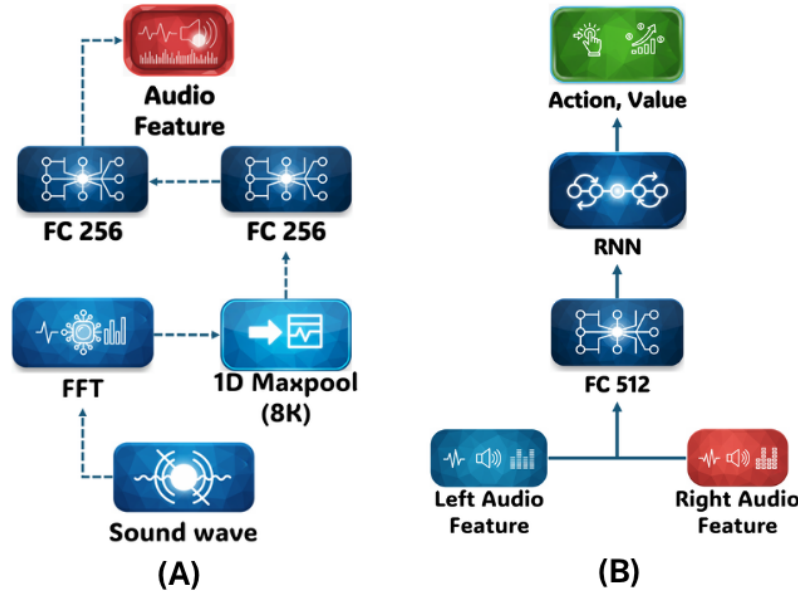


Figure 3. Illustration of (A) the audio encoder and (B) the agent architecture.

3.2.2. DareFightingICE

For DareFightingICE, we employed the deep reinforcement learning *Blind AI* [9] to perform the evaluation. This agent provides three alternative methods for processing the audio data from the DareFightingICE platform, corresponding to three distinct audio encoders: a one-dimensional Convolutional Neural Network (1D-CNN), a Fast Fourier Transform, and a Mel spectrogram (Mel).

1D-CNN: The input audio is first downsampled by selecting every eighth sample, thereby reducing computational complexity. The downsampled signal is then processed through two one-dimensional convolutional layers. This procedure yields a 32×5 audio-feature vector.

Fast Fourier Transform: An FFT transforms the input audio signal into the frequency domain, after which the magnitudes are converted to their natural logarithms. The resulting data are subsequently downsampled, producing a one-dimensional audio feature vector with 256 elements.

Mel: A short-time Fourier transform (STFT) converts the input audio into a frequency-domain spectrogram by applying a series of Fourier transforms to overlapping, windowed segments of the signal. The resulting spectrogram is then mapped onto the Mel scale. For our configuration, we use a 25 ms window, a 10 ms hop size, and 80 Mel-frequency components. Finally, the Mel spectrogram is passed through two two-dimensional convolutional layers, yielding a $32 \times 40 \times 1$ audio-feature vector.

Network architecture: The blind AI model begins with an encoder chosen from the three aforementioned options. Its output is passed to a Gated Recurrent Unit (GRU), which captures temporal dependencies in the audio sequence. Finally, three fully connected layers produce 40 output units, each corresponding to a possible action. The three encoders' architecture are shown in Figure 4.

Training. The Blind AI was trained for 900 rounds on DareFightingICE (with each game consisting of three rounds) against MCTSAI23i, a weakened Monte Carlo Tree Search (MCTS) agent [49]. We used the same hyperparameter as the previous study [9]. Blind AI's performance was then assessed over 90 evaluation rounds against the same opponent. We reported the win ratio and the average hit-point (HP) difference at the end of each round.

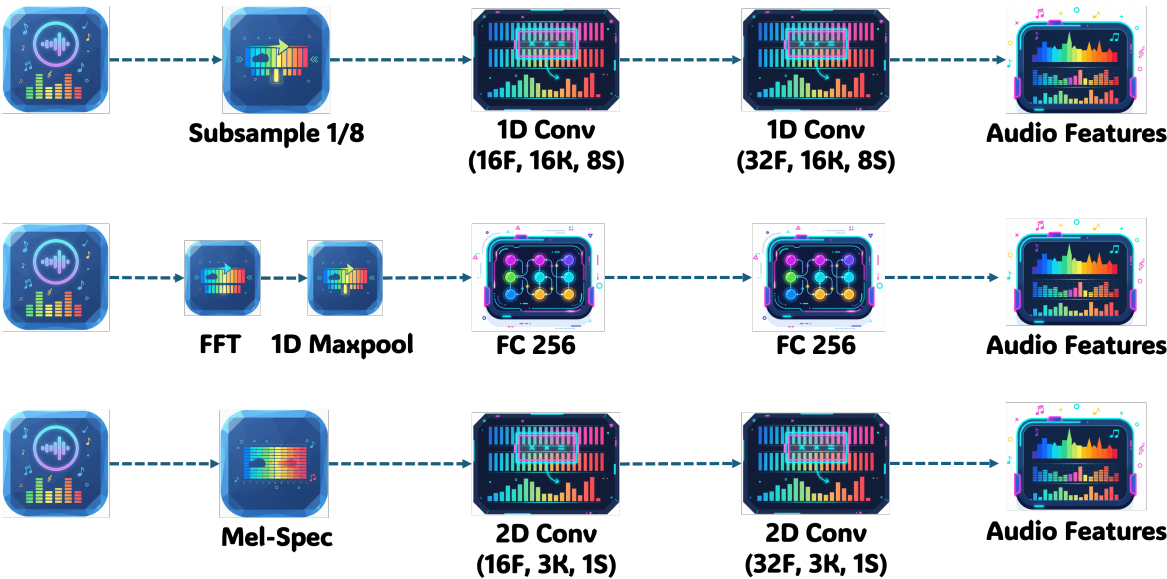


Figure 4. DareFightingICE’s blind AI encoders: 1D-CNN (top), Fast Fourier Transform (middle) and Mel (bottom)

3.3. Human-Participant Study
3.3.1. Procedure

In this study, we selected to utilize sighted participants as proxies for visually impaired individuals to collect data on our game. Employing sighted participants under conditions that simulate blindness is a widely accepted practice in the field of game accessibility research. Previous studies have consistently used blindfolded sighted users to assess early versions of accessible games, as recruiting a substantial number of VIPs can pose significant challenges. [50,51]. Furthermore, sound and game design should be refined before testing with such an audience. Nonetheless, we recognize the limitation of our study in not incorporating VIPs at this stage. However, research indicates that visually impaired individuals tend to perform better than sighted individuals when blindfolded, implying that any performance-related findings are likely to improve when tested with visually impaired individuals. [50,52]. The study was conducted by the principles of the Declaration of Helsinki and its subsequent amendments, as well as the research guidelines of the American Psychological Association. Ethical committee approval was not required under Japan’s national regulations regarding privacy and informed consent, specifically the Act on the Protection of Personal Information (APPI).

The study followed the same structure for both games. At the start, participants were given a tutorial period to familiarize themselves with the platform and controls. After the tutorial phase, participants played in vision mode and then in non-vision mode. In the non-vision mode, the device screen was turned black to simulate the absence of visuals. The flowchart of both games is illustrated in Figure 5.



