UNIARTS HELSINKI

SIBELIUS ACADEMY

Shaping Non-Linear Soundscapes:

Creative and Technical Approaches for Game Audio Mixing

Master's Thesis
7.4.2025
Mikko Saarinen
Department of Music Technology
University of the Arts Helsinki

UNIARTS HELSINKI

SIBELIUS ACADEMY

Title	Number of Pages
Shaping Non-Linear Soundscapes: Creative and Technical Approaches for Game Audio Mixing	61

Author	Semester
Mikko Saarinen	Spring 2025

Department

Music Technology

Abstract

This thesis examines the audio mixing process for video games and is aimed at anyone interested in the field of interactive audio. The research focuses on the creation of a mix that aligns with a game's core design and artistic vision while meeting today's technical standards and requirements.

Written from a sound designer's perspective, the thesis explores workflows, tools, and techniques for mixing in non-linear media, offering a practical and holistic lens that considers the technical, artistic, and collaborative aspects of interactive audio. This study examines how a well-crafted mix enhances gameplay clarity, immersion, and emotional impact, ensuring that players receive essential auditory feedback while experiencing a compelling soundscape.

A significant focus is placed on practical implementation, including the role of audio middleware, real-time mix adjustments, and technologies such as object-based audio and 3D spatialization. Case studies of existing games, along with reflections on personal industry experience, provide real-world examples of interactive mixing solutions. By bridging academic research and industry practice, this thesis serves as both an educational resource and a reference for professionals, offering insights into the evolving landscape of game audio mixing in modern interactive media.

Keywords

computer games, mixing, audio design, game music, middleware, interactive art

This work has been checked in the plagiarism detection program Turnitin

Glossary

Ambience	Background environmental sounds that set the mood and sense of place.
Attenuation	The reduction of sound volume over distance.
Audio Object	A sound source with spatial metadata for 3D audio rendering.
Binaural Rendering	3D sound processing that simulates spatial hearing over headphones.
Bus	A signal path that groups and processes multiple audio signals.
Ducking	Lowering the volume of one sound when another plays.
Dynamic Range	The span between the softest and loudest audio elements.
Foley	Sounds that represent character movement, clothes, gear, etc.
Game Engine	The core software environment in which games are developed.
HDR	High Dynamic Range audio system that prioritizes important sounds dynamically.
HRTF	Head-Related Transfer Function; used for realistic spatial audio.
Implementation	The process of integrating audio assets into the game.
LUFS	Loudness Units relative to Full Scale; a measurement of perceived loudness.
Ludic	Audio serving a gameplay function, such as feedback or cues.
Mastering	Finalizing and optimizing the full game mix for playback.
Middleware	Software that connects sound assets to the game engine with interactive functionality.
Mixing	The process of balancing, shaping, and organizing audio elements within a game.
Non-linear Media	Media that doesn't follow a fixed timeline, e.g. games.
Occlusion	The dampening of sound caused by obstacles between source and listener.
RTPC	A system that changes audio parameters in real time based on game data.
Sidechaining	A mixing technique where one audio signal controls another.
Soundscape	The combined sonic environment of a game.
Spatialization	The positioning of sound in 3D space.
Voice Count	The number of simultaneous sounds the audio engine plays.

1. Preface

Throughout my life as a sound designer, composer and musician, I've always found the mixing stage of any type of media exciting and fascinating. Whether it's a piece of music, a short film or a video game, this stage of production is where all elements come together, forming something greater than the sum of their parts. To me this applies especially to video games, where the end result is an intricate blend of systems, visuals, and sounds, all interacting in real time. At its best, these technologies combine to deliver an unforgettable gameplay experience for the player.

So far in my game audio journey I've had the opportunity to work on mobile, console, PC and VR games. I've gained experience with different tools such as game engines, audio middleware, and I've collaborated with development teams ranging from a few individuals to large groups. I have developed a mental image of what game audio production entails from start to finish.

The research question guiding this thesis is: how to create a mix that aligns with a game's core design and artistic vision while also meeting today's technical standards and requirements. As quickly became apparent, there is no single correct answer to this question. Therefore, I have approached the subject from multiple angles—technical, artistic, and game design—seeking to provide a broad perspective on the game audio mixing process.

My motivation for writing this thesis is to learn more about the craft of game audio mixing, by research and reflection. I've gathered information by reading existing academic literature, articles and interviews, attended conferences and watched recordings of presentations by professionals. I use existing games as examples and analyze their soundscapes from the perspective of mixing. By reflecting on my own work experiences, I share practical examples of solutions and shed light on the reasoning behind decisions. I hope this work provides useful information and inspiration for anyone interested in the realm of game audio.

2. The Role of Sound in Video Games

Video games contain various types of sound elements—such as music, foley, ambience, UI (user interface) sounds, and dialogue—that together form the overall soundscape. The sounds in a game serve different purposes. Video games are interactive, meaning that the user engages in direct interaction with the game. At its core, the role of game audio is to provide auditory feedback based on the user's actions. The human brain processes sound faster than visual stimuli, making sound an effective way to communicate game events to the user (Aditya et al. 2015).

2.1. Generative & Procedural

Simplified, game audio can be either generative or procedural. In the generative approach, the audio engine creates a soundscape by playing individual sound files and modulating different parameters, such as pitch, volume, playback speed, and panning, according to predetermined rules and logic. Procedural sound, on the other hand, is not tied to pre-recorded sound files but is synthesized in real time. This approach allows for more dynamic and responsive sounds that adapt precisely to game events. Today, games often use a combination of these techniques. (Farnell 2008, 2)

Each sound file is attached to an audio event, which is then triggered by the game engine. For instance, a footstep sound is called whenever the player character's walking animation reaches the frame where their foot touches the ground.

2.2. Adaptivity

A typical characteristic of video game audio is adaptivity, which refers to how the soundscape adjusts to the player's actions. A simple example of this is when the player presses a jump button on the controller, and the character in the game jumps with an accompanying sound effect. In theory, this can happen at any point during gameplay and as many times as the player desires. Games are therefore nonlinear, unlike movies, which play out in a predetermined sequence from start to finish during every viewing.

In addition to adaptivity, video game audio can also be dynamic, meaning that the soundscape evolves and reacts to the game's environment or conditions. For instance, the game might activate a low-pass filter on the whole soundscape whenever the player character dives underwater. This real-time signal processing creates a muffled underwater sound effect. In some shooter games, when the player's weapon magazine is running low, the volume, pitch and timbre of gunfire sounds may subtly change. This allows the player to time their reloads based solely on auditory cues. Similarly, the tempo of the game's music may increase as the timer for a given task runs out, creating a sense of urgency or even panic in the player. Studies have shown that players who play without sound have a lower pulse rate than those who play with sound (Gouveja 2013, 13). For example, the anxiety-inducing and oppressive music and sound design of *Bloodborne* (From Software 2015) has even impaired my own performance in the game.

2.3. Narrative & Ludic

Video games often take place in fictional worlds with their own history, culture, and geography. The sound designer plays a crucial role as a storyteller, using narrative sound design to enhance the illusion of a living virtual world. These sounds reinforce world-building, character development, and thematic elements. For example, the sound of a gate opening can reveal the time period in which the game is set or the type of technology present in the world. Is the gate freshly maintained or old and worn? Does it sound ominous or welcoming?

In addition to its narrative role, game audio also serves a ludic function by providing essential feedback and cues about gameplay events. Sounds that offer useful information to the player—such as guiding movement, signaling enemy presence, or confirming successful actions—fall into this category (Zlobin 2021). For instance, in a stealth game, enemy footsteps growing louder can indicate danger, while a distinct chime might confirm a completed objective.

Thus, game audio operates on both narrative and ludic levels, helping players navigate a 3D space, reinforcing the story, setting the mood, and enhancing the overall believability of the game world. To be most effective, audio content must seamlessly integrate with the game's visual style, genre, and mechanics.

2.4. The Mix

The audio mix plays a pivotal role in how all of the audio information is presented. Where should the player's attention be at any given time? For instance, a narrative beat might be led by dialogue, while an action sequence could require the player to focus on the location of enemies or other threats. At other times, environmental sounds and music may immerse the player further into the game's world and story. At its best, through a well-crafted audio mix, a game's soundscape seamlessly supports both gameplay and immersion. The term immersion refers to the player's mental state in which their awareness of the surrounding reality fades, and they become deeply absorbed in the virtual experience (Agrawal et al. 2019, 1).

3. Principles of Game Audio

3.1. Game Engine

A video game is a complex software system that integrates various components such as code, graphics, sound, physics simulation, artificial intelligence, and animation. These elements are typically designed to provide the user with a satisfying and entertaining gaming experience. The game engine, in turn, is a development environment that brings these components together and offers numerous features to support the game development process. Today, some of the most popular game engines include Unity (Unity Technologies) and Unreal Engine (Epic Games). (Toftedahl & Engström 2019)

Game engines are customizable to meet the needs of game developers. Some game studios use entirely proprietary, custom-built software. For example, the Finnish company Remedy Entertainment utilizes its own Northlight game engine. While game engines are generally capable of handling audio data, there are audio middleware solutions specifically designed to specialize in the functionality of game audio. As video game productions grow increasingly complex and larger in scale, the demands on audio production also continue to rise.

3.2. Audio Middleware

Audio middleware provides sound designers with a graphical user interface that enables the creation of complex audio functionalities without text-based programming. These functionalities include managing and playing audio files, defining behavior logic, supporting various file formats, streaming, simulating 3D sound, real-time digital signal processing, and even low-level audio programming (Gouveia 2013, 31). Both tools allow sound designers to categorize sounds, route them into mixer buses, and use a graphical mixing desk to control the entire soundscape. These mixing desks enable designers to adjust audio levels, apply effects, and manage the balance of the soundscape in a highly visual and intuitive way. This functionality supports middleware's role as an external audio engine by giving sound designers precise control over how audio is managed and delivered in real time.

The role of audio middleware is to provide sound designers with efficient tools and to function as an external audio engine that follows the commands given by the game engine. For instance, when the foot of a walking game character touches the ground, the game engine sends a command to the audio middleware to play a footstep sound. The command may also include additional information, such as the type of surface the character is walking on, the character's position in 3D space relative to the audio listener, the environment in which the character is moving, and the speed at which the character is moving. Using this information, the middleware can play the desired sound and, if necessary, spatialize it—that is, simulate how the sound resonates in a three-dimensional space.

The most widely used audio middleware in game development are Wwise (WaveWorks Interactive Sound Engine) by Audiokinetic and FMOD Studio by Firelight Technologies. Both offer real-time mixing, DSP effects, spatial processing, and advanced interactive music systems. These middleware solutions integrate into game engines as plugins and are optimized for multiple gaming platforms, including PC, PlayStation, Xbox, and Nintendo. While Wwise and FMOD serve the same fundamental purpose, they differ in their graphical user interfaces and workflow structures.

Although it is entirely possible to implement complex audio systems without middleware, doing so requires extensive programming expertise. Middleware provides a more accessible and flexible framework for sound designers to implement and manage audio without relying solely on code. While the underlying principles remain the same, this text assumes that all workflow examples are based on the use of audio middleware.

4. Roles in Game Audio

4.1. The Audio Team

The roles of audio professionals in game development can range from highly specialized to broadly generalized, depending on the scope of the project and the size of the team. In smaller teams, a single individual may handle all aspects of audio, including creating sound effects, composing music, managing implementation, and delivering the final mix.

In contrast, large-scale productions, such as so-called Triple-A (AAA) games, often involve extensive audio teams with highly specialized roles. These teams may include an audio director, audio lead, sound designers, technical sound designers, audio implementers, audio programmers, composers, music editors, dialogue designers, dialogue coordinators, dialogue supervisors, foley editors, and localization specialists, among others.

4.2. The Mix Owner

Interestingly, the role of a dedicated mixing engineer is still uncommon in game development. Some industry professionals argue that mixing is best handled as a collective effort, where a mix supervisor oversees the process rather than a single person handling all of the hands-on mixing (Riviere 2023, 14). Currently, the responsibility for the final mix most often falls to the audio lead.

Still, introducing this role could provide a fresh perspective, as a new pair of ears might bring valuable insights to the overall mix. Games are long-term projects that often span several years, and those working closely on the audio throughout this time can become too attached or desensitized to certain aspects of the soundscape. Incorporating a specialized mixing engineer could enhance the process and result in a more cohesive audio experience. That said, the complexity of modern games presents significant challenges for an "outsider mixing engineer" joining at the end of a project.

Games are built on intricate layers of interconnected systems developed over years, and understanding the full scope of these systems can be a daunting task—not to mention the thousands of sound effects, dialogue, and music that are added and iterated upon throughout the

multi-year production (Riviere 2023, 1). However, this issue could be mitigated by involving the mixing engineer early enough in the project, allowing them to grasp the overall audio vision and the inner workings of the game. As technology and tools evolve, the roles within game audio development may also shift. This progression could steer the process more toward system and content curation rather than traditional sound creation. Such changes might open the door for new specialized roles, such as a designated game mixing engineer, who focuses solely on balancing and refining the interactive soundscape.

5. Linear & Non-Linear

5.1. Traditional Linear Workflows in Film Sound

In this section, I describe the fundamental differences between linear and non-linear media—both in the final product and the production process—focusing on the perspective of mixing. While there are undeniable similarities between film and game sound production, my goal is to highlight the key differences. Understanding these distinctions is crucial for adopting a non-linear mindset in game audio.

Film is a century-old medium, whereas video games have existed only for a handful of decades, with the first known example dating back to 1940s (Egenfeldt-Nielsen et al. 2013 ,58). As a result, much of the common thinking around sound production is still rooted in linear workflows. The tools used in film sound design evolved from analogue film cameras and tape machines, and even modern digital audio workstations (DAWs) reflect this heritage. Most editing software follows a linear timeline-based approach, and many plugins are designed to resemble vintage hardware, reinforcing traditional ways of working.