Figure 5. The diagram presents the flow of the user study for both games.

3.3.2. Participants

We recruited 37 participants (5 females, 31 males, 1 non-binary individual, aged 18-44), who were categorized into three groups based on self-reported gaming experience: **Somewhat familiar**, **Familiar**, and **Very familiar**. The study followed a counterbalanced design where each participant played both games. Within each game, they completed two blocks: one with full vision and one in an audio-only condition. The order of games and conditions was randomized to mitigate learning effects.

Participants received a brief training session before each audio-only block to familiarize themselves with the sound design. Out of 37 participants, 40% were very familiar, 38% were familiar, and 22% were somewhat familiar with FPS games, and 27% were very familiar, 35% were familiar, and 38% were somewhat familiar with fighting games. Due to the modest number of participants, we combined the familiarity groups into the above three.

3.3.3. Measures

Participants completed a Game User Experience Satisfaction Scale (GUESS)[53] questionnaire after each game type. Each participant played both game genres in a randomized order to ensure fairness. The selected GUESS factors were Audio Aesthetics, Usability/Playability, Play Engrossment, and Enjoyment. We chose three relevant questions for each of these factors. GUESS uses a 7-point Likert scale, with Cronbach’s alpha values ranging from 0.75 to 0.88, indicating strong reliability across the subscales. Moreover, objective performance data were logged for all sessions. The evaluation metrics for DareFightingICE included win rate and the average health difference between the two players, while those for SonicDoom comprised average time-to-completion, average kills, total deaths, and average remaining health. Subjective experience was assessed using the GUESS questionnaire [53].

3.4. Data Analysis

In order to compare participants’ evaluations across the three Expertise groups, we used non-parametric tests, as the Shapiro–Wilk normality tests indicated several significant deviations from normality and group sizes were unbalanced. Although averaged Likert-type scores are often treated as continuous, non-parametric methods provide a more conservative and robust choice. Accordingly, we ran Kruskal–Wallis tests separately for each GUESS subscale, followed by Dunn post-hoc tests with Holm correction [54]. We reported η^2H as an omnibus effect size and Cliff’s δ for pairwise contrasts. For the cross-game comparison, where both game and group were included as factors, we applied an Aligned Rank Transform (ART) ANOVA, which is specifically designed for factorial designs with non-normal data.

4. Results

This section presents the empirical findings from our study. We detail the objective performance metrics for both **human participants** and **blind AI agents** across **the two genres of games**, followed by an analysis of the subjective player experience data from the GUESS questionnaires. All comparisons between AI and human players are made against the human performance in the audio-only (blind) condition.

4.1. DareFightingICE

Performance in the DareFightingICE environment was evaluated using two primary quantitative metrics: (I) **Win Ratio**, defined as the proportion of matches won by participants, and (II) average **Health Point (HP) difference**, calculated as the mean disparity in remaining health values across all completed matches against the standardized MCTS opponent. The individual performances of each player are given in the [A](#).

4.1.1. Human Performance Across Modalities

The analysis of human performance in DareFightingICE highlights the impact of expertise and sensory conditions. According to Table 1, player expertise was a significant performance factor, and the removal of vision uniformly decreased player effectiveness. In the **vision** condition, win ratios were high across all groups, ranging from **54.0%** (Somewhat familiar) to **74.0%** (Very familiar). In the **Blind** condition, a clear performance drop was observed for all players. Very familiar players demonstrated the most successful skill transfer, maintaining a **73.0%** win ratio, while somewhat familiar players’ win

ratio fell to **36.0%**. The very familiar and familiar group performed the same in the vision condition; however, the performance of familiar players declined significantly from 74% to 56% in blind mode.

Table 1. *DareFightingICE*: Human Performance by Expertise and Modality

Expertise	Non-vision Mode (Audio-Only)		Vision Mode (Vision + Audio)	
	Win Ratio (%)	HP Difference	Win Ratio (%)	HP Difference
Very familiar	73.0	+60.09	74.0	+69.78
Familiar	56.0	+21.88	74.0	+61.22
Somewhat familiar	36.0	-13.89	54.0	+28.51
Overall Avg.	53.0	+18.13	67.0	+51.31

4.1.2. Human vs. AI Comparison

A comparative analysis between human players and the blind AI provides further insight into performance differences across expertise levels. In Table 2, specifically sections A and B show that the performance of the blind AI in *DareFightingICE* was comparable to the overall average of human players in the blind condition. The overall human average win ratio was **53.0%**. However, a more detailed comparison of Tables 1 and 2 reveals a key insight: both the **very familiar (73.0%)** human player group individually achieved a higher win ratio than the AI’s average performance. The **familiar (56.0%)** group performed similarly to average AI performance across all encoders. Only the somewhat familiar group (36.%) performed worse.

Table 2. Comparison of Performances (A: Blind AI vs Human, B: Blind AI with Sound Design, Winner sound design from the 2022 *DareFightingICE* Sound Design and AI Competition [46].)

A. Blind AI vs Human			B. Blind AI on all Encoders		
Type	Win (%)	HP Diff	Type	Win (%)	HP Diff
Blind AI	54.0	24.64	1D-CNN	54.0	12.66
Human	53.0	18.13	FFT	53.0	27.03
			Mel-spec	56.0	34.23

4.2. *SonicDoom*

Performance in *SonicDoom* was quantified through separate measurements across the scenarios. For the basic scenario, average completion time and average kill count were employed as metrics, given that it is a straightforward task where speed correlates with better performance. In the deadly corridor scenario, we utilized average survival time, average health, average kills, and average deaths as indicators of performance. The shift from average completion time to average survival time was motivated by observations that, even with increased HP in non-vision mode, nearly all participants were unable to reach the green vest at the endpoint. Consequently, the deadly corridor scenario evolved into a survival challenge, where longer survival indicated greater progress. This, along with average kills and deaths, provides nuanced insights into adaptability within non-vision mode. The individual performances of each player are given in the A.

4.2.1. Human Performance Across Modalities

In order to measure the impact of sensory deprivation, we analyzed human performance across both blind and non-blind conditions. It is critical to note that players were given a much larger health pool in the blind "Deadly Corridor" scenario (500 HP) compared to the vision condition (100 HP). Therefore, a direct comparison of raw survival times is misleading; instead, we focus on combat effectiveness as measured by kills and deaths.

As shown in Tables 3 and 4, in the complex "Deadly Corridor" scenario, the loss of vision clearly degraded the combat effectiveness of experienced players. The average kill count for Expert ('Very familiar') players dropped significantly from 4.48 with vision to 3.19 in the blind condition. A similar decline was observed for Intermediate ('Familiar') players, whose kills fell from 3.30 to 1.98. The same was observed for Novice ('Somewhat familiar') players, who achieved more kills in non-blind mode (2.44) than blind mode (0.78).

Table 3. *SonicDoom* Performance Metrics: AI vs. Human (Blind Condition)

Deadly Corridor Scenario				
Player Type	Avg. Survival Time (s) ↑	Avg. Health ↑	Avg. Kills ↑	Avg. Deaths ↓
Very familiar	29.27	9.02	3.19	0.85
Familiar	28.70	-9.33	1.98	1.0
Somewhat familiar	13.17	-13.50	0.78	1.0
Overall	25.57	-2.79	2.21	0.94
Blind AI Agent	5.57	-1.02	2.5	0.92
Basic Scenario				
Player Type	Avg. Completion Time (s) ↓	-	Avg. Kills ↑	-
Very familiar	5.35	-	0.64	-
Familiar	6.11	-	0.62	-
Somewhat familiar	9.25	-	0.37	-
Overall	6.48	-	0.55	-
Blind AI Agent	1.55	-	0.31	-

In the "Basic" scenario, where starting health was not a factor, the impact of removing vision was unambiguous: it uniformly hindered performance as illustrated in Tables 3 and 4. Expert players were over three times slower in the blind condition, with their completion time increasing from 1.56 seconds with vision to 5.35 seconds with audio alone. This trend of significantly slower completion times was consistent across all skill levels. Overall, the findings indicate that while a larger health pool can offset some immediate risks in audio-only combat, the loss of vision demonstrably reduces offensive performance for skilled players and severely hampers navigational efficiency for everyone.

4.3. Human vs. AI Comparison in SonicDoom

A comparative analysis of the AI agent and human participants reveals a stark divergence in strategy and effectiveness, with the AI's performance profile shifting dramatically depending on the availability of visual data and the complexity of the task.

4.3.1. Blind (Audio-Only) Condition Analysis

In the audio-only condition, the data clearly shows that experienced humans significantly outperform the blind AI in complex, high-threat situations. This information is illustrated in Table 3. In the "Deadly Corridor" scenario, the blind AI's performance was critically deficient. It survived for an average of only 5.57 seconds, which is less than a fifth of the survival time of expert ("Very familiar") human players (29.27s). While the AI's kill count (2.5) was higher than that of intermediate ("Familiar") and novice ("Somewhat familiar") players, it was substantially lower than the experts' 3.19 kills. This demonstrates that the AI's reactive audio-based policy is insufficient for navigating dynamic combat, whereas expert humans leverage superior cognitive skills to survive longer and perform more effectively.

Conversely, in the simpler "Basic" scenario, the blind AI showcased its specialized efficiency. The agent completed the level in an average of 1.55 seconds, a time that is over three times faster than the quickest human experts (5.35s). This suggests that the AI has developed a highly optimized policy for less complex tasks but lacks the adaptive and strategic capabilities required for more complex combat scenarios.

Table 4. *SonicDoom* Performance Metrics: AI vs. Human (Non-Blind Condition)

Player Type	Deadly Corridor Scenario			
	Avg. Survival Time (s) ↑	Avg. Health ↑	Avg. Kills ↑	Avg. Deaths ↓
Very familiar	10.23	16.51	4.48	0.58
Familiar	7.61	10.36	3.30	0.79
Somewhat familiar	3.31	-21.60	2.44	0.96
Overall	7.74	5.92	3.09	0.74
Blind AI Agent	1.22	154.78	3.02	0.20

Player Type	Basic Scenario			
	Avg. Completion Time (s) ↓	-	Avg. Kills ↑	-
Very familiar	1.56	-	1.0	-
Familiar	2.69	-	1.0	-
Somewhat familiar	7.81	-	0.80	-
Overall	3.34	-	0.96	-
Blind AI Agent	0.15	-	1.0	-

4.3.2. Non-Blind (Vision) Condition Analysis

When provided with visual data, the AI’s performance surpasses that of all human players, dramatically exceeding their capabilities. This information is illustrated in Table 4. In the **"Deadly Corridor,"** the Blind AI demonstrated near-invincibility. Its average health of **+154.78** is an order of magnitude higher than the expert players’ **+16.51**, and its death rate was a mere **0.20**. Interestingly, expert humans still achieved a higher kill count (**4.48**) compared to the AI (**3.02**), suggesting the AI adopted a strategy of perfect evasion and defense. In contrast, humans employed a more aggressive, high-risk/high-reward approach. Nonetheless, the AI’s ability to preserve its health showcases a level of mastery unattainable by humans.

This superhuman capability was most evident in the **"Basic" scenario**. The Blind AI completed the level in just **0.15 seconds**. This is ten times faster than the expert human players (**1.56s**), highlighting the AI’s profound superiority in reaction speed and execution efficiency when unconstrained by sensory limitations.

4.4. Subjective User Experience (GUESS Scores)

As seen in Table 5, across both games, none of the subscales showed significant group differences after correction. For *SonicDoom*, Usability/Playability came closest to significance ($H(2) = 3.42$, $p = 0.181$, Holm-adjusted $p = 0.724$), but still did not reach the threshold; all other subscales were clearly non-significant. Effect sizes were small ($\eta^2H \leq 0.04$), and Cliff’s δ values were at most small-to-moderate, with wide confidence intervals crossing zero. For *DareFightingICE*, the overall pattern was similar: no significant Kruskal–Wallis tests, with the largest trend in Audio Aesthetics ($H(2) = 3.45$, $p = 0.178$, Holm-adjusted $p = 0.712$). Again, effect sizes were small and pairwise contrasts non-significant.

In the cross-game ART ANOVA, results confirmed that Audio Aesthetics significantly distinguished the two games ($p = 0.003$), with one game being rated higher overall across groups. This information is illustrated in Figure 6. Usability/Playability showed a trend toward a game $\times$ group interaction ($p = 0.093$), suggesting that familiarity might differentially shape perceptions of usability/playability across games, although this did not reach significance. No other main or interaction effects were detected.

In plain terms, participants rated the two games similarly in terms of usability, engrossment and enjoyment, but differed systematically in their evaluation of audio aesthetics. The raw scores and descriptive statistics of GUESS subscales are reported in B.

Table 5. *GUESS* group comparisons per Game and Subscale (Kruskal–Wallis with Holm correction).

Subscale	Kruskal–Wallis		Effect size
	KW H(df)	p (Holm)	eta ² [H] (95% CI)
SonicDoom			
Usability/Playability	2.59 (2)	0.274 (0.822)	0.02 [-0.05, 0.29]
Play Engrossment	0.67 (2)	0.715 (1.000)	-0.04 [-0.06, 0.22]
Enjoyment	0.83 (2)	0.661 (1.000)	-0.03 [-0.06, 0.22]
Audio Aesthetics	3.45 (2)	0.178 (0.712)	0.04 [-0.05, 0.34]
DareFightingICE			
Usability/Playability	3.42 (2)	0.181 (0.724)	0.04 [-0.05, 0.36]
Play Engrossment	0.81 (2)	0.666 (1.000)	-0.03 [-0.06, 0.26]
Enjoyment	1.32 (2)	0.518 (1.000)	-0.02 [-0.06, 0.25]
Audio Aesthetics	0.79 (2)	0.674 (1.000)	-0.04 [-0.06, 0.23]