However, contemporary film and TV production have embraced non-linear and non-destructive workflows. Editors and sound designers can move freely along a timeline, modifying elements dynamically. Some directors, such as Denis Villeneuve and Alejandro González Iñárritu, integrate sound designers early in production—even during principal photography—allowing audio to influence creative decisions rather than simply complementing visuals in post-production. This approach elevates sound from a supporting role to an active storytelling tool. (Mangini 2021)
The term "audio post-production" traditionally differentiates post-production sound from production audio (i.e., sound captured on set). However, in modern filmmaking, these distinctions are increasingly blurred. Many professionals have transitioned from film sound to game development, bringing this terminology and linear workflow mentality with them.

5.2. A Non-Linear Perspective

To shift toward a truly non-linear mindset, I suggest a simple mental exercise: remove the word "post" from "audio post-production" and think of it simply as audio production. This subtle change reframes sound as an integral part of game development, rather than an afterthought. Game sound design is not just about creating sound—it's about shaping the game itself through sound. Ideally, audio should be developed in tandem with other game systems, evolving alongside them.

5.3. Fixed vs. Dynamic Soundscapes

A film is always the same from start to finish, no matter how many times it is played. Every aspect —the script, dialogue, set decoration, costumes, editing, music composition, sound design, and final mix—is meticulously pre-planned. A film director has full control over these parameters, sculpting a linear audiovisual experience that typically lasts for a few hours.

For a re-recording mixer, who is responsible for the final mix of a film, their job is to craft a polished soundtrack within a fixed cinematic structure. Re-recording, also known as mixing or dubbing, is the process of combining separate audio elements—dialogue, sound effects, and music —into a seamless, cohesive soundtrack (Yewdall 2011, 482). This stage ensures that all elements blend harmoniously while maintaining clarity, balance, and artistic intent.

Video games, however, present an entirely different challenge, as each play session varies in length, pacing, and content. Consider an open-world simulation game, where the player has freedom to navigate the virtual world and engage with different in-game systems. Many elements may be randomized or procedurally generated, resulting in dynamic and unpredictable soundscapes.

5.4. Predictable Mixing in an Unpredictable Medium

One example is *Red Dead Redemption 2* (Rockstar Games, 2018), a Western-themed open-world action-adventure game. The player can revisit the same location at different times of day, experiencing natural variations in environmental soundscapes—birds singing in the morning, insects buzzing at midday, and wolves howling at night. The game's dynamic weather system

further alters the soundscape, introducing wind gusts, rain, or full-fledged thunderstorms. Character movement also plays a role: the rustling of a leather jacket, the clinking of spurs, footsteps crunching over dry grass or wet mud contribute to the immersive sound design. Moreover, random encounters may suddenly shift the atmosphere—a quiet hunt in the woods could escalate into an explosive gunfight with a group of bandits.

Image 1. Screenshot from *Red Dead Redemption 2*.

These unpredictable interactions create a unique challenge for mixing. In a film, dialogue is carefully placed and mixed for clarity, but in a game, an important conversation may unfold mid-battle or while the player is galloping on horseback. Unlike in linear media, where every scene is precisely planned, interactive audio must account for countless variables, including where the player is looking, their actions, and how the game world is responding in real-time.

The unpredictability of player-driven experiences requires mixers to use predictable mixing strategies—anticipating common gameplay behaviors to create clarity, guide the player's attention, and avoid overwhelming the listener with excessive auditory information. Techniques such as using contrast (e.g. a moment of silence before an important event) or subtly directing player focus through sound cues help maintain coherence in an inherently unpredictable mix. (Riviere 2023, 1)

5.5. An Iterative Process

Not only is gameplay non-linear, but the mixing process and production itself follow an iterative structure. Game development relies on a constant feedback loop, where elements are designed, tested, refined, and improved based on ongoing reviews. This process extends to audio, where sound elements must remain flexible and adaptive to player interactions.

Films also incorporate test screenings and temporary mixes to gather audience feedback before finalizing the sound mix. However, in game development, this process is far more prolonged and iterative, lasting throughout the entire production cycle.

5.6. The Role of QA

The QA (Quality Assurance) team plays a critical role in this iteration cycle. In large-scale productions, different disciplines often have dedicated QA teams, focusing on areas such as gameplay, audio, and accessibility. Through playtesting, developers gather essential feedback to refine the experience. In smaller teams, however, testing responsibilities are often shared across the entire team. Regardless of scale, iteration and refinement are fundamental to game development, requiring a different approach to mixing than in traditional linear media.

6. Aspects of a Good Mix

There is no single correct answer to the question, “What makes a good mix for a video game?” In the end, mixing is inherently subjective, shaped by artistic intent, game design priorities, technical limitations and player expectations. In this chapter, I explore existing theories, concepts and ideas from industry professionals while reflecting on existing games as well as my own experiences. Although there is no universal definition of a “good mix,” we can analyze, categorize and prioritize game sounds to determine which elements are most crucial for the player to hear at any given moment during gameplay.

6.1. Perspective Matters

It is also important to recognize that video games are a broad and diverse form of interactive media, encompassing a wide variety of styles and genres. One of the most significant factors shaping a game’s soundscape is perspective. Is it a 2D or 3D game? First-person or third-person? Side-scrolling, isometric, or top-down?

In a three-dimensional first- or third-person game that strives for realism, the player is closer to the action and more immersed in the world of the game. This places them at the center of the soundscape, where they are surrounded by hundreds of sound sources. This is not to say that a more distant gameplay perspective cannot still create a sense of immersion for the player. In first-person games especially, spatial audio cues allow the player to hear movement all around them, creating a sense of presence and envelopment. As the player approaches a sound source in 3D space, it becomes louder, wider, and richer in detail, enhancing the sense of physicality.

By contrast, a 2D side-scroller with stylized retro pixel art may embrace an 8-bit audio aesthetic, featuring a limited dynamic range, lower signal-to-noise ratio, and a restricted voice count (Burke 2024, 6). While these hardware limitations no longer exist, they helped define a distinctive audio style that is still deliberately imitated in modern games for nostalgic or artistic purposes.

6.2. Genre

The genre of a game also plays a crucial role in determining how it should be mixed. A fast-paced action game may prioritize keeping the player character's actions at the forefront, ensuring that enemy sounds are clearly locatable for gameplay responsiveness. Survival horror games, on the other hand, emphasize atmosphere, using environmental sounds, reverberation, and spatialized audio cues to build tension.

For example, *The Last of Us Part II* (Naughty Dog, 2020) uses high-contrast mixing to intensify its stealth and combat sequences. During tense, quiet stealth sections, the mix prioritizes enemy movements, distant echoes, unsettling snarls, and the restrained breathing of the protagonist, creating an oppressive sense of vulnerability. When combat erupts, gunfire is brutally loud and impactful, almost uncomfortably so, reinforcing the weight of violence in the game's world.

Image 2. Screenshot from *The Last of Us Part II Remastered*.

In contrast, a game like *Devil May Cry 5* (Capcom, 2019) takes a completely different approach. As a high-energy action game, it maintains constant intensity, with a tighter, more compressed mix that keeps the momentum flowing. Unlike *The Last of Us Part II*, where gunfire feels deliberately

jarring, *Devil May Cry 5* ensures that gunshots, melee strikes, and enemy reactions are always prominent and satisfying, aligning with its arcade-like combat design. A wide dynamic range would not suit a game that demands nonstop action, so a punchy, more compressed mix better serves its gameplay experience.

Image 3. Screenshot from *Devil May Cry 5*.

6.3. Competitive Games

For competitive games the mix should always be subservient to gameplay. Listening to the game becomes a skill that players develop. (Ridgeway 2024) For example, in *Call of Duty: Modern Warfare II* (Infinity Ward, 2022), enemy footsteps have a slightly different attenuation curve than teammates' footsteps, ensuring they stand out more clearly and allowing players to react accordingly. In live-service games, updates and patches can affect the mix, possibly disrupting the balance players rely on. Radical changes in the mix can impact high-level competitive players, making proper testing and iteration crucial before release.

Image 4. Screenshot from *Call of Duty: Modern Warfare II*.

6.4. The Ludo-Narrative Plane

Zlobin (2021) proposes that game sound design isn't just about designing sound, but about designing games via sound. One way to approach game audio is by considering whether sounds serve a ludic or narrative function. This perspective shifts the focus from fidelity and aesthetics to a design-driven approach to analyzing a soundscape. While this chapter primarily touches on sound design, this framework is equally useful in mixing and understanding the overall game audio process.

Ludic functions guide the player's focus and provide information about gameplay states and events essential to progression. Narrative functions, on the other hand, reinforce storytelling, character development, and world-building. However, these categories are not mutually exclusive—a sound can embody both ludic and narrative roles simultaneously.

Zlobin further suggests that both ludic and narrative sounds can have negative effects if poorly implemented. An anti-ludic sound misinforms or distracts the player, while an anti-narrative sound disrupts the story or alters its intended emotional impact. To visualize this relationship, we can

think of a Cartesian ludo-narrative plane, where sounds exist fluidly between four quadrants: Synergistic, Misinforming, Destructive, and Utilitarian.

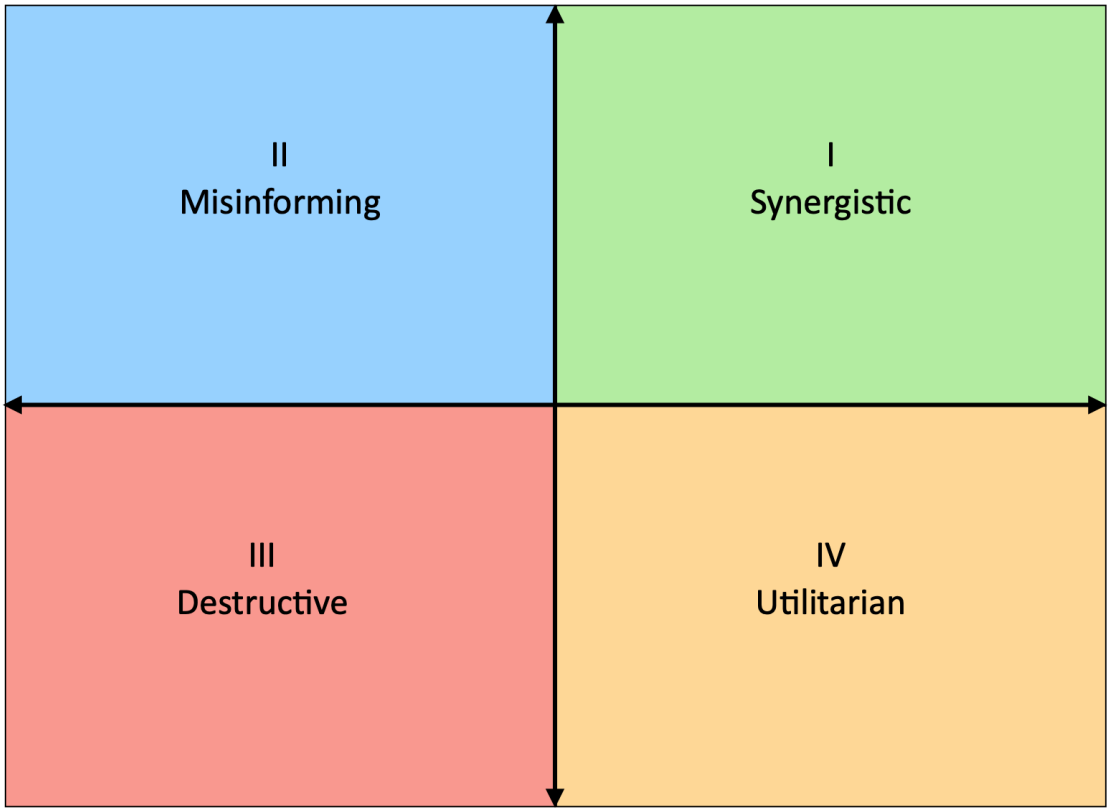

Image 5. The Ludo-Narrative Plane.

A synergistic sound enhances both gameplay and narrative. An example of this can be found in *The Last of Us Part II*, where dying enemies emit disturbing death screams. This sound not only provides ludic feedback by confirming that a threat has been eliminated, but also reinforces the narrative's emotional weight, making the player question their actions. (Zlobin 2021)

A misinforming sound, on the other hand, supports the narrative but may negatively affect gameplay. For instance, in an intense action sequence, critical dialogue lines may overlap with important ludic audio cues, potentially disrupting the player's ability to react. However, in some genres—such as horror—misinforming sounds can be beneficial. Unsettling environmental noises with no direct gameplay function create tension, enhancing the horror experience. Whether misinforming sounds are effective or disruptive depends entirely on context.

A destructive sound negatively impacts both gameplay and narrative, often due to audio bugs or poor implementation. These are unintentional and should be eliminated during mixing and testing. A utilitarian sound, on the other hand, serves a purely gameplay-driven purpose and is often non-diegetic, such as UI related sounds. (Zlobin 2021)

Jacobsen (2016, 14) refers to the ludic and narrative functions as “informant diegesis”. A sound in a game can be either diegetic or non-diegetic while always containing information about the current situation in the interaction process. Another benefit for analyzing the function of game sounds is to being able to communicate with the development team about the possible requirements for implementation. This type of design driven audio analysis gives us the language for developing games via sound. (Jacobsen 2016, 17)

The ludo-narrative framework not only helps evaluate sound effects individually but also clarifies which sounds should be prioritized at any given moment. Once this is clear, the person mixing the game can make informed decisions about which sounds should take precedence.