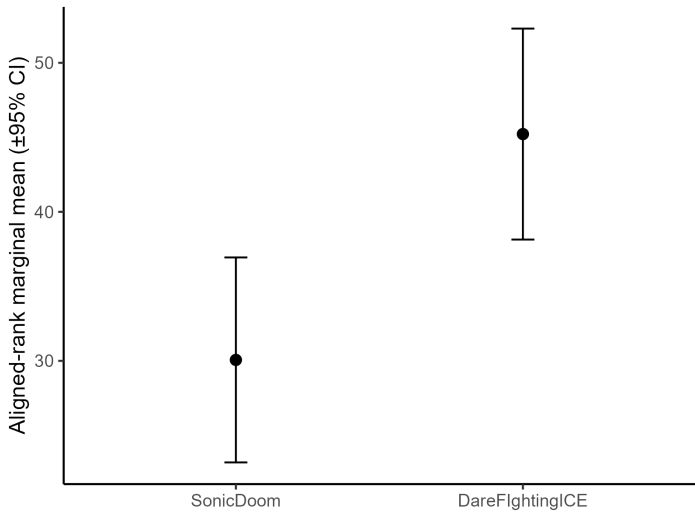


Figure 6. Audio Aesthetics: Game effect from ART model ($p = 0.003$)

5. Discussion

The main contribution of this study is a symmetrical comparison of human expertise transfer and "blind" AI performance in audio-only gaming environments, revealing a nuanced, genre-dependent relationship between cognitive adaptability and computational efficiency. Our findings move beyond the simple question of whether AI can play games with audio, instead interrogating *how* its performance profile diverges from that of human players and what this implies for the future of AI development and accessible game design. The finding is that the superiority of either human or AI is not absolute but is contingent on the nature of the task, specifically the balance between reactive execution and abstract, strategic reasoning.

5.1. The Divergence of Human and AI Expertise

In the **DareFightingICE** setup, highly experienced human players significantly outperformed the reinforcement learning agent. This result suggests that the abstract skills honed through years of visual gameplay, such as understanding spacing, predicting opponent actions, and managing resources, are effectively transferable to the auditory domain. Expert players are not merely reacting to audio cues; they are integrating these cues into pre-existing, sophisticated mental models of fighting game dynamics. This aligns with research on predictive processing in expert gameplay, where top players rely on internal models to anticipate events rather than simply reacting to them [55]. The AI, trained

via PPO on only audio, developed a competent reactive policy but lacked this deeper, predictive understanding, leading to a performance ceiling that seasoned humans could surpass.

Conversely, the **SonicDoom FPS setup** starkly illustrated the AI's brittle but powerful expertise. In the basic scenario, the AI achieved superhuman speed, executing a straightforward task with an efficiency that no human could match. This reflects the core strength of RL: optimizing a policy for a well-defined problem space through millions of iterations [7]. However, this strength became a critical weakness in the complex "Deadly Corridor" scenario. The AI's catastrophic failure to navigate this dynamic, high-threat environment highlights its inability to form a robust cognitive map or engage in the kind of spatial reasoning that humans perform implicitly. While expert humans could leverage sound to navigate, prioritize threats, and survive, the AI's policy collapsed, unable to generalize from its training to a task requiring long-term planning and situational awareness. This finding corroborates recent studies in auditory navigation, which show that standard RL agents struggle to build coherent spatial representations from raw audio streams without specialized architectures, such as those incorporating attention mechanisms or external memory [56].

A noteworthy finding is the performance parity between the blind AI and the overall human average in DareFightingICE, with both achieving a win ratio of approximately 53-54%. This result is significant because it establishes a clear benchmark for AI competence in this audio-only task. It suggests that the AI's natively learned policy is as effective as the general skill set that the average human player can transfer to an audio-only context. However, this average masks the full story; the AI's performance was comparable to the "Familiar" player group but was decisively surpassed by "Very familiar" experts, who achieved a 73% win rate. This implies that while the AI successfully mastered a baseline reactive strategy sufficient to match an average player, it could not replicate the advanced predictive and abstract reasoning that distinguishes expert human performance. The AI learned to react to the sounds, but the experts understood the fight.

5.2. Implications for Accessible Game Design and AI Development

These findings carry significant implications for the design of audio-first games. The success of human players in DareFightingICE underscores the efficacy of a **semantically rich sound design**, where discrete, high-clarity audio cues map directly to specific gameplay events and states. This "semantic sonification" appears more conducive to human learning and skill transfer than the more naturalistic, but also more ambiguous, continuous soundscape of an FPS. For designers aiming to create accessible experiences, particularly for VIPs, this suggests prioritizing clarity and informativeness over pure realism. As our study also showed that novice players struggled significantly in the audio-only condition across both genres, there is a clear need for structured audio-centric tutorials that explicitly teach players how to interpret the game's soundscape.

For AI researchers, our results serve as a benchmark. The performance gap between the AI's success in simple tasks and its failure in complex ones indicates that current "blind" agents possess reactive intelligence but lack abstract reasoning. To create agents that can truly master complex, audio-based environments, future work must move beyond standard policy optimization. Promising avenues include **memory-augmented networks** that can build and maintain an internal state of the environment over time, **curriculum learning** that trains agents on progressively more difficult tasks, and integrating **world models** that allow the agent to "imagine" and plan for future outcomes based on auditory input [57].

5.3. Limitations and Future Directions

This study, while providing a robust framework, has several limitations that open doors for future research. First, we utilized blindfolded sighted participants as a proxy for VIPs, a common methodology in the field. While necessary for early-stage research, VIPs often possess heightened auditory processing skills and unique cognitive strategies. Future studies should aim to include VIPs to validate and extend our findings. Second, our AI was based on a PPO architecture; more advanced

models, such as those leveraging Transformers for audio processing, might yield superior performance and narrow the gap with human experts.

Finally, our investigation was confined to two specific genres. Future work should expand this comparative framework to other domains, such as real-time strategy or puzzle games, which require different cognitive skills, including long-term planning and logical deduction. Such research would further extend the distinct contours of human and artificial intelligence, pushing us closer to creating both more capable AI and more universally accessible virtual worlds. Moreover, future studies can integrate various body signals to strengthen the capacity of AI agents to adapt in real time to the complex interplay of human emotions, behaviors, and physiological states, thereby enabling more personalized and context-aware interactions.

6. Conclusions

Our symmetrical, cross-genre comparison reveals that the performance relationship between human players and AI agents in audio-only environments is fundamentally task-dependent. In scenarios requiring complex, adaptive strategy, such as the *SonicDoom* "Deadly Corridor," the predictive and cognitive skills of expert humans significantly surpassed the reactive policy of the blind AI. Conversely, in tasks demanding optimized navigational efficiency or pure reaction speed, such as the *SonicDoom* "Basic" scenario or when granted full vision, the AI demonstrated superhuman capabilities. This divergence confirms that while visually-honed gaming skills are transferable to the auditory domain, they manifest as a different, more robust form of expertise compared to the AI's specialized but brittle proficiency. Furthermore, the performance parity in *DareFightingICE*, where the blind AI's win rate (54.0%) matched the overall human average (53.0%), serves as a crucial benchmark. It demonstrates that current reinforcement learning methods can produce an agent with a level of competence equivalent to the average human player in a reaction-based audio-only task. Yet, the AI's failure to challenge expert players underscores that its policy, while effective, is less sophisticated than the deeply ingrained, predictive mental models of seasoned humans.