6.5. Loudness Standards

By analyzing overall loudness, frequency balance, and dynamic range, we can assess whether a mix possesses commonly desired qualities such as clarity, punch, and depth. Adhering to technical guidelines, such as loudness standards, ensures the mix translates well across different listening devices and environments, maintaining consistency and intelligibility for the player.

The European Broadcasting Union's (EBU) R128 recommendation specifies an average program loudness of –23 LUFS (Loudness Units relative to Full Scale) with a Maximum True Peak Level of –1 dBTP to prevent distortion caused by inter-sample peaks (EBU 2020). While originally designed for broadcast, these standards are increasingly relevant in game audio to provide a consistent auditory experience across various platforms. The Loudness Range (LRA) metric, also part of EBU R128, measures the variation in loudness over time, helping sound designers tailor the dynamic range to suit different playback environments, from high-end home theaters to mobile devices.
(Audiokinetic 2024)

6.6. Clarity

According to Bridgett (2024), a good mix consists of three key elements: clarity, impact, and spectacle. Clarity ensures that the mix has a full, well-balanced frequency response, where each sound element has its own space in the sound field without masking others. Masking occurs when louder sounds in the same frequency range obscure quieter ones, making them difficult to perceive (Laaksonen 2013, 34). To maintain clarity, it is crucial to critically assess which sounds contribute to the overall experience and remove unnecessary audio elements. Sound designers must identify competing frequencies and eliminate excessive auditory clutter, ensuring that only the most essential sounds remain. The most important sounds from a gameplay perspective are prioritized, and dialogue remains intelligible. Ideally, the player should be able to understand what is happening in the game solely by listening.

Wild Rumble (Mean Mink, 2025) is a 2D roguelike soccer game featuring cartoony animal characters with unique special abilities. I worked on the project as a sound designer, and one of the key challenges was crafting a soundscape that complemented the game's chaotic and comedic nature while maintaining clarity from a gameplay perspective.

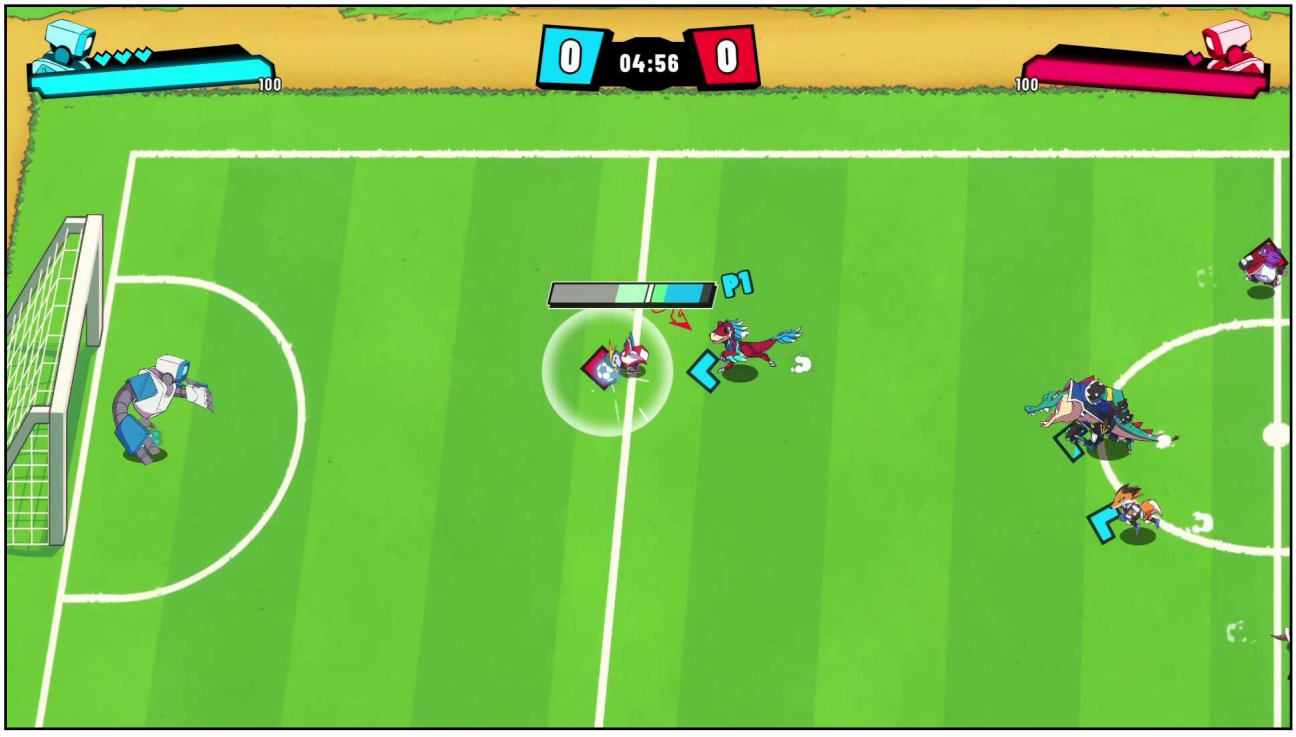

Image 6. Screenshot from *Wild Rumble*.

In the game, characters are constantly fighting for possession of the ball, rushing, and tackling each other. My goal was to ensure that, no matter how frantic the action became, the player could always subliminally perceive the ball's location on the screen. To achieve this, I emphasized all sounds related to the ball bouncing and ensured that every kick or reception was accompanied by a distinct, hard thump that cut through the chaos. The challenge was striking a balance where these sounds weren't overtly loud or distracting but instead worked subtly on a subconscious level to enhance gameplay readability.

It Takes Two (Hazelight Studios, 2021) is a split-screen cooperative action-adventure game in which two players control a couple on the verge of separation, who are suddenly transformed into living dolls. To return to their human forms—and mend their relationship—they must work together to overcome challenges. The game's narrative and gameplay are deeply intertwined, reinforcing the theme of cooperation at both a mechanical and storytelling level.

Image 7. Screenshot from *It Takes Two*.

The vertically split-screen format introduces unique mixing challenges, as two distinct perspectives and audio listeners coexist within the same 3D space throughout the game. This was addressed by splitting the mix into two halves, with each player's audio occupying a separate region. While the game's mix was authored in a surround multichannel format, the LCR (Left, Center, Right) channels

were used as the foundation for clarity and separation. Sounds related to the left-side player character—such as movement, abilities, and dialogue—were panned between the L and C channels, with a slight bleed into the R channel. The same approach was applied to the right-side character, ensuring that each player's audio remained distinct while still feeling cohesive.

Ambiences were authored in a quadraphonic format (L, R, Lr, Rr), allowing them to occupy the rear channels as well. A potential challenge arose when players moved into separate in-game environments, which meant that two different quadraphonic ambience tracks would overlap. To prevent clashing or excessive clutter, the solution involved lowering the volume of the right-side player's ambience in the L and Lr channels, and vice versa, effectively reducing interference between the two spaces while maintaining immersion.

Instead of working against the technical and creative constraints of the split-screen format, the Hazelight audio team embraced these limitations, allowing the narrative to dictate which side of the mix should take prominence at any given moment. This approach led to deliberate, sometimes dramatic shifts in the mix, ensuring that the player's attention was always drawn to new and engaging elements.

Despite being a split-screen game, *It Takes Two* follows a largely linear structure, which gave the audio team greater control over mix clarity by removing unnecessary sonic elements and maintaining clarity and focus throughout gameplay. The game was mixed with two players sitting side by side, simulating a realistic gameplay scenario where neither was positioned at the traditional listening sweet spot. This approach ensured that the mix was balanced for cooperative play, rather than being optimized for a single centered listener. (Hazelight, 2021)

6.7. Impact

A mix with impact is one that breathes, has constant movement, and is never static. It has enough dynamic range to support both quiet, intimate moments and loud, bombastic ones. (Bridgett 2024)

One example of this is *Inside* (Playdead, 2016), a 2D puzzle platformer game, where the player character’s footsteps act as the sonic anchor of the mix. Footsteps are mixed loud, front, and center to evoke an eerie sense of solitude and emptiness. However, when the player moves

continuously for a long period, the footstep volume is gradually and imperceptibly lowered, making room for other sounds in the mix.

Image 8. Screenshot from *Inside*.

Similarly, in *Ghost of Tsushima* (Sucker Punch, 2020), this subtle yet effective mixing technique was used to manage ear fatigue and maintain a dynamic soundscape during long traversals. The development team implemented a Real-Time Parameter Control (RTPC) in Wwise that tracked how long the player had been running or galloping on horseback. After 10–15 seconds, the system would gradually lower the volume of non-footstep foley sounds, allowing other elements to cut through the mix while keeping the experience fresh and engaging. (Meyer 2022)

This approach prevents foley sounds from becoming irritating to the player. If the character moves continuously, the foley volume attenuation is adjusted over a long period of time in a way that is significant yet imperceptible. However, whenever the character steps onto a new surface material —such as soil, wood, metal, or grass—the foley volume momentarily increases by 1.5 dB to emphasize the change. Essentially, the mix simulates how human auditory perception naturally adapts to new stimuli. (Riviere 2024, 105)

For example, if you're walking and suddenly step into muddy terrain, you initially notice the change, but within seconds, your brain accepts it as the new normal and stops focusing on it. This ability to filter out unnecessary information and prioritize relevant stimuli is known as the "cocktail party effect" (Farnell 2008, 90).

Image 9. Screenshot from *Ghost of Tsushima DIRECTOR'S CUT*.

Inside features a large set piece where the player must navigate a hazardous environment while avoiding a mysterious, lethal sonic boom that blasts through the air at regular intervals. To survive, the player must listen closely to the rhythm of the charging noise and the ensuing boom, strategically timing their movements to reach safe zones behind walls and other structures.

The setting is a vast research hall with an immense, echoing reverb, reinforcing the scale and atmosphere of the space. Every time the player reaches a safe area, the reverb is dampened by a low-pass filter, creating a muffled effect that brings the character's nervous breathing to the forefront. This subtle shift provides a sense of momentary, intimate safety. However, the second the player steps out of cover, the filter abruptly opens up, restoring the full, sharp reverberation—an auditory signal that danger is imminent.

Another interesting method used in *Inside* is what Andersen (2024) describes as "inverted ducking", where instead of sound effects being reduced in volume when the music level rises, they

are actually pushed up in volume to match the music's energy, making the moment more intense on all fronts. (Andersen 2024)

On a first-person game project I worked on, a dynamic mixing system was designed to adjust the ratio between the mid and side signals of NPC foley sounds based on their proximity to the player. As an NPC moved closer to the audio listener, the mix gradually introduced more of the side signal, increasing the perceived width of the sound.

Although the same foley performances were used for both the player character and NPCs, different microphone setups and recording perspectives were employed to ensure a distinct sonic identity for each. This demonstrates how mixing considerations can be integrated from the source recording stage, contributing to a more dynamic and immersive final mix.

6.8. Spectacle

A spectacle mix moment is when something unexpected or unusual happens in the soundscape and attracts the player's attention. For example, during an important narrative beat, the SFX bus volume is significantly lowered, allowing music to dominate the mix and emphasize the emotional weight of the moment.

A prime example of this can be found in *Death Stranding* (Kojima Productions, 2019). Throughout the game, players become accustomed to hearing the main character's heavy breathing, cloth movement, and equipment rattling as they traverse difficult terrain, accompanied by ambient wind, rain, and thunder. However, during pivotal moments—particularly as the player approaches their final destination in a mission—all of these environmental sounds fade away, giving way to music, creating a sense of satisfaction, relief, and progress.

Image 10. Screenshot from *Death Stranding DIRECTOR'S CUT*.

An undisclosed project I worked on features a spectacle mix moment where the player is pulled into a cosmic portal that distorts time and space. To sonically represent this effect, the mix was designed to dynamically respond to the player's proximity to the portal. As the player approaches, the music slowly ducks away, while its signal is simultaneously sent to an effect bus with a 10-second-long reverb and a flanger effect, creating an otherworldly sense of space and movement.

Additionally, the proximity of the portal subtly lowers the music's playrate, causing a slight pitch drop, while an ethereal shimmering layer of the portal's sound gradually rises in pitch and volume, reinforcing the feeling of being drawn in. A deep, rumbling noise emanates from the massive portal, implemented as an ambisonic file with a spread parameter in the audio middleware. With a spread value of 0, the sound remains mono, but as the value increases toward 100, it expands into a full surround sound field. This creates the illusion of the portal enveloping the player as they are pulled inside.

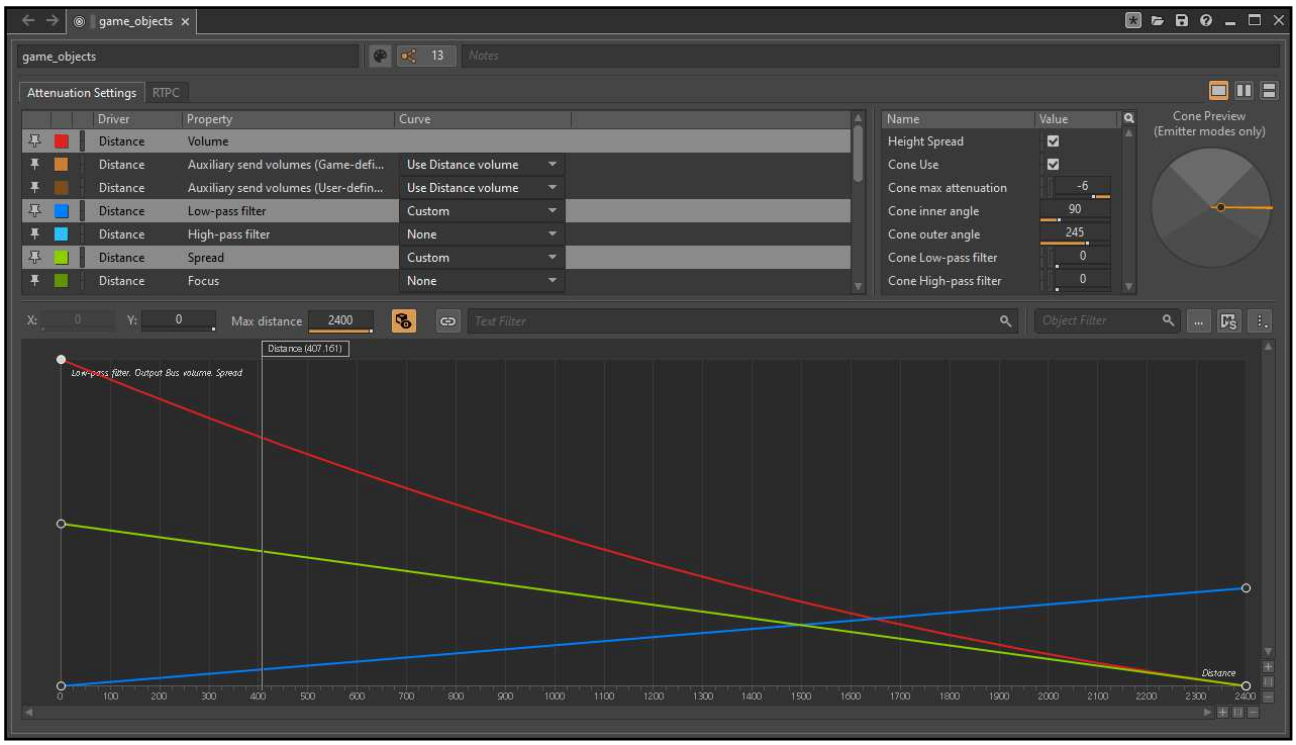

Image 11. Screenshot from Wwise, displaying attenuation curves for Volume, Low-pass filtering and Spread which are dynamically controlled by a Distance parameter.

Dolby Atmos height channels, LFE, or even controller speaker output and haptic feedback can further elevate these spectacle mix moments. However, these moments must be used in moderation, as constant drastic changes in the mix can unintendedly confuse the player.

6.9. Immersion, Emotion & Intensity

In *Shadow of the Tomb Raider* (Eidos Montréal, 2018), the development team aimed to create a soundscape that immerses players in fear and darkness, reflecting Lara Croft's (the main character) internal struggles and the game's thematic elements. Compared to open-world games, its linear structure allows for greater control over the dramatic shape of the mix, enabling the team to carefully sculpt audio moments for maximum emotional impact.

Additionally, the game utilizes spatialization techniques to enhance musical immersion. Instead of keeping the score as a traditional stereo or background layer, musical elements are dynamically positioned within 3D space, blurring the line between diegetic and non-diegetic audio. For instance, percussive rhythms may feel like they emerge from the surrounding environment, while distant chanting or instrumental textures shift in location based on the player's movement and camera perspective. This spatialized approach strengthens the connection between music and the

world, making the soundscape feel more organic and reactive rather than a separate, imposed layer of audio.

Image 12. Screenshot from *Shadow of the Tomb Raider.*

The implementation of Dolby Atmos further enhances this sense of depth and realism, utilizing height channels and object-based audio to create a three-dimensional sound environment. This allows elements like echoing cave ambiences, rainfall, or whispering voices to move freely within the vertical space, enhancing spatial awareness and immersion. (Kerr et al. 2018)

Even though less linear games make it harder to shape a predetermined intensity curve, it is still an important consideration. A mix should never be pushing at 100% intensity all the time—instead, leaving headroom for big moments makes the overall experience more dynamic and exciting.

6.10. A Triangle of Needs

Bridgett proposes a “triangle of needs” as a framework for understanding the current state of a mix. Each corner of the triangle represents a different mixing priority: Emotional, Ludic, and Immersive. Emotional focuses on the story, characters, and point of view, Ludic emphasizes gameplay mechanics and player feedback, and Immersive represents the game’s world and environmental depth.

Imagine a marker that moves dynamically within this triangle in real-time, adjusting based on what the player “needs” to hear at any given moment. Since human auditory perception is limited, mixing should prioritize subtraction over addition, ensuring that only the most relevant sounds stand out. There should always be a clear distinction between foreground and background elements, rather than all three mix priorities competing simultaneously. As a result, the marker should never remain fixed in the center but instead shift fluidly to balance gameplay, narrative, and immersion as needed. (Bridgett 2024)

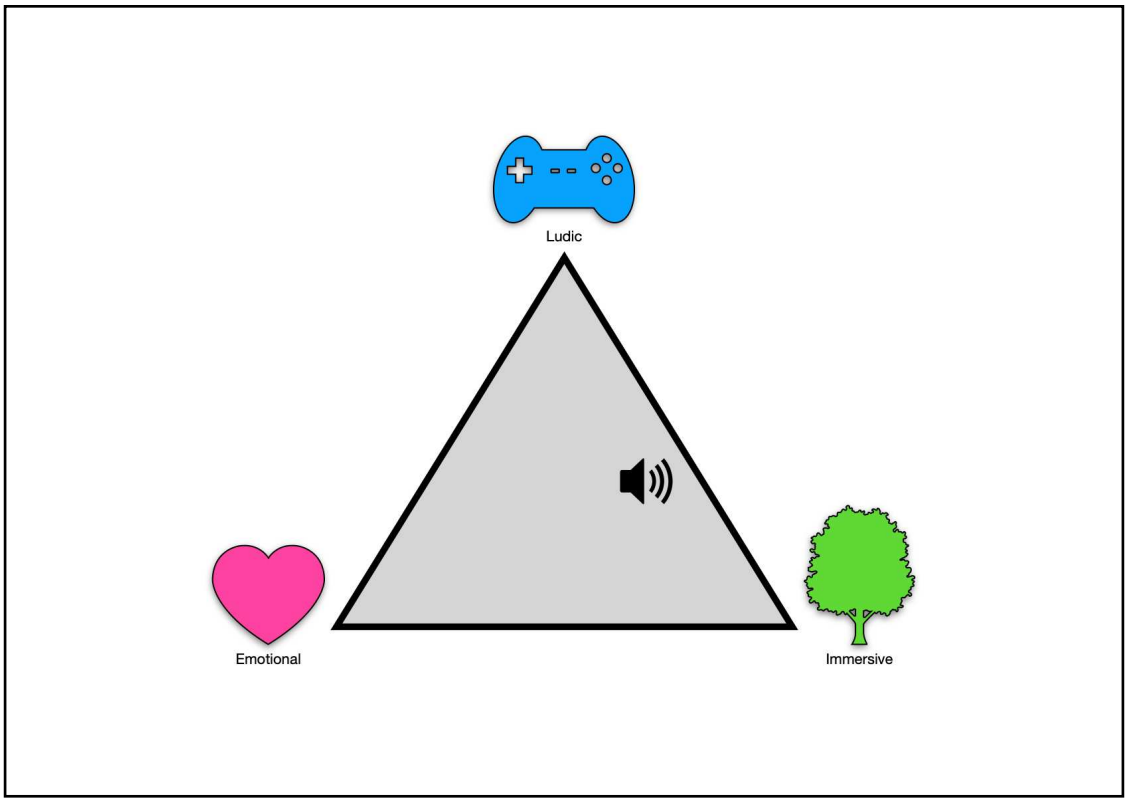

Image 13. Bridgett’s Triangle of Needs visualized.

6.11.Beyond Precision

A great mix is more than just technically precise—it is dynamic, emotionally engaging, and adaptable. It enhances the gameplay experience, guides the player's perception, and at times, breaks conventions to create unforgettable moments. Essentially a good mix provides the player sonically with everything they need to understand and enjoy the game. Lastly, it's important to remember that rules are meant to be broken—sometimes, unconventional approaches lead to the most unique and compelling results.

7. Arcade Roots & Spatial Futures

In the following chapter, I examine two games, *Tormentor X Punisher* (E-Studio, 2017) and *Returnal* (Housemarque, 2021), which differ in size, scope, gameplay, perspective, and style. However, both share a common lineage in the arcade game tradition, relying on fast-paced, challenging gameplay and meaningful auditory feedback.

Before analyzing these games, I first discuss the influence of arcade games, which is particularly relevant to *Tormentor X Punisher*. I then explore the game itself, examining how its mixing choices contribute to its gameplay and aesthetic. Following this, I shift focus to spatial audio, a key component in the sound of *Returnal*, before analyzing how the game utilizes 3D sound features.

7.1. Influence of the Arcade

> "*The sound of the early video game arcades is probably embedded in the consciousness of everyone who was a child during the late 1970s and early 1980s. To walk into an arcade was to experience an overwhelming onslaught of crashes, laser guns, synthesized speech, and electronic beeping music, all competing for our attention.*" (Wolf 2012, 119)

Video games gained commercial popularity through arcades. The arcardes are extremely loud environments because each game had to compete against a high level of ambient noise. Studies have shown that noise levels in arcades often reached between 88-90 dB (A-weighted) during peak hours (Mirbod et al. 1992, 1), creating a sonic battlefield where every game fought for auditory dominance. Sound not only played a crucial role in the enjoyment and success of early video games but also contributes to the nostalgia surrounding that era. The distinctive sounds of arcades have left a lasting influence on how games played on home consoles sound today.

7.2. Tormentor X Punisher

An extreme example of an arcade-like game with a tight, compressed, and loud mix is *Tormentor X Punisher* (E-Studio, 2017). This top-down twin-stick shooter throws hordes of demons and monsters at the player from all directions, creating an intense and chaotic experience. The soundscape is an aggressive barrage of punchy, satisfying gunfire, grotesque splatter explosions, and the goofy, exaggerated screams of demons. Adding to the mayhem, the main character

delivers announcer-style, profanity-filled commentary on the player's performance, all underscored by a relentless, pounding industrial metal soundtrack.

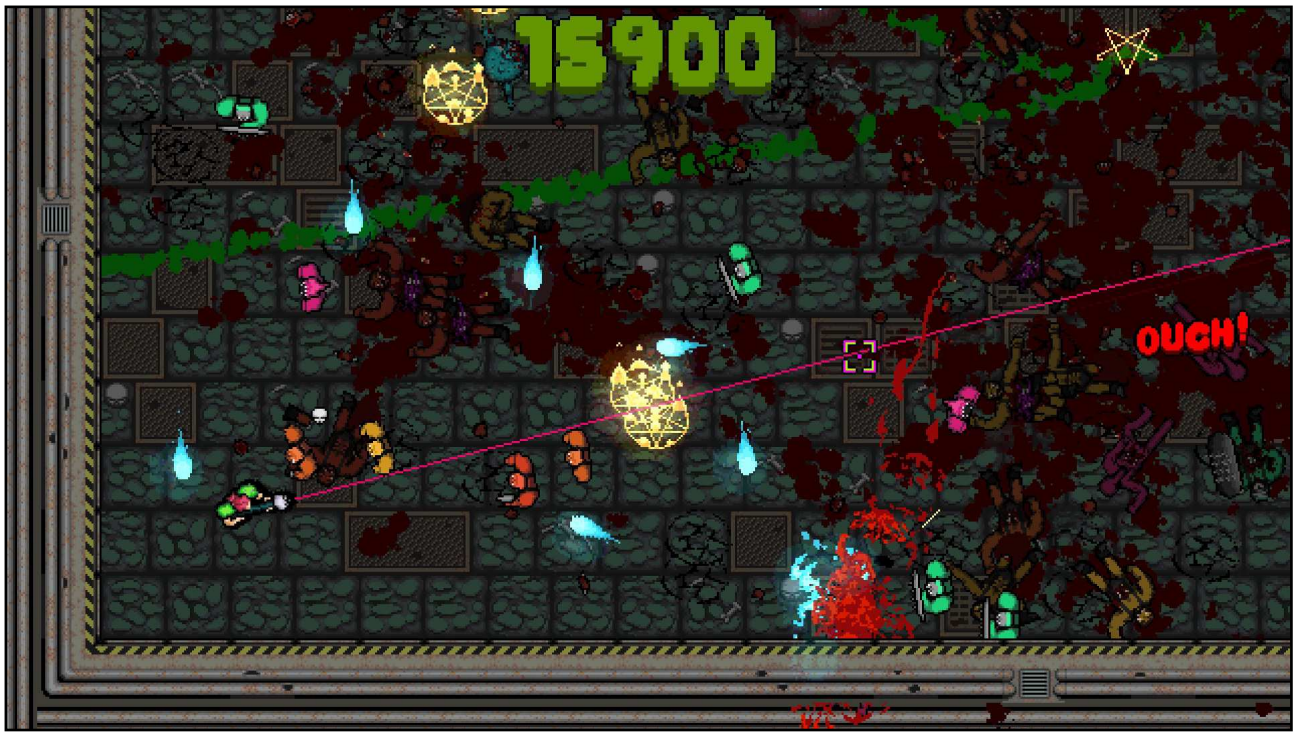

Image 14. Screenshot from *Tormentor X Punisher*.

Despite the chaotic and aggressive nature of the soundscape, the mix is carefully crafted to ensure that all key ludic sounds remain audible at all times. Turner (2016) explains that he wanted to eliminate the need for a traditional HUD, ensuring that players remained fully engaged in the fast-paced gameplay without unnecessary distractions. As a result, the game employs synergistic sound design, where every sound conveys crucial gameplay information while simultaneously reinforcing the game's narrative.