These findings yield actionable insights for both accessible game design and AI research. For developers, our work validates the use of blind AI agents as powerful tools for benchmarking the clarity of discrete audio cues and simple navigation paths, while underscoring the irreplaceable role of expert human testers in validating complex, strategic gameplay. For the AI community, this study highlights the critical need to advance reinforcement learning beyond reactive policies toward models that incorporate predictive processing and internal spatial mapping. Ultimately, bridging the gap between the AI's narrow proficiency and the adaptability of human expertise represents the next frontier for creating truly intelligent agents for complex, real-world interactive tasks [58–60].

Author Contributions: Ibrahim Khan: Contribution: Conceptualization, Methodology, Software, Formal Analysis, Investigation, writing—original draft preparation, Review and Editing, Data Curation, Visualization Ibrahim Khan is a main author responsible for overall coordination and execution of the research. He designed the methodology, developed the software, performed formal analyses, and conducted investigations. He also curated the dataset, created visualizations, and prepared the original draft. Van Thai Nguyen: Contribution: Methodology, Software, Writing—Review and Editing, Formal Analysis Van Thai Nguyen supported the methodological design and validated the results of this research to ensure that the findings were accurately represented. In addition, he contributed to reviewing and editing the manuscript to improve its clarity and quality. Cvetković Tijan Juraj: Contribution: Validation, Writing—Review and Editing, Formal Analysis : Cvetković Tijan Juraj validated the results of this research to ensure that the findings were accurately represented. In addition, Cvetković Tijan Juraj contributed to reviewing and editing the manuscript to improve its clarity and quality. Ruck Thawonmas: Contribution: Conceptualization, Methodology, Resources, Writing—Review and Editing, Supervision, Funding acquisition Ruck Thawonmas supervised the research to ensure it remained on track with academic standards and best practices. He supported the conceptual and methodological design, provided necessary resources for the study. He also offered guidance throughout the research process and contributed to reviewing and editing the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: The study was conducted by the principles of the Declaration of Helsinki and its subsequent amendments, as well as the research guidelines of the American Psychological Association. Ethical committee approval was not required under Japan’s national regulations regarding privacy and informed consent, specifically the Act on the Protection of Personal Information (APPI).

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

Data Availability Statement: The data presented in this study are not publicly available due to ethical and privacy restrictions. The informed consent agreement provided to all participants stipulated that the data collected would be used for research purposes only and would not be shared publicly to protect their privacy. Therefore, the dataset generated and analyzed during the current study cannot be deposited in a public repository.

Acknowledgments: The authors wish to express their sincere gratitude to all the individuals who participated in this study; their time and effort were essential to this research. We would also like to extend our special thanks to Mustafa Can Gursesli for his valuable contributions and insightful feedback throughout the course of this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

1D-CNN	one-dimensional Convolutional Neural Network
AI	artificial intelligence
ANOVA	Analysis of Variance
APPI	Act on the Protection of Personal Information
ART	Aligned Rank Transform
BGM	background music
FFT	Fast Fourier Transform
FPS	first-person shooter
GRU	Gated Recurrent Unit
GUESS	Game User Experience Satisfaction Scale
HP	Health Point
HRTF	Head-Related Transfer Function
MCTS	Monte Carlo Tree Search
Mel	Mel spectrogram
PPO	Proximal Policy Optimization
RL	reinforcement learning
STFT	short-time Fourier transform
VIPs	visually impaired players

Appendix A. Individual Player Performance Data

This appendix provides a detailed breakdown of the performance of each participant across both the *DareFightingICE* and *SonicDoom* experiments. The data is presented by player familiarity ("Very familiar," "Familiar," "Somewhat familiar") and experimental condition.

Appendix A.1. DareFightingICE: Individual Performance Metrics

Table A1. Very Familiar: Blind vs. Non-Blind

Player No.	Blind		Non-Blind	
	Win Rate (%)	HP diff	Win Rate (%)	HP diff
1	78	52.78	78	68.33
2	56	20.78	67	28.89
3	100	151.22	100	219.00
4	78	115.89	89	100.33
5	67	14.56	67	13.11
6	78	93.11	100	145.56
7	89	64.33	100	128.67
8	67	18.67	33	-32.89
9	44	8.67	33	-35.00
10	78	60.89	78	61.78
Average	73	60.09	74	69.78

Table A2. Familiar: Blind vs. Non-Blind

Player No.	Blind		Non-Blind	
	Win Rate (%)	HP diff	Win Rate (%)	HP diff
1	56	6.22	100	130.11
2	67	44.44	78	99.00
3	44	14.11	67	44.78
4	89	83.56	100	89.67
5	44	1.33	67	63.22
6	67	66.56	56	26.89
7	33	-12.67	67	94.00
8	44	21.56	56	24.33
9	78	50.89	78	23.11
10	56	12.78	89	49.44
11	33	-35.67	78	42.11
12	56	-6.89	67	63.56
13	67	38.22	67	45.67
Average	56	21.88	74	61.22

Table A3. Somewhat Familiar: Blind vs. Non-Blind

Player No.	Blind		Non-Blind	
	Win Rate (%)	HP diff	Win Rate (%)	HP diff
1	44	-28.11	73	88.31
2	22	-22.00	89	112.89
3	44	8.44	56	20.89
4	33	-36.44	89	140.22
5	11	-38.44	56	35.67
6	67	38.22	67	45.67
7	11	-40.00	89	147.67
8	56	25.78	56	53.44
9	22	-30.89	40	-50.00
10	44	-17.56	56	30.22
11	22	-27.00	22	-59.78
12	22	-20.89	0	-100.22
13	57	-6.89	33	-32.89
14	44	1.33	33	-32.89
Average	36	-13.89	54	28.51

Appendix A.2. SonicDoom: Individual Performance Metrics

Table A4. Very Familiar: Blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	2.350	1.000	32.340	-8.000	3.670	1.00
2	11.230	1.000	34.730	34.000	5.330	0.67
3	3.410	1.000	27.810	-6.000	3.330	1.00
4	7.950	0.330	18.650	-14.000	1.000	1.00
5	8.820	0.000	12.830	36.000	2.670	0.67
6	8.820	0.000	20.170	-14.000	2.670	1.00
7	4.820	0.670	34.810	-14.000	4.000	1.00
8	5.950	1.000	46.740	10.000	4.330	0.67
9	1.540	1.000	37.800	-12.000	2.670	1.00
10	2.540	1.000	35.800	-12.000	2.670	0.33
11	6.740	0.330	15.070	66.000	3.670	0.67
12	4.393	0.593	30.217	15.055	3.021	1.00
13	4.152	0.581	30.454	16.564	2.979	0.67
14	3.912	0.569	30.691	18.073	2.937	1.00
15	3.672	0.557	30.928	19.582	2.895	0.85
Average	5.353	0.642	29.269	9.018	3.189	0.85