Turner (2016) further explains that enemy vocalizations serve multiple purposes. They act as spawn indicators and pre-attack warnings, guiding the player's attention, while also injecting demonic personality into the game's world. Similarly, the player character's dialogue provides both comedic and functional feedback, such as guiding the player to reload their machine gun, which humorously requires a shotgun blast to complete the process. The higher frequency vocalizations of both enemies and the player character are designed to cut through the mix, ensuring that critical auditory cues remain distinguishable amid the lower-frequency gunfire and explosions. This

demonstrates how mixing considerations begin at the sound design stage, shaping the final mix from the outset.

The game also features boss enemies, where the music momentarily fades out upon their arrival. This brief silence naturally guides the player's attention toward the incoming threat, helping them mentally prepare for the challenge. When the boss battle music resumes, the contrast makes its impact even stronger. The entire soundtrack is structured at 120 BPM, maintaining a consistent rhythmic flow throughout the experience. Notably, the machine gun's fire rate aligns with an eighth-note triplet subdivision of the music's tempo. As the gunfire slows down toward the last few bullets, the sound gradually becomes more muffled, subtly signaling to the player that it is time to reload. This is another example of how sound itself communicates gameplay mechanics without requiring on-screen indicators.

The fast-paced action, flashy visual effects, and violently shaking camera further amplify the sense of chaos as the player blasts through waves of enemies. This intensity creates an incredibly visceral game feel, which, for some players, may affect the longevity of gameplay sessions. However, *Tormentor X Punisher* is designed with a pick-up-and-play mentality, throwing the player straight into the action upon launch. This design choice makes it well-suited for short, high-energy gameplay bursts.

The arcade-inspired mix of *Tormentor X Punisher* deliberately disregards traditional loudness standards and dynamic range recommendations, reinforcing its chaotic identity. This bold approach has been recognized in the industry, earning the game the Firestarter Award for Best Sound Design at the Game Music Awards and a nomination for Excellence in Audio at the Independent Games Festival in 2018. The game serves as a prime example of how a collection of seemingly simple design choices can work together synergistically, creating a cohesive and deeply engaging experience. The game design, audio design, visuals, and narrative all function in harmony, proving that effective sound design is not just about fidelity, but about gameplay, emotion and immersion.

7.3. Spatial Audio

Spatial audio has become a widely used term in modern video games. Technologies such as ambisonics and object-based audio enable the creation of immersive 3D audio experiences,

allowing for more realistic spatial positioning of sounds. By leveraging HRTFs (Head-Related Transfer Functions) and binaural rendering, it is possible to simulate human directional hearing, enhancing player immersion and situational awareness.

However, while spatial audio is a powerful tool with exciting possibilities, it is important to consider both its benefits and potential challenges in implementation. Studies have shown that binaural rendering technology can improve player reaction times in competitive gaming (Brandmeyer et al. 2021). On the other hand, there is evidence that 3D sound may hinder the perception of visual events by dispersing visual attention (Mendonça et al. 2015).

Virtual reality (VR) motion sickness, also known as cybersickness, occurs due to a mismatch between visual, vestibular, and proprioceptive inputs. Spatial audio cues can help mitigate these symptoms by improving spatial awareness, reducing sensory conflicts, and reinforcing the user's perception of space and movement. One example is the use of a constant, audible environmental soundscape, such as an ambisonic ambience that rotates with the player's head movements, providing the brain with a reference point in space (Kostrova, 2020).

It is ultimately up to the sound designer to determine how to best utilize spatialization tools in a way that serves the gameplay experience. While these tools can greatly enhance immersion, they can also introduce unintended consequences. Binauralization plugins vary in quality, and as they act as filters that alter the sound, they may either enhance the auditory experience or degrade the clarity and integrity of the original sound source.

Due to human physiology, everyone perceives sound slightly differently, meaning that HRTFs and 3D audio effects work differently for each person. Currently, it is rare to have personalized HRTFs across all applications, so this must be taken into account when mixing a game while monitoring through a binauralizer. Careful implementation and mixing decisions are essential to ensuring that spatial audio contributes meaningfully to both gameplay and immersion rather than becoming a distraction or a technical limitation.

7.4. Wwise 3D Audio Pipeline

In this chapter, I present an overview of how 3D audio is managed within middleware, specifically using Wwise. The role of audio middleware is to author, manage, and optimize sound assets,

ensuring they are correctly mixed and delivered to the game engine. However, when dealing with 3D spatial audio, the process extends beyond Wwise's internal handling, requiring integration with platform-specific audio engines, such as the PlayStation 5's Tempest 3D AudioTech. Other possible options for spatialization include Dolby Atmos and Windows Sonic for Headphones.

When a player launches a game, the PS5 system software detects their current audio setup, whether it's stereo headphones, a 7.1.4 surround speaker configuration, or a soundbar. Wwise receives system-level configuration data from the PS5, allowing it to adjust output routing accordingly.

For example, if a user plays with a 7.1.4 surround speaker setup, Wwise outputs a 7.1.4 mix to match the system's configuration. If the player switches to headphones, the same mix is downmixed to stereo, but if 3D audio is enabled in the PS5 settings, Wwise passes spatial data to the Tempest 3D Audio Engine, which then processes binauralization using HRTFs or speaker-based rendering. Binauralization effectiveness varies between individuals due to physiological differences in ear shape and head size.

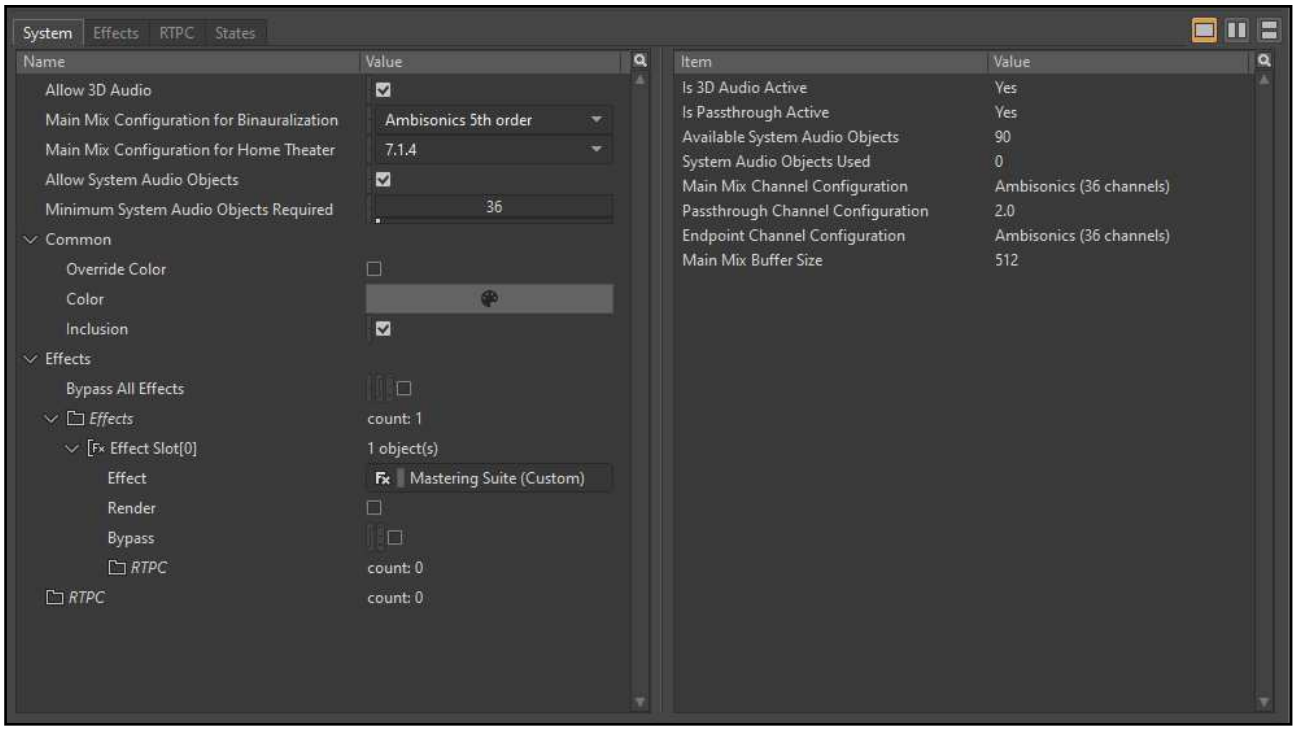

Image 15. Screenshot from Wwise, displaying the System Audio Device configuration.

At the final output stage, Wwise manages three distinct audio streams before they are sent to the PS5's system-level renderer:

1. **Passthrough** – This stream bypasses spatial processing entirely, keeping the sound in its original form. It is commonly used for stereo music tracks, UI sounds, or sounds that should remain unaffected by environmental effects.

2. **Main Mix** – This is the primary audio bus that undergoes binauralization when headphones are detected. If 3D audio is disabled, the Main Mix is simply downmixed to stereo or routed to the appropriate speaker configuration, basically becoming Passthrough.

3. **Audio Object** – These are individual sound sources that contain precise spatial metadata, allowing them to be positioned dynamically in 3D space. The PS5's Tempest engine processes these objects, using the system's settings to determine whether to binauralize them for headphones or spatialize them for speakers.

Image 16. Screenshot from Wwise, displaying Passthrough, Main Mix and Audio Object streams.

All three streams—Passthrough, Main Mix, and Audio Objects—are ultimately mixed at the system endpoint, where they can be mastered using the Wwise Mastering Suite to ensure cohesion across different playback environments.

Image 17. Screenshot from Wwise, displaying the Mastering Suite plugin.

7.5. Distinguishing Game Objects & Audio Objects

An audio object is a buffer containing metadata that defines its spatial and rendering properties. This metadata includes 3D position, distance, azimuth, elevation, focus, spread, spatialization type, and plugin-specific parameters. Within the Wwise rendering pipeline, this metadata is created, modified, and processed before the final audio signal is sent to the system end-point, whether headphones or speakers, for playback.

Game objects refer to actors or entities within the game world that trigger Wwise events. When an event is executed, the corresponding game object sends data back to Wwise, providing information such as 3D position, orientation, occlusion, obstruction, state, switch settings, and RTPC values. This dynamic exchange allows Wwise to continuously adjust spatialization and

processing parameters in real time, ensuring that audio accurately reflects in-game actions and environmental conditions.

While object-based audio has been used in games for decades, its modern implementation differs in its ability to leverage metadata for precise spatial rendering. Unlike traditional channel-based spatialization methods, such as 7.1 surround sound, which offer low-resolution spatial positioning, object-based audio enables independent sound sources to be positioned dynamically in 3D space with higher accuracy. (Ashby et al. 2021)

7.6. The 3D Sound of Returnal

This chapter draws primarily from the *A Sound Effect Blog*'s article by Walden (2021): "*Returnal: How Its Glorious, Dark Electronic Sound Was Made (and making the most of PS5's new audio engine)*". The article contains interviews with members of the *Returnal* audio team: Loic Couthier, Simon Gumbleton, Ash Read, Peter Hanson, Lewis Everest, and Harvey Scott.

In *Returnal* (Housemarque, 2021), players take on the role of Selene, an Astra scout who, while searching for the origin of a mysterious signal, crash-lands on Atropos—a shape-shifting alien planet. The game is a third-person roguelike shooter with bullet hell elements, delivering an intense challenge wrapped in a bleak and haunting narrative.

Image 18. Screenshot from *Returnal.*

7.7. A Hybrid Approach

Returnal takes advantage of Wwise audio middleware and the PlayStation 5's proprietary 3D audio technology, utilizing a hybrid mix that integrates multiple spatialization methods. The game's mix structure includes 5th order ambisonics, audio objects, 7.1 passthrough, stereo haptics, and controller speaker, with a total of 540 buses and 260 auxiliary buses, reflecting the complexity of its priority-based mixing system.

Audio objects offer precise spatial localization, enhancing situational awareness in combat, particularly for tracking enemy positions. This is crucial in *Returnal*, where enemies can attack from multiple directions, including above. However, while audio objects provide pinpoint spatialization, they are more CPU-intensive and cannot benefit from real-time bus processing, requiring careful optimization.

Ambisonics, on the other hand, provide a broader, more diffuse spatial representation. While they are less CPU-intensive than audio objects, they still require four times more processing power than 7.1 channel audio. Additionally, ambisonic beds take up significant voice count—a single third-order ambisonics ambient layer consumes 16 channels. In *Returnal*, ambisonics were primarily used for immersive environmental sounds, while audio objects were reserved for gameplay-critical elements with a strong ludic function.

7.8. Perception and Player Awareness

One of *Returnal*'s signature gameplay mechanics involves dodging hundreds of lethal projectiles fired by enemies. These projectiles were carefully mixed to ensure clear and immediate player awareness. Each projectile's movement sound consists of two distinct layers. The first layer is a material layer, representing elements like fire, energy, or water. The second layer is a "threat layer", which becomes more audible as the projectile nears the player.

The threat layer was specifically designed to target the 4-6 kHz frequency range, an area where human hearing is naturally most sensitive. To enhance player reaction and stress response, these sounds were constructed using distorted and granulated human screams, sourced from both adult

and child vocalizations. This approach leverages the natural human reaction to high-stress, fear-inducing sounds, heightening tension and urgency during gameplay.
Despite *Returnal* being a shooter game, after extensive playtesting and experimentation, Couthier decided against the conventional approach of using the player character's weapon as anchor of the mix. Instead, the focus shifted to ensuring that the player could hear everything happening around them, which was essential for survival in the game's chaotic combat encounters.

With hundreds of sounds playing simultaneously, all of which are in most cases eventually summed into a two-channel binaural audio stream, the mix had to account for the human brain's limited capacity for auditory attention (Oberem 2020, 9). This led to the implementation of a priority-based mix system, where enemy sounds were assigned different priority levels. Higher-priority sounds would dynamically duck lower-priority sounds during playback, ensuring that the most crucial audio cues remained perceptible amidst the mayhem.

Much of the mixing process involved determining which sounds required the highest priority in different gameplay contexts. Extensive testing was conducted to identify the most critical auditory cues at any given moment. However, simply removing lower-priority sounds was not a viable solution, as it risked making the mix feel bland and lacking in energy and sense of danger. Striking a balance between clarity and sonic intensity was a major challenge.

To further manage mix clarity, *Returnal* utilized culling features and Wwise's voice limiting functions to reduce auditory clutter while preserving the necessary gameplay feedback. This approach ensured that even in the game's most intense combat sequences, the mix remained informative, responsive, and immersive.

8. The Game Mixing Process

Mixing in video games presents unique challenges compared to traditional linear media like film or television. In an interactive environment, sounds must dynamically adapt to gameplay, ensuring clarity, immersion, and responsiveness while maintaining the artistic and technical integrity of the mix. Unlike a pre-mixed linear soundtrack, a game mix is constantly evolving based on player actions, spatial positioning, and gameplay state changes.

This chapter provides a comprehensive exploration of the practical aspects of game mixing, covering pre-production planning, mix structure design, spatialization strategies, sound prioritization techniques, and mastering workflows. While mixing approaches vary depending on the genre, platform, and creative vision of the game, this chapter highlights fundamental principles and real-world techniques that can be applied across different types of projects.

8.1. Defining the Audio Vision

Pre-production involves establishing the audio vision, which is based on the original game idea and the detailed Game Design Document (GDD). The GDD provides crucial information for the entire development team, including the game's core ideas, gameplay mechanics, themes, aesthetics, and project guidelines.

In response to the GDD, the audio director creates a Sound Design Document, defining a unified audio vision. This document outlines key aspects such as team roles, tools, audio middleware, plugins, and technical audio guidelines, including loudness standards, file formats, and naming conventions.

Having unified naming conventions is particularly beneficial, as it improves file management, batch processing, automation potential, overall project organization, and readability (Uzer 2024). These elements are all integral to efficient mixing and managing a game audio project.

Together, these documents form the foundation for a cohesive and consistent audio direction for the project. In large teams where individuals work in different environments, maintaining consistent SPL and LUFS standards for sound monitoring is common practice. This ensures that all

team members can make creative decisions with confidence, knowing their work accurately translates across the project. (Broomhall 2024)

8.2. Communicating through Reference

At this stage, it is common practice to create an audio target video - a simulated representation of how the final product should sound. This serves as both a creative and technical reference, outlining the desired audio experience while also illustrating the necessary audio system functionalities. These may include automated filtering, sidechaining, and real-time effects such as delays and reverbs.

By presenting this vision in a linear video format, it becomes easier to communicate complex audio concepts to programmers and other non-audio team members. This ensures that the development team understands what type of audio systems need to be implemented for the project. Throughout production, the audio target video serves as a reference and style guide, helping the team maintain consistency and stay aligned with the original audio vision. (Riviere 2023, 16)

8.3. Planning the Final Mix Early

During pre-production, it is beneficial to start considering the final mix and its key characteristics. What will the dynamic range be like? Is the game a large-scale, cinematic experience that emphasizes spectacle and immersion, naturally requiring a wider dynamic range similar to that of a film? Or is it a handheld game with a smaller scale, 2D graphics, and a focus on gameplay—where a more compressed mix might be preferable to ensure clarity on both small speakers and headphones?

8.4. Cross-Disciplinary Collaboration

In order to achieve a mix that supports both gameplay and narrative, establishing effective communication and collaboration with other disciplines is just as important as the mixing itself. The game design, narrative, art, animation, and visual effects (VFX) teams must work closely with the audio department to ensure a cohesive experience. Since audio is an invisible element, extra

care must be taken to keep all disciplines informed and aligned with the intended sonic goals of the project.

8.5. Structuring the Mix

Sub-mixes and sound categories should be carefully planned to ensure a well-structured mix. Where do the sub-mixes sit in terms of frequency spectrum and loudness? AAA studios such as Sony Santa Monica Studio, Naughty Dog and Remedy implement this approach by defining predetermined loudness standards for different sub-mixes, including ambiences, environmental sound effects, dialogue, loud special SFX, and extreme loudest special SFX. In this system, the loudness level of an audio asset is decided before it is even imported into the audio middleware.

Even dialogue performance intensities may be loudness standardized e.g. shouting opposed to normal speaking. *God of War Ragnarök* (Sony Santa Monica Studio, 2022) contains 25 000 lines of dialogue in 13 languages. Having loudness standards for different dialogue intensities improves consistency and overall workflow when it comes to mixing. (Broomhall, 2024)

In audio middleware, a mixer structure—referred to as a hierarchy in Wwise—is established and prepared. Later in this text, I will present a more detailed overview of how a typical mixer hierarchy might be structured. The core idea is to categorize sounds into different mix buses based not only on their sound type but also from a ludic perspective, ensuring that critical gameplay information is always delivered clearly to the player.

8.6. Mixing with Purpose

For instance, a project might include both a ‘Gameplay Bus’ and a ‘Critical Gameplay Bus’, each serving a distinct purpose in the mix. The ‘Gameplay Bus’ may contain general in-game sound elements, such as environmental sound effects, minor character actions, foley, and non-essential UI feedback. These sounds contribute to immersion but are not necessarily vital for gameplay decisions.

On the other hand, the ‘Critical Gameplay Bus’ is reserved for high-priority sounds that directly impact gameplay and require the player's immediate attention. This could include enemy attack cues, low-health warning sounds, key interaction prompts, or footsteps of nearby threats in a

stealth game. The mixing system can dynamically adjust the balance between these buses, ensuring that crucial sounds remain intelligible even in dense, action-packed scenarios.
By implementing this kind of audio hierarchy, sound designers can ensure that gameplay-relevant audio always cuts through the mix, preventing important cues from being masked by secondary sound elements. This approach helps to balance immersion and clarity, keeping the player informed and engaged. It also allows the game to "mix itself" to some extent, ensuring a more consistent and controlled soundscape from the outset.

8.7. Creative Alignment

Each game project has a unique high-level vision that the mix must creatively and technically support. For instance, *God of War Ragnarök* (Santa Monica Studio, 2022) focuses on intimate storytelling, exploration, and expressive combat, while *Returnal* (Housemarque, 2021) emphasizes cosmic horror, time-loop distortion, and emotional loss. Ensuring that the mix aligns with these core pillars throughout development is essential. Achieving this requires strong communication and collaboration—not only within the audio team but also with developers and creative directors.

Image 19. Screenshot from *God of War Ragnarök*.

8.8. Music & Sound FX Synergy

Another crucial consideration is how the instrumentation of the musical score interacts with the sound effects. If the soundtrack is in the style of electronic dance music with bass-heavy beats, there may be little room left for low-end-rich sound effects. This can, of course, be managed through mixing techniques like multiband EQ sidechaining, ensuring a balanced mix. However, pre-production is the ideal time to establish a plan where all sonic elements have their own space, working together in harmony. These constraints can also spark creative ideas regarding the style, genre, instrumentation, and arrangement of the music.

A great example of this is *Black Myth: WuKong* (Game Science, 2024), an action role-playing game inspired by Chinese mythology. The game's battle sound design features dynamic, punchy, bass-heavy impacts, whooshes, and magical shimmers, while the music predominantly utilizes traditional East-Asian acoustic instruments such as guitars, flutes, drums, percussion, and vocals. The high-pitched instruments naturally cut through the mix, complementing the game's world and aesthetic. Additionally, the composition and arrangement leave space for sound effects, resulting in moments where a punch, a drum hit, airy whooshes, and the scratchy slides of a string instrument merge into a complete, exhilarating soundscape that elevates the entire experience.

Image 20. Screenshot from *Black Myth: WuKong*.

8.9. Starting from the DAW

The mixing process begins at the content level, in the DAW, where sound effects are created and designed. However, the act of mixing begins the moment when two or more sounds play in a live game build. (Riviere 2023, 25)

Each sound effect is created with a specific purpose and belongs to a certain category, which may have predetermined loudness levels or spectral rules. For instance, character dialogue might be restricted from containing frequencies below 100Hz, as typical human speech fundamental frequency ranges between 110 – 300Hz (Rossing et al. 2014, 343). This high-pass filtering is to ensure clarity and prevent low-end muddiness in the mix. Similarly, explosions could be classified as the extreme loudest sounds in the game, with a maximum short-term loudness cap of –16 LUFS and a true peak limit of –3 dBTP.

Spectral characteristics may also be defined early in the process to maintain consistency across the mix. For example, UI sound effects could be designed with a bright, pristine high-end, meaning their higher frequencies are enhanced to cut through other elements. Establishing these guidelines early on ensures that sounds sit harmoniously within the mix while maintaining clarity and coherence across the game's audio environment.

When mastering sounds in the DAW some prefer to normalize all SFX to one loudness level first and then make further adjustments in the engine according to taste. Establishing submix loudness standards is especially beneficial in large-scale productions, where tens or even hundreds of people contribute to the sound design. In smaller projects, however, it is often entirely practical to handle leveling later in the process, allowing for more flexibility and creative control during implementation.

Individual audio files should still be mastered to a sufficiently loud level to take full advantage of the available bit depth, maximizing the resolution of the dynamic range. Otherwise, the signal may need to be boosted later, which can increase the audibility of noise and artifacts due to a lower signal-to-noise ratio and pre-existing quantization errors. Additionally, lossy compression applied by the game engine will further degrade audio quality if the bit depth is not fully utilized. To prevent digital clipping, mastering individual SFX to a maximum level of -1 dBTP is recommended, with -3 dBTP often used as an added safety margin. (Huber et al. 2018, 207)

Once new SFX are added to the game, it is then tested with the new sounds, and mix adjustments are made. At this stage of production it is important to establish a constant test and feedback loop, where new features are thoroughly tested out and investigated how they interact with one another. New features are built to the game, which might affect the earlier decisions made about the sounds.

8.10. The Intrinsic & Extrinsic Space

While game engines and audio middleware are capable of real-time sound spatialization, it is not uncommon to apply spatial processing for acoustic simulation during the sound design phase in the DAW. Certain sounds require precise control over reverb and spatial characteristics, making it preferable to pre-render (bake) spatial effects into audio files before implementation. Instead of relying on real-time DSP reverb effects in-engine, pre-rendered effects help reduce CPU load and ensure consistent spatial characteristics, particularly for sounds that do not require dynamic environmental adaptation.

In electroacoustic music, a sound object's spatial and spectral properties can be categorized into intrinsic and extrinsic spaces. This method of analyzing and describing the shape of sound over time is known as spectromorphology (Smalley 1997, 107). Intrinsic space refers to a sound object's inherent spatial and spectral characteristics, while extrinsic space refers to the acoustic environment in which the sound exists.

According to an interview with Andersen, adding a short, first-reflection-like delay or reverb to a sound can alter the perceived size of the sound source (Riviere 2023, 109). Another example is the use of decorrelation techniques, such as applying an all-pass filter, to blur or spread the perceived sound image. If designing the sound of a massive, mysterious, magical orb floating in the air, these techniques can be used to shape the perceived spatial characteristics of the sound in the DAW before it is placed into the game engine's acoustic space.

However, while pre-rendering offers greater control over how a sound behaves in the mix, it also reduces flexibility and limits real-time dynamic adjustments. Managing multiple variations of the same sound—each processed for different acoustic environments—can increase project complexity, requiring careful asset organization and implementation. These trade-offs must be

considered, as the decision to pre-render or process sounds dynamically ultimately depends on the balance between control, efficiency, and the dynamic needs of the game.

8.11. Mixer Structure

At the game engine or audio middleware level, traditional mixing tools such as volume sliders, panners, EQs, and filters are readily available. While each engine differs in terminology and visual layout, the fundamental concept of a mixing desk signal flow remains consistent across platforms. A key aspect of game mixing is categorizing and organizing sounds effectively. Designing a mixer structure with a logical hierarchy ensures that gameplay dynamics are supported in real-time. Sounds are routed through an audio bus structure, which processes them in a hierarchical manner, allowing for group-based mixing adjustments and dynamic mix control.

Image 21. Screenshot from *Wild Rumble*'s FMOD Studio project, displaying the mixer view.

A top-down approach to building a mixer structure offers flexibility, as buses can be added incrementally as the project evolves. The process typically starts with larger, high-level sound

categories, such as ‘Environment’, ‘Player’, and ‘NPC’, before breaking them down into more granular subcategories. For example, the ‘Player’ bus can be divided into ‘Foley’, ‘Voice’, and ‘Abilities’, with ‘Foley’ further categorized into ‘Cloth’, ‘Equipment’, and ‘Footsteps’. (Riviere 2023, 51)

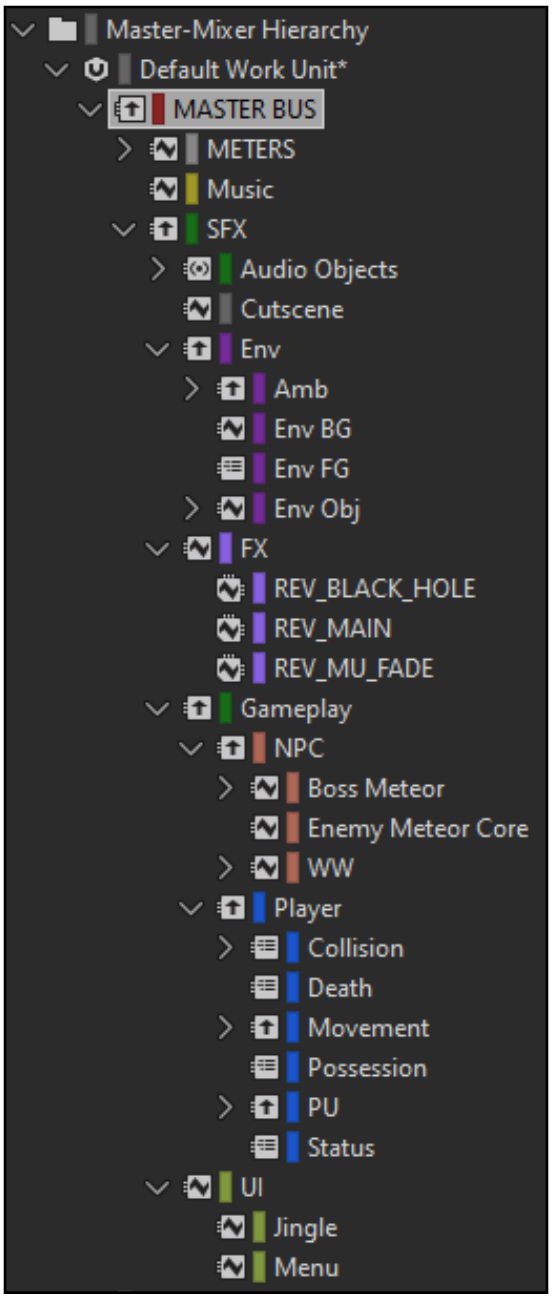

Image 22. Screenshot from Wwise, displaying a Master-Mixer Hierarchy.