Table A5. Familiar: Blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	8.820	0.000	12.150	-10.000	0.330	1.00
2	3.720	1.000	30.430	-12.000	4.670	1.00
3	7.950	0.333	18.650	-14.000	1.000	1.00
4	5.390	1.000	13.120	-16.000	2.000	1.00
5	5.340	0.667	36.820	-20.000	2.000	1.00
6	5.950	1.000	36.740	10.000	4.330	1.00
7	8.950	0.333	12.650	-14.000	1.000	1.00
8	5.080	0.333	24.550	-10.000	0.330	1.00
9	5.810	0.667	39.700	-10.000	2.670	1.00
10	5.884	0.648	35.605	-8.000	1.926	1.00
11	5.794	0.659	37.508	-7.467	1.904	1.00
12	5.704	0.670	39.411	-6.933	1.882	1.00
13	5.614	0.681	31.314	-6.400	1.859	1.00
14	5.524	0.693	33.217	-5.867	1.837	1.00
Average	6.109	0.620	28.705	-9.333	1.981	1.00

Table A6. Somewhat Familiar: Blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	3.880	0.670	23.880	-14.000	2.330	1.00
2	5.290	1.000	16.480	-22.000	2.000	1.00
3	7.410	0.330	18.130	-12.000	2.670	1.00
4	8.250	0.330	14.760	-14.000	0.330	1.00
5	10.015	0.160	11.885	-13.000	0.500	1.00
6	11.538	0.000	9.314	-12.000	0.000	1.00
7	13.061	0.172	6.743	-11.000	0.544	1.00
8	14.584	0.340	4.172	-10.000	1.072	1.00
Average	9.254	0.375	13.171	-13.500	1.181	1.00

Table A7. Very Familiar: Non-blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	1.120	1.000	10.310	34.000	5.330	0.33
2	1.270	1.000	4.810	-16.000	3.330	1.00
3	1.410	1.000	11.660	22.000	6.000	0.00
4	4.570	1.000	3.110	-12.000	1.000	1.00
5	1.000	1.000	7.860	24.000	5.000	0.33
6	1.370	1.000	8.780	-12.000	3.670	1.00
7	2.230	1.000	8.810	2.000	3.330	0.67
8	1.960	1.000	7.140	-12.000	3.000	0.00
9	0.950	1.000	11.470	46.000	6.000	1.00
10	0.950	1.000	15.800	56.000	6.000	0.67
11	1.390	1.000	12.130	8.000	4.670	0.33
12	1.367	1.000	19.660	24.073	4.832	0.33
13	1.319	1.000	11.763	25.964	4.921	1.00
14	1.271	1.000	9.957	27.855	5.009	0.00
15	1.223	1.000	10.150	29.745	5.097	1.00
Average	1.560	1.000	10.227	16.509	4.479	0.58

Table A8. Familiar: Non-blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	2.590	1.000	14.240	-4.000	4.670	0.67
2	1.480	1.000	8.020	-12.000	4.330	1.00
3	4.570	1.000	3.110	-12.000	1.000	1.00
4	1.390	1.000	3.120	6.000	3.670	0.67
5	5.820	1.000	3.280	-10.000	1.000	1.00
6	1.960	1.000	7.140	-12.000	3.000	1.00
7	4.570	1.000	3.110	-12.000	1.000	0.33
8	1.750	1.000	13.130	24.000	5.000	0.33
9	1.390	1.000	10.970	34.000	5.000	0.67
10	2.551	1.000	7.869	20.389	3.407	0.67
11	2.494	1.000	7.974	24.422	3.452	1.00
12	2.437	1.000	8.078	28.456	3.496	1.00
13	2.380	1.000	8.183	32.489	3.540	0.67
14	2.323	1.000	8.287	36.522	3.585	1.00
Average	2.690	1.000	7.610	10.310	3.300	0.79

Table A9. Somewhat Familiar: Non-blind Mode

Player No.	Basic		Deadly Corridor			
	Time Taken	Enemies Killed	Time Taken	Health	Enemies Killed	Deaths
1	1.580	1.000	7.810	20.000	5.000	1.00
2	1.260	1.000	10.870	16.000	4.670	1.00
3	1.780	1.000	6.220	-4.000	2.330	1.00
4	8.820	0.000	3.200	-16.000	0.670	1.00
5	8.920	1.000	2.405	-28.000	0.665	0.67
6	11.144	1.000	0.557	-40.800	2.198	1.00
7	13.368	0.733	1.291	-53.600	1.731	1.00
8	15.592	0.705	3.139	-66.400	2.264	1.00
Average	7.810	0.800	4.440	-21.600	2.440	0.96

Insights from Individual Performance

A closer examination of the individual player data reveals several key insights into skill transfer, adaptability, and the inherent variability of player performance. The averages presented in the main body of the study are clarified by these individual results, highlighting specific trends and noteworthy outliers.

Expertise Directly Correlates with Adaptability in *DareFightingICE*

- **Resilience of Experts:** The “Very familiar” group’s average win rate barely changed between conditions (74% Non-Blind vs. 73% Blind). Player 3 exemplifies this, achieving a perfect 100% win rate in both modes, indicating a near-perfect transfer of fighting game fundamentals to the auditory domain.
- **Performance Cliff for Non-Experts:** In contrast, the “Familiar” group’s average win rate dropped sharply from 74% to 56% when deprived of vision. The “Somewhat familiar” group saw their performance decline from a winning average (54%) to a losing one (36%), with their average HP difference shifting from +28.51 to -13.89. This suggests that while basic competence can be achieved with vision, adapting to audio-only gameplay is a distinct skill that relies heavily on deep-seated expertise.
- **Anomalous Performances:** Interestingly, two players in the “Very familiar” group (Players 8 and 9) achieved a higher win rate in the blind condition than in the non-blind condition. This suggests

that for some individuals, focusing solely on audio cues may be more effective than processing combined audio-visual information, potentially reducing cognitive load or eliminating visual distractions.

Visual Deprivation Severely Hampers Navigational and Combat Efficiency in *SonicDoom*

- **Drastic Slowdown in Simple Tasks:** In the “Basic” scenario, completion times increased dramatically for all groups in blind mode. “Very familiar” players, who averaged a lightning-fast 1.56 seconds with vision, slowed to 5.35 seconds when relying on audio alone. This highlights a severe reduction in target acquisition and movement speed.
- **Reduced Combat Effectiveness:** While survival times in the “Deadly Corridor” were longer in blind mode, this is a reflection of the increased health pool provided in the experiment. A more accurate measure of performance, the average number of enemies killed, decreased across the board. “Very familiar” players’ kill count dropped from 4.48 to 3.19, and “Familiar” players’ from 3.30 to 1.98. This demonstrates that players were less capable of engaging threats effectively without visual cues.

High Degree of Performance Variability Within Experience Groups

- **Standout Performers and Struggling Players:** In the “Somewhat familiar” *DareFightingICE* group, Player 6 performed exceptionally well, achieving a 67% win rate in both conditions—on par with the “Familiar” group average. Conversely, Player 12 from the same group had a 0% win rate in the non-blind condition, and Player 9 struggled with a negative HP difference even with vision.
- **Navigational Disparities in *SonicDoom*:** The blind “Deadly Corridor” scenario revealed a wide gap in navigational aptitude. Within the “Somewhat familiar” group, Player 1 survived for an average of 23.88 seconds, approaching the average of more experienced groups, while Player 8 survived for only 4.17 seconds. This indicates that the ability to interpret spatial audio for navigation and survival varies greatly from person to person, independent of their general FPS experience.

Appendix B. GUESS Scores

The raw scores and descriptive statistics of GUESS subscales are reported below.