Sub-buses should not be added arbitrarily but with a clear purpose. While more buses provide greater control, they also increase complexity, which can make the mix more difficult to manage.

Before adding additional buses, it is crucial to consider the reasoning behind it—is it necessary for separate processing, or would a simpler structure be more efficient? Anticipating potential mix scenarios in advance can help determine which subcategories require independent processing, such as different reverb treatments, dynamic range adjustments, or priority-based mixing strategies.

In an undisclosed project I worked on, we created an 'Enemy' bus, which was further divided into different enemy types. Each enemy type was assigned a priority level based on how threatening they were to the player. To reinforce this hierarchy in the mix, we implemented a ducking system, where more dangerous enemies would automatically reduce the volume of less threatening ones, ensuring that the player's focus remained on the most immediate threats.

There are cases where different layers of a single audio event may require distinct spatialization treatments and custom bus routing. For example, an explosion could be split into low-frequency and high-frequency content layers. The low-frequency layer represents the powerful burst of energy, while the high-frequency layer conveys debris and destruction.

Low frequencies are perceived as less directional by the human auditory system because of their longer wavelengths, which result in less distinct interaural level differences (ILD) and interaural time differences (ITD) (Aro 2006, 30). With this in mind, the low-frequency layer could be routed to the Passthrough bus, allowing it to remain omnidirectional and unaffected by spatialization processing, while the high-frequency layer could be sent to the Audio Object bus, where it can be precisely spatialized to enhance the sense of destruction and movement in 3D space.

HRTFs are filters, meaning they introduce coloration and can cause comb filtering due to phase shifting. As a result, transients may lose their attack and become mushier, while low-end frequencies are filtered out. (Ashby et al. 2021) It is important to consider the trade-offs of using HRTFs, as they increase spatial resolution but can degrade sonic quality. This is another example of why implementation, spatialization, and mixer structure should be carefully considered early in the sound design phase.

Modularity and flexibility are essential when designing a mix structure, as it is common to remove, add, or re-route buses throughout development. Since gameplay features may evolve, the mix

structure must also be adaptable, ensuring that it can accommodate changes and support new audio requirements as the project progresses.

8.12. Avoiding Listening Fatigue

Each sound file is attached to an audio event, which is then triggered by the game engine. For instance, a footstep sound is called whenever the player character's walking animation reaches the frame where their foot touches the ground.

Human hearing is highly sensitive to changes in pitch, timbre, amplitude, and timing, making it particularly attuned to patterns and repetition. The ability to distinguish the rustling of the wind from the movement of an approaching predator has provided a crucial evolutionary advantage. The short-term auditory memory system, known as echoic memory, retains sounds for a few seconds, enabling rapid comparison between current and recent auditory information. (Farnell 2008, 88) Because of this, players can quickly detect if a sound is repeated identically over time, leading to fatigue or breaking immersion.

To avoid repetition and listening fatigue in games, most sounds that occur frequently—such as footsteps—have multiple variations. Instead of repeating a single sample, the system might cycle through 20 different variations. Additionally, randomized volume and pitch modulation can be applied to create further variation, making the sound feel more organic and dynamic. Even slight randomization helps create a more natural listening experience.

The line between sound design and mixing can sometimes be blurry. While adding pitch variation is an essential sound design technique, the way these sounds interact within the game's mix is still influenced by mixing choices.

8.13. Sound Prioritizing Techniques

Taking advantage of the bus structure for category-based sound prioritization can be achieved using run-time signal modulation. Techniques such as sidechaining, ducking, and rule-based mix states help declutter the mix, making subtle or dramatic adjustments depending on the gameplay context (Riviere 2023, 56).

A multiband sidechaining approach can be used for a more refined and subtle mix adjustment. For example, when dialogue is present, only the specific frequency range of speech in the ambience track's spectrum is attenuated, preserving the rest of the environmental sounds. In contrast, during an action-heavy gameplay sequence, a more aggressive and broad-frequency ducking effect may be preferable. This not only improves clarity but can also introduce a pumping effect, adding a sense of intensity and momentum to the overall soundscape.

A rule-based mix state could dynamically alter the mix based on gameplay conditions. For instance, if the player character is injured and health is low, the mix could apply a low-pass filter to the SFX and Music buses, reducing high-frequency detail while pushing the character's voice forward to emphasize labored breathing and pain reactions. Additionally, an auxiliary send could introduce a disorienting reverb, reinforcing the player's weakened or unstable state.

In VR audio design, player eye tracking can be leveraged for adaptive mixing. When the player looks at a key object or interaction point, its volume can be subtly increased, while low-pass filtering is applied to less prioritized sound emitters, naturally guiding the player's auditory focus without needing visual indicators.

8.14. The High Dynamic Range System

Another common approach in sound prioritizing is the HDR (High Dynamic Range) System which is a technique designed to manage the limited dynamic range of digital audio while preserving the clarity and balance of important sounds. In the real world, the human ear can perceive a dynamic range from a 30dB whisper up to a jet engine which is around 140 dB (Audiokinetic 2024). Originally developed by DICE in 2006, HDR audio was introduced as a way to map the large dynamic range of reality onto the smaller one allowed by 16-bit, 96dB dynamic range playback systems (Huber et al. 2010, 208).

HDR works by dynamically adjusting the mix within a movable volume window—the loudest sounds raise the window, while sounds below a set threshold are automatically culled. Beyond its practical function, HDR mixing can also be used creatively. In *Battlefield 4* (DICE 2013), for example, the HDR system simulates the acoustic energy of battle, making the environment "recover" from loud sounds before allowing quieter details to return. This pumping effect mimics the way cheap audio limiters work in consumer recording equipment, reinforcing the sensation of sonic intensity

in combat scenarios. By prioritizing key sounds dynamically, HDR ensures that important gameplay audio remains clear and immersive, even in highly chaotic soundscapes. (Stevens et al. 2010)

8.15. HDR in Wwise

The HDR system is available in Wwise and implementing it early in production is crucial for maintaining control over Voice Volume gain staging and ensuring a consistent, clean audio pipeline. Since Wwise determines HDR behavior based on a sound's final Voice Volume, various factors—such as parent Actor-Mixers, RTPCs, States, and Bus Volume adjustments—can unintentionally affect how sounds interact within the HDR system.

Establishing HDR from the beginning helps prevent unintended Voice Volume changes that could disrupt how sounds are prioritized in the mix. When gain staging is planned early, the audio team can set clear rules, making it easier to maintain consistency across all assets and ensuring that HDR functions as intended. It also reduces the risk of technical issues later in development, avoiding the need to rework complex sound routing or fix unpredictable volume changes caused by conflicting system parameters.

If HDR is introduced midway through development, the team may have to spend significant time cleaning up the existing mix, troubleshooting unexpected volume shifts, and adjusting routing logic to ensure sounds behave correctly within the HDR framework. This process can be far more challenging than starting with HDR principles in place from the outset. (Previty 2024)

8.16. Attenuation Curve

An attenuation curve defines how a sound emitter behaves based on its proximity to the audio listener, which acts as a virtual microphone or set of ears within the game world.

The first element in the chain is volume attenuation, where the sound grows louder as the emitter moves closer and fades as it moves away. Next, a low-pass filter can be applied, with its cutoff frequency modulated by distance, mimicking how higher frequencies dissipate over distance in real life. Both of these effects simulate how sound naturally loses energy over time and space.

Beyond volume and filtering, additional spatialization techniques can enhance realism. A spread effect can dynamically adjust a stereo file's mid/side ratio, making the sound mono at a distance and gradually widening into stereo as the emitter approaches. This helps create a sense of proximity and emitter size.

Other parameters can be automated not only by distance but also by angle—for instance, adjusting volume or filtering based on whether the sound source is in front, behind, or to the side of the listener. This type of directional filtering helps prioritize important sound sources in front of the player, preventing player distraction or unwanted noise from behind that could clutter the mix.

Different sound effects require customized spatial settings. For example, an NPC delivering critical dialogue should not be attenuated so heavily that their voice becomes inaudible from a distance. Meanwhile, a barking dog with no critical gameplay relevance can follow a more realistic attenuation curve, fading naturally as the player moves away.

In a project I worked on, the unit scale of the entire game was changed (i.e. the measurement system determining distances in the game world) at a fairly late stage of production, which rendered all my previously adjusted attenuation curves inaccurate. I had to go back and readjust all of the custom attenuation curves made for multiple different objects. Sounds that the player was able to hear earlier at a maximum distance were no longer audible.

Such changes, if not properly communicated, can result in redundant troubleshooting and a loss of work hours. Most likely, this type of incident does not occur as frequently in more experienced and professional environments where producers oversee each discipline. However, it is important to acknowledge that even small changes in other disciplines can unexpectedly affect audio. This is why constant playtesting is crucial.

8.17. Reverb

An audio effects bus with real-time delay and reverb system can be created to enhance the player's sense of physical location and depth. There are multiple approaches to implementing reverb, and choosing the right method depends on both creative intent and technical constraints. One key consideration is whether to use algorithmic or convolution reverb, keeping in mind the DSP

processing power required. For example, using a multichannel convolution reverb can be CPU-intensive.

Another decision is whether the system should be state-based or dynamic. A state-based system switches between pre-configured reverb settings, with each instance custom-tailored to different game levels or environments. In contrast, a dynamic system adjusts reverb parameters in real-time based on the player's movement and environment, aiming to simulate the actual physical space. This approach would require ray tracing and additional programming support for implementation.

In a state-based system, the sound designer could place trigger boxes in the game world. When the player enters a trigger zone, the reverb setting smoothly transitions to match the new environment. This method allows for precise artistic control over how spaces sound. The dynamic approach, on the other hand, can create more seamless transitions between spaces but may result in less control over the final sound. While dynamic systems can make environments feel more organic and reactive, state-based reverb allows for handcrafted acoustic design, offering opportunities for greater artistic expression.

Ultimately, the development team must determine which approach aligns best with both the creative direction and technical feasibility of the game. While there is no single "correct" method, understanding the strengths and trade-offs of different mixing techniques is essential in creating a balanced, immersive soundscape.

8.18. Mastering

A good starting point for mastering is ensuring that the game sounds balanced and well-mixed on its own, without relying on master bus processing. However, it is recommended to keep a master limiter active throughout development to catch peaks that exceed –1 dBTP, preventing unintended clipping.

The first step in mastering is to control overall loudness and dynamics, ensuring a consistent and controlled mix across various gameplay scenarios. Once this is achieved, the next priority is to prepare the game for different playback systems, including home theater setups, TV speakers, soundbars, and headphones.

It is also possible to implement a "midnight mode", in which dynamic range is compressed to maintain clarity at low listening volumes, making dialogue and critical sounds more intelligible without excessive loudness spikes. The Wwise Mastering Suite, for example, provides tools such as a master limiter, multiband compressor, multichannel volume control, and master EQ, allowing the mix to be optimized for different end-points.

For instance, EQ adjustments may be necessary for flatscreen TVs, which typically lack low-end response and may emphasize harsh high frequencies. A home theater setup might require a –3 dB attenuation on rear channels to maintain front-heavy focus on dialogue and primary gameplay sounds. Additionally, applying a subtle cut in the low-mid frequency range can help reduce muddiness and improve overall mix clarity.

A multiband compressor can be particularly useful for midnight mode, ensuring that louder peaks are controlled while preserving important details in quieter sounds. However, it is crucial to remember that mastering affects all sounds in the game, so drastic changes should be avoided to prevent unintended alterations to the mix’s balance and dynamics.

A good approach to mastering is to start with the highest-end configuration, such as a surround sound mix, and then gradually adapt the mix for lower-end playback setups, addressing one version at a time. This step-by-step process helps maintain clarity and focus, ensuring that each mix translates optimally without overwhelming the decision-making process. (Riviere 2023, 122)

9. Conclusion

This thesis has explored the multifaceted nature of game audio mixing, emphasizing how it blends technical precision, artistic intent, and interactivity. Unlike linear media, where a mix follows a predetermined structure, game audio is inherently non-linear, requiring a dynamic approach that adapts in real-time to player actions, environmental changes, and gameplay events. Through an analysis of mixing principles, workflow strategies, my own experiences and real-world case studies, this research has demonstrated how a well-executed mix can enhance gameplay clarity, immersion, and emotional impact.