Table A1. Descriptive statistics of GUESS subscale scores (all participants).

subscale	n	Mean ± SD	Median [IQR]	Min–Max
Usability/Playability	74	15.18 ± 3.34	15.00 [13.00–18.00]	4.00–21.00
Play Engrossment	74	13.65 ± 3.49	14.00 [11.00–16.00]	6.00–21.00
Enjoyment	74	14.74 ± 3.78	15.00 [13.00–18.00]	4.00–21.00
Audio Aesthetics	74	16.22 ± 2.73	17.00 [15.00–18.00]	7.00–21.00

Table A2. Descriptive statistics of GUESS subscales by game and group (Median [IQR]).

Game	Subscale	VeryFamiliar	Familiar	SomewhatFamiliar
SonicDoom	Usability/Playability	15.50 [11.00–17.00]	14.00 [13.00–16.00]	16.00 [15.00–18.00]
SonicDoom	Play Engrossment	14.00 [12.50–15.00]	12.50 [10.25–16.00]	12.00 [10.00–15.00]
SonicDoom	Enjoyment	16.50 [15.25–18.00]	15.50 [14.00–17.75]	16.00 [13.00–18.00]
SonicDoom	Audio Aesthetics	16.00 [15.00–18.00]	14.50 [13.25–15.75]	16.00 [15.00–18.00]
DareFightingICE	Usability/Playability	17.00 [14.50–19.00]	15.00 [13.00–17.50]	13.00 [13.00–14.25]
DareFightingICE	Play Engrossment	15.00 [12.00–17.00]	13.00 [12.25–14.00]	16.00 [11.00–18.25]
DareFightingICE	Enjoyment	15.00 [13.50–19.00]	15.00 [10.00–16.00]	15.00 [14.25–15.75]
DareFightingICE	Audio Aesthetics	17.00 [16.00–20.00]	17.00 [16.25–18.00]	18.00 [15.00–18.00]

References

- Collins, K. *From Pac-Man to pop music: interactive audio in games and new media*; Routledge, 2017.
- Gursesli, M.C.; Tarchi, P.; Guazzini, A.; Duradoni, M.; Yaras, A.T.; Akinci, E.; Yurdakul, U.; Calà, F.; Tonacci, A.; Vilone, D.; et al. Silent vs. Sound: The Impact of Uniform Auditory Stimuli on Eye Movements and Game Performance. In Proceedings of the 2025 IEEE Gaming, Entertainment, and Media Conference (GEM). IEEE, 2025, pp. 1–6.
- Andrade, R.; Rogerson, M.J.; Waycott, J.; Baker, S.; Vetere, F. Playing Blind: Revealing the World of Gamers with Visual Impairment. In Proceedings of the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–14. <https://doi.org/10.1145/3290605.3300346>.
- Donohue, S.E.; Woldorff, M.G.; Mitroff, S.R. Video game players show more precise multisensory temporal processing abilities. *Attention, perception, & psychophysics* **2010**, *72*, 1120–1129.
- Massiceti, D.; Hicks, S.L.; van Rheede, J.J. Stereoscopic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PloS one* **2018**, *13*, e0199389.
- Connors, E.C.; Yazzolino, L.A.; Sánchez, J.; Merabet, L.B. Development of an audio-based virtual gaming environment to assist with navigation skills in the blind. *Journal of visualized experiments: JoVE* **2013**, ., 50272.
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533.
- Vinyals, O.; Babuschkin, I.; Czarnecki, W.M.; et al.. Grandmaster level in *StarCraft II* using multi-agent reinforcement learning. *Nature* **2019**, *575*, 350–354. <https://doi.org/10.1038/s41586-019-1724-z>.
- Van Nguyen, T.; Dai, X.; Khan, I.; Thawonmas, R.; Pham, H.V. A deep reinforcement learning blind AI in DareFightingICE. In Proceedings of the 2022 IEEE Conference on Games (CoG). IEEE, 2022, pp. 632–637.
- Lake, B.M.; Ullman, T.D.; Tenenbaum, J.B.; Gershman, S.J. Building machines that learn and think like people. *Behavioral and Brain Sciences* **2017**, *40*, e253. <https://doi.org/10.1017/S0140525X16001837>.
- Gursesli, M.C.; Lombardi, S.; Duradoni, M.; Bocchi, L.; Guazzini, A.; Lanata, A. Facial emotion recognition (FER) through custom lightweight CNN model: performance evaluation in public datasets. *IEEE Access* **2024**, *12*, 45543–45559.
- Khan, I.; Van Nguyen, T.; Dai, X.; Thawonmas, R. Darefightingice competition: A fighting game sound design and ai competition. In Proceedings of the 2022 IEEE Conference on Games (CoG). IEEE, 2022, pp. 478–485.
- Kempka, M.; Wydmuch, M.; Runc, G.; Toczek, J.; Jaśkowski, W. Vizdoom: A doom-based ai research platform for visual reinforcement learning. In Proceedings of the 2016 IEEE conference on computational intelligence and games (CIG). IEEE, 2016, pp. 1–8.
- Smuts, A. Are video games art? *Contemporary Aesthetics (Journal Archive)* **2005**, *3*, 6.
- Gursesli, M.C.; Martucci, A.; Mattiassi, A.D.; Duradoni, M.; Guazzini, A. Development and validation of the psychological motivations for playing video games scale (PMPVGs). *Simulation & Gaming* **2024**, *55*, 856–885.
- Gursesli, M.C.; Guazzini, A.; Thawonmas, R.; Valenti, C.; Duradoni, M.; Thawonmas, R. Internet Gaming Disorder and Psychological Distress: a PRISMA systematic review. *Heliyon* **2025**, *11*.
- Bostan, B. Player motivations: A psychological perspective. *Computers in Entertainment (CIE)* **2009**, *7*, 1–26.
- Przybylski, A.K.; Rigby, C.S.; Ryan, R.M. A motivational model of video game engagement. *Review of general psychology* **2010**, *14*, 154–166.
- Thiparpakul, P.; Mokekhaow, S.; Supabanpot, K. How can video game atmosphere affect audience emotion with sound. In Proceedings of the 2021 9th International Conference on Information and Education Technology (ICIET). IEEE, 2021, pp. 480–484.
- Guillen, G.; Jylha, H.; Hassan, L. The role sound plays in games: a thematic literature study on immersion, inclusivity and accessibility in game sound research. In Proceedings of the Proceedings of the 24th International Academic Mindtrek Conference, 2021, pp. 12–20.
- Paterson, N.; Naliuka, K.; Jensen, S.K.; Carrigy, T.; Haahr, M.; Conway, F. Spatial audio and reverberation in an augmented reality game sound design. In Proceedings of the 40th AES Conference: Spatial Audio, Tokyo, Japan, October 8-10, 2010.
- Haehn, L.; Schlittmeier, S.J.; Böffel, C. Exploring the impact of ambient and character sounds on player experience in video games. *Applied Sciences* **2024**, *14*, 583.
- Nacke, L.E.; Grimshaw, M.N.; Lindley, C.A. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with computers* **2010**, *22*, 336–343.

24. Broderick, J.; Duggan, J.; Redfern, S. The importance of spatial audio in modern games and virtual environments. In Proceedings of the 2018 IEEE Games, Entertainment, Media Conference (GEM). IEEE, 2018, pp. 1–9.
25. Semionov, K.; McGregor, I. Effect of various spatial auditory cues on the perception of threat in a first-person shooter video game. In Proceedings of the Proceedings of the 15th International Audio Mostly Conference, 2020, pp. 22–29.
26. Sadowska, D.; Sacewicz, T.; Rebiś, K.; Kowalski, T.; Krzepota, J. Examining physiological changes during counter-strike: Global offensive (CS: GO) performance in recreational male esports players. *Applied Sciences* **2023**, *13*, 11526.
27. Ng, P.; Nesbitt, K.; Blackmore, K. Sound improves player performance in a multiplayer online battle arena game. In Proceedings of the Australasian Conference on Artificial Life and Computational Intelligence. Springer, 2015, pp. 166–174.
28. Ng, P.; Nesbitt, K. Informative sound design in video games. In Proceedings of the Proceedings of the 9th Australasian conference on interactive entertainment: matters of life and death, 2013, pp. 1–9.
29. Schaffert, N.; Mattes, K.; Effenberg, A.O. A sound design for acoustic feedback in elite sports. In Proceedings of the International Symposium on Computer Music Modeling and Retrieval. Springer, 2009, pp. 143–165.
30. Bandai Namco Entertainment. Tekken 7. Bandai Namco Entertainment, Minato, Tokyo, 2015.
31. NetherRealm Studios. Mortal Kombat 11. NetherRealm Studios, Chicago, Illinois, US, 2019.
32. Double Helix Games. Killer Instinct. Double Helix Games, Irvine, California, US, 2013.
33. Khan, I.; Van Nguyen, T.; Thawonmas, R. Fighting to the beat: Multi-instrumental adaptive background music approach. *Entertainment Computing* **2025**, p. 100985.
34. Agrimi, E.; Battaglini, C.; Bottari, D.; Gnecco, G.; Leporini, B. Game accessibility for visually impaired people: a review. *Soft Computing* **2024**, *28*, 10475–10489.
35. Games, G. Shades of Doom, 2001. Accessed: 2025-03-14.
36. AGRIP. AudioQuake, 2003. Accessed: 2025-03-14.
37. the Terraformers Team. Terraformers, 2003. Accessed: 2025-03-14.
38. Nair, V.; Karp, J.L.; Silverman, S.; Kalra, M.; Lehv, H.; Jamil, F.; Smith, B.A. Navstick: Making video games blind-accessible via the ability to look around. In Proceedings of the The 34th Annual ACM Symposium on User Interface Software and Technology, 2021, pp. 538–551.
39. Nair, V.; Zhu, H.; Song, P.; Wang, J.; Smith, B.A. Surveyor: Facilitating Discovery Within Video Games for Blind and Low Vision Players. In Proceedings of the Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–15.
40. Marples, D.; Gledhill, D.; Carter, P. The effect of lighting, landmarks and auditory cues on human performance in navigating a virtual maze. In Proceedings of the Symposium on Interactive 3D Graphics and Games, 2020, pp. 1–9.
41. Tan, S.L.; Baxa, J.; Spackman, M. The role of music in video games. In *Playing video games: Motives, responses, and consequences*; Routledge, 2010; pp. 141–157.
42. Nacke, L.E.; Grimshaw, M.N.; Lindley, C.A. The effects of sound and graphics on immersion in video games. In Proceedings of the Proceedings of the Audio Mostly Conference, 2010, pp. 1–7.
43. Mirza, B.; Hoque, M.E.; Mahmud, H. The impact of audio-visual stimuli on gaming performance. *Entertainment Computing* **2019**, *31*, 100308.
44. Lipscomb, S.D.; Zehnder, S.M. Interactive music: Crossing boundaries. In Proceedings of the Proceedings of the 2005 International Computer Music Conference, 2005, pp. 747–750.
45. Khan, I.; Van Nguyen, T.; Gursesli, M.C.; Thawonmas, R. Sonic Doom: Enhanced Sound Design and Accessibility in a First-Person Shooter Game. In Proceedings of the 2025 IEEE Conference on Games (CoG). IEEE, 2025, pp. 1–8.
46. Organizers, D.C. The DareFightingICE Sound Design and AI Competition Winners details, R22. <https://www.ice.ci.ritsumei.ac.jp/~ftgaic/index-R22.html>, 2022. Results of a competition held at the **IEEE Conference on Games (CoG) 2022**. Accessed: September 29, 2025.
47. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms, 2017, [[arXiv:cs.LG/1707.06347](https://arxiv.org/abs/1707.06347)].
48. Petrenko, A.; Huang, Z.; Kumar, T.; Sukhatme, G.; Koltun, V. Sample factory: Egocentric 3d control from pixels at 100000 fps with asynchronous reinforcement learning. In Proceedings of the International Conference on Machine Learning. PMLR, 2020, pp. 7652–7662.

49. Ishihara, M.; Miyazaki, T.; Chu, C.Y.; Harada, T.; Thawonmas, R. Applying and improving Monte-Carlo Tree Search in a fighting game AI. In Proceedings of the Proceedings of the 13th international conference on advances in computer entertainment technology, 2016, pp. 1–6.
50. Cheiran, J.F.; Nedel, L.; Pimenta, M.S. Inclusive games: a multimodal experience for blind players. In Proceedings of the 2011 Brazilian Symposium on Games and Digital Entertainment. IEEE, 2011, pp. 164–172.
51. Smith, B.A.; Nayar, S.K. The RAD: Making racing games equivalently accessible to people who are blind. In Proceedings of the Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–12.
52. Zeng, L.; Weber, G. A pilot study of collaborative accessibility: How blind people find an entrance. In Proceedings of the Proceedings of the 17th international conference on human-computer interaction with mobile devices and services, 2015, pp. 347–356.
53. 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.
54. Ruxton, G.D.; Beauchamp, G. Time for some a priori thinking about post hoc testing. *Behavioral ecology* **2008**, *19*, 690–693.
55. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **2013**, *36*, 181–204.
56. Chen, C.; Schissler, C.; Garg, S.; Kobernik, P.; Clegg, A.; Calamia, P.; Batra, D.; Robinson, P.; Grauman, K. Soundspaces 2.0: A simulation platform for visual-acoustic learning. In Proceedings of the ., 2022, Vol. 35, pp. 8896–8911.
57. Ha, D.; Schmidhuber, J. World models. *arXiv preprint arXiv:1803.10122* **2018**, *2*.
58. Hunicke, R. The case for dynamic difficulty adjustment in games. In Proceedings of the Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, 2005, pp. 429–433.
59. Gursesli, M.C.; Selek, M.E.; Samur, M.O.; Duradoni, M.; Park, K.; Guazzini, A.; Lanatà, A. Design of cloud-based real-time eye-tracking monitoring and storage system. *Algorithms* **2023**, *16*, 355.
60. Gursesli, M.C.; Calà, F.; Tarchi, P.; Frassinetti, L.; Guazzini, A.; Lanatà, A. Eyetracking correlated in the matching pairs game. In Proceedings of the 2023 IEEE Gaming, Entertainment, and Media Conference (GEM). IEEE, 2023, pp. 1–6.

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.