While much of the discussion has focused on frameworks, methodologies, and industry practices, it is equally important to acknowledge that mixing is not just about adhering to rules. Some of the most striking and memorable soundscapes in games come from intentionally breaking conventions, whether through exaggerated compression, extreme loudness, unconventional or surprising solutions, or prioritizing chaos over clarity. A mix does not always need to be polished or balanced—sometimes unpredictability, or overwhelming intensity can be used as creative tools to reinforce a game's unique aesthetic and emotional impact.

This reinforces the idea that there is no singular "correct" way to mix a game. Some projects require subtlety, while others demand loudness, visceral aggression, or imaginative and surreal audio treatments. The craft of mixing lies in knowing when to follow conventions and when to break them to serve the unique identity of a game. Whether using object-based audio for precise spatialization or a compressed, overdriven mix to create relentless intensity, the role of mixing extends beyond technical execution—it is a fundamental part of game design and artistic expression.

In conclusion, this work has aimed to bridge the gap between academic study and industry practice. My goal was to examine the vast topic of game audio and mixing from multiple angles, combining technical insights with broader creative and systemic viewpoints. As game audio continues to evolve, the tools and techniques available to sound designers will expand, offering new creative possibilities while also presenting new challenges. The future of mixing may see AI-assisted workflows, increasingly dynamic mix systems, and more convincing spatial audio solutions, all of which will push the boundaries of interactive sound design. However, at its core, the goal of

game mixing will always remain the same: to create an engaging, emotional, and immersive soundscape that enhances the player experience.

References

Aditya, J., Ramta, B., Avnish, K., KD, S. International Journal of Applied and Basic Medical Research May – August 2015. Part 5. ”A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. “The online version of the medical journal. Accessed 11.1.2025. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4456887/pdf/IJABMR-5-

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., and Forchhammer, S. 2019. “Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences. “ AES 147th Convention. Accessed 11.1.2025. https://www.researchgate.net/publication/336414398_Defining_Immersion_Literature_Review_and_Implications_for_Res

Andersen, M., Bridgett, R., Maged, K.R., Ridgeway, E., Björkman, E. 2024. “Creative Perspectives in Game Audio”. Personal notes. Workshop. Den Danske Film Skole.

Aro, E. 2006. *Tilaääni*. Riffi-julkaisut. ISBN 951-98245-6-1

Ashby, S., Kastbauer, D. 2021. “Approaching the 3D Audio Ideal | The Necessary Adoption of Object-Based Rendering”. YouTube video of Gamesoundcon 2020 presentation. Audiokinetic Youtube channel. Accessed 18.3.2025. https://youtu.be/Y2dm_ISSEtU?si=kp1ldZdPekTEBksH

Audiokinetic. 2024. “Monitoring the signal level”. Wwise Online User Manual. Accessed 6.3.2025. https://www.audiokinetic.com/en/library/edge/?source=Help&id=monitoring_signal_level#loudness_measurement_scopes

Audiokinetic. 2024. “Understanding HDR”. Wwise Online User Manual. Accessed 26.2.2025 https://www.audiokinetic.com/en/library/edge/?source=Help&id=understanding_hdr

Brandmeyer, A., Swedlow, N., Hertensteiner, M.,Crum, P. 2021. Benefits of immersive spatial audio on athlete reaction times in e- sports. In ACM CHI Workshop. EHPHCI: Esports and High Performance HCI, May 08–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

Broomhall, J. 2024. “Mixing AAA Videogames”. AudioTechnology Online Magazine Issue 93. Accessed 26.2.2025. https://www.audiotechnology.com/features/mixing-aaa-videogames

Burke, K.R., 'Hard Limitations and Soft Possibilities: A “Systematic” History of Early Video Game Sound Technology', in William Gibbons, and Mark Grimshaw-Aagaard (eds), The Oxford Handbook of Video Game Music and Sound, Oxford Handbooks (2024; online edn, Oxford Academic, 20 June 2024), accessed 18 Mar. 2025. https://doi-org.ezproxy.uniarts.fi/10.1093/oxfordhb/9780197556160.013.17

EBU. 2020. “R 128 LOUDNESS NORMALISATION AND PERMITTED MAXIMUM LEVEL OF AUDIO SIGNALS”. Online publication. Accessed 6.3.2025. https://tech.ebu.ch/docs/r/r128.pdf

Egenfeldt-Nielsen, Simon, et al. Understanding Video Games : The Essential Introduction, Taylor & Francis Group, 2012. ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/uahelsinki/detail.action?docID=1181119

Farnell, Andy. 2008. *Designing Sound*. Digital prepress run 0 edition. MIT Press.

Gouveia, David. 2013. *Getting Started with C++ Audio Programming for Game Development*. E-book. Packt Publishing Limited.

Halden, J., Couthier, L., Gumbleton, S., Read, A., Hanson, P. Everest, L., Scott, H. 2021. "*Returnal: How Its Glorious, Dark Electronic Sound Was Made (and making the most of PS5's new audio engine)*". A Sound Effect. Online Blog Post. Accessed 11.3.2021.
https://www.asoundeffect.com/returnal-game-audio/?utm_source=copy&utm_medium=website&utm_campaign=sshare

Hazelight, Eriksson, P., Sjöberg, J.E., Kristiansson, G., Mongeau, A.S. 2021. "Behind the Sound of 'It Takes Two' | A Q&A with the Hazelight Audio Team". Ak Blog. Accessed 16.3.2025.
https://www.audiokinetic.com/en/blog/behind-the-sound-of-it-takes-two-a-qa-with-the-hazelight-audio-team/

Huber, D.M., Runstein, R. 2018. *Modern Recording Techniques*. Routledge. 9th Edition. Ebook. Accessed 26.2.2025.
https://www-taylorfrancis-com.ezproxy.uniarts.fi/reader/read-online/19c648b2-8b76-42c0-a360-49e00f81303b/book/epub?context=ubx

Jacobsen, B. 2016. "Informant Diegesis in Videogames". Master's thesis. Aarhus University.

Kerr, C., Bridgett, R., D' Oliveria, B. 2018. "Building an immersive soundscape in Shadow of the Tomb Raider - full Q&A". Online journal. Accssed 23.2.2025.
https://www.gamedeveloper.com/audio/building-an-immersive-soundscape-in-shadow-of-the-tomb-raider---full-q-a

Laaksonen, J. 2013. *Äänityön kivijalka*. Second edition. Riffi-julkaisut.

Mangini, M. 2021. "Dune - Sound Design approach". The Vernacular. Online blog post. Accessed 26.3.2025.
https://markmangini.com/Mark_Mangini/Blog/Entries/2021/11/15_Dune_-_Sound_Design_approach.html

Mendonça, C., Rummukainen, O., Pulkki, V. 2015. 3D Sound can have a negative impact on the perception of visual content in audiovisual reproductions. The 21st International Conference on Auditory Display. Accessed 12.3.2025.
https://www.researchgate.net/publication/279808657_3D_Sound_can_have_a_negative_impact_on_the_perception_of_visual_content_in_audiovisual_reproductions

Meyer, B. 2022. "The Sound Design of Ghost of Tsushima: Horses". Online blog post. Accessed 20.2.2025.
https://bradleymeyer.com/wp-core/2022/09/10/the-sound-design-of-ghost-of-tsushima-horses/

Mirbod S.M., Inaba, R., Yoshida, H., Nagata, C., Komura, Y., Iwata, H. 1992. "Noise exposure level while operating electronic arcade games as a leisure time activity". Industrial Health. Scientific journal. Accessed 8.3.2025.
https://www.jstage.jst.go.jp/article/indhealth1963/30/2/30_2_65/_pdf/-char/en

Oberem, J. 2020. "Examining auditory selective attention: From dichotic towards realistic environments". Institute of Technical Acoustics RWTH Aachen University. Accessed 26.3.2025.
https://library.oapen.org/bitstream/handle/20.500.12657/57090/external_content.pdf;jsessionid=CC407AA8B3B5EF66E0938C8685EEE2D7?sequence=1

Previty, A. 2024. "Wwise HDR: Overview and Best Practices for Game Mixing". AK Blog. Accessed 26.2.2025.
https://www.audiokinetic.com/en/blog/wwise-hdr-overview-and-best-practices-for-game-mixing/

Riviere, A. 2023. *Game Audio Mixing*. E-Book. Available from: VitalSource Bookshelf, Taylor & Francis.

Rossing, T.D., Moore, R.F., Wheeler, P.A. 2014. *The Science of Sound*. Third Edition. Pearson Education Limited.

Smalley, D. 1997. Spectromorphology: explaining sound-shapes. Organised Sound: An international journal of music technology. 1997-08, Vol.2 (2), p.107-126. Accessed 13.3.2025.
https://www-cambridge-org.ezproxy.uniarts.fi/core/journals/organised-sound/article/spectromorphology-explaining-soundshapes/A18EBE591592836FC22C20FB327D3232?utm_campaign=shareaholic&utm_medium=copy_link&utm_source=bookmark

Stevens, R., Reybould, D. 2015. "The reality paradox: Authenticity, fidelity and the real in Battlefield 4". Article in The Soundtrack. Accessed 26.2.2025.
https://www.researchgate.net/publication/286640535_The_reality_paradox_Authenticity_fidelity_and_the_real_in_Battlefield_4

Toftedahl, M., Engström, H. 2019. A Taxonomy of Game Engines and the Tools that Drive the Industry. Division of Game Development University of Skövde Skövde, Sweden. DiGra. Accessed 24.1.2025.
https://www.researchgate.net/publication/337165844_A_Taxonomy_of_Game_Engines_and_the_Tools_that_Drive_the_Industry

Turner, J. 2016. "SGC16: Joonas Turner - The Design of Tormentor X Punisher".Sweden Game Conference 2016. YouTube Video. Sweden Game Arena YouTube channel. Accessed 12.3.2025.
https://youtu.be/5iTiEFC73uQ?si=nLLy3BccNwkR9OVy

Uzer, C. 2024. "Naming Convention Best Practices". Blog post. Accessed 2.3.2025.
https://canuzer.com/naming-convention-best-practices/

Wolf, M.J.P. 2012. *Before the Crash : Early Video Game History*. Wayne State University Press. ProQuest Ebook Central. Accessed 8.3.2025.
http://ebookcentral.proquest.com/lib/uahelsinki/detail.action?docID=3416474

Yewdall, David Lewis. 2011. *Practical Art of Motion Picture Sound.* Ebook. Taylor & Francis Group ProQuest Ebook Central. Accessed 24.2.2025.
http://ebookcentral.proquest.com/lib/uahelsinki/detail.action?docID=709490

Zlobin, D. 2021. "Ludic and narrative sound in games". Blog post. Accessed 23.2.2025.
https://uxdesign.cc/ludic-and-narrative-sound-in-games-3fc52dbbfee1

Appendices

Image 1.	Rockstar Games. 2019. Retrieved 30.3.2025. https://store.steampowered.com/app/1174180/Red_Dead_Redemption_2/
Image 2.	Naughty Dog. 2025. Retrieved 30.3.205. https://store.steampowered.com/app/2531310/The_Last_of_Us_Part_II_Remastered/
Image 3.	Capcom. 2019. Retrieved 30.3.2025. https://store.steampowered.com/app/601150/Devil_May_Cry_5/
Image 4.	Activision. 2022. Retrieved 30.3.2025. https://store.steampowered.com/app/1962660/Call_of_Duty_Modern_Warfare_II/
Image 5.	Recreated according to Denis Zlobin's model. Original version: https://miro.medium.com/v2/resize:fit:1400/format:webp/1*AtS0dAW28PAZJwQiJ7QquQ.png
Image 6.	Mean Mink. 2025. Retrieved 30.3.2025. https://store.steampowered.com/app/2004060/Wild_Rumble/
Image 7.	Hazelight Studios. 2021. Retrieved 30.3.2025. https://store.steampowered.com/app/1426210/It_Takes_Two/
Image 8.	Playdead. 2016. Retrieved 30.3.2025: https://www.theguardian.com/technology/2016/jul/05/inside-review-limbo-developer-playdead-xbox-pc-platform-puzzle
Image 9.	Sucker Punch Productions. 2024. Retrieved 30.3.205: https://store.steampowered.com/app/2215430/Ghost_of_Tsushima_DIRECTORS_CUT/
Image 10.	Kojima Productions. 2022. Retrieved 30.3.2025: https://store.steampowered.com/app/1850570/DEATH_STRANDING_DIRECTORS_CUT/
Image 11.	Screenshot from Audiokinetic Wwise ver. 2024.1.3. audio middleware.
Image 12.	Eidos Montréal. Retrieved 30.3.2025. https://store.steampowered.com/app/750920/Shadow_of_the_Tomb_Raider_Definitive_Edition/
Image 13.	Rob Bridgett's Triangle of Needs visualized.
Image 14.	E-Studio. 2017. Retrieved 30.3.2025. https://store.steampowered.com/app/500670/TormentorPunisher/
Image 15.	Screenshot from Audiokinetic Wwise ver. 2024.1.3. audio middleware.
Image 16.	Screenshot from Audiokinetic Wwise ver. 2024.1.3. audio middleware.
Image 17.	Screenshot from Audiokinetic Wwise ver. 2024.1.3. audio middleware.
Image 18.	Housemarque. 2023. Retrieved 30.3.2025. https://store.steampowered.com/app/1649240/Returnal/
Image 19.	Santa Monica Studio. 2024. Retrieved 30.3.2025. https://store.steampowered.com/app/2322010/God_of_War_Ragnark/
Image 20.	Game Science. Retrieved 30.3.2025. https://store.steampowered.com/app/2358720/Black_Myth_Wukong/
Image 21.	Screenshot from FMOD Studio ver. 2.01.23 audio middleware.
Image 22.	Screenshot from Audiokinetic Wwise ver. 2024.1.3. audio middleware.