Footsteps in video games

Anton Diemar Sherar

ms17106

Bachelor Thesis

Media and Sonic Communication

18/12/2020

Supervisor: Niels Böttcher

Keystrokes: 49246

Anton Diemar Sherar, ms17
18/12/2020

Abstract

This thesis aims to investigate footstep sounds in video games, and how they can be used in service to gameplay and player performance, specifically in online multiplayer games. It goes into the history of footstep sounds in games, as well as the challenges and limitations met by developers, and the resulting innovations. It explains how footsteps distinguish themselves in a multiplayer game compared to single-player games, with a perspective on enhancing gameplay. It encompasses interviews conducted on experts from the games industry, explaining their approach to footsteps and the challenges faced when developing audio systems in their respective games. It includes an analysis of these games, and proposes a footstep system taking direct inspiration from said games, concluding in a playtest to see how the system fares when used by players. As a consequence of footstep sounds being very low volume in nature, it is more difficult to test its effect on players in a realistic setting, where many other sounds grab the player's attention. It is concluded that while there are many areas where footstep sounds have varying degrees of impact on player performance, specifically with regards to localizing important objects in game, a highly controlled environment is required for proper testing, due to the nature of multiplayer games, where so many variables change with every test. Further evidence and testing is suggested for future studies.

Table of contents

Introduction

Throughout the history of video games, footsteps have slowly risen in both priority and quality with regards to gameplay. From synthesized 8-bit sounds in old games to high fidelity recordings in thousands of different variations and layers in present games, footsteps have come very far since the dawn of video games.

With video games becoming a big and profitable industry in the past decades, it is no surprise to anyone, that audio has become increasingly important. In recent years, online multiplayer games have become the most popular and profitable genre within the industry, and new paths of innovation and challenges have come along with it.

The motivation behind this study is primarily the authors' wish to gain a deeper understanding of the importance of game audio, and what roles it can fill besides cosmetics and storytelling, specifically with regards to footstep sounds, which is one of the most frequently used sounds in games:

> "... the pervasive presence of footsteps throughout the entirety of a game makes it hard to ignore and certainly is a decision that must be consciously made." (Kastbauer, 2010)

The author is currently working on an online multiplayer game and deemed it fit to use this game for testing the theories, sound designs, and implementations based on the research conducted. Hopefully, these studies can help future sound designers and programmers who work on audio in similar multiplayer games with gaining a better understanding of the different ways of improving gameplay through sound.

Problem statement

This thesis will seek to answer the following statement:

How can one grant players competitive advantages in online multiplayer games through footstep sound design?

To find the answer to this problem, the following sub-questions have been raised:

- *In which ways does sound play a different part in multiplayer games than single-player games?*
- *Which challenges do competitive games introduce for a sound designer?*
- *How complex can a sound designer make a footstep system in online multiplayer games without compromising more important gameplay-related elements?*

Delimitation

In researching a grand topic such as audio in video games, it is important to specify and focus on a narrow genre of games. This thesis will therefore only be focusing on aspects of games that are relevant with relation to the problem statement and associated sub-questions. Therefore, this thesis will primarily focus on online multiplayer games, specifically discussing elements that are relevant to footstep and foley within a competitive context.

Method

The author planned to find answers to the previously asked questions by following the following structure; First, by diving into the history of footstep sounds and the

technical limitations of the past, the author hoped to gain an understanding of the role of footstep sounds. Next, the author looked into the knowledge and implementations of footsteps as a gameplay mechanic, with a particular focus on multiplayer games. Afterward, a research section analyzing the different approaches to footstep sounds in three case studies was explored. Interviews were conducted, with some of the sound designers and programmers behind the games explaining their various approaches to footsteps in their games. Finally, applying all the research gathered throughout the project and analysis conducted on the case studies, the author sought to test out and implement various theories and designs in footstep sound. This was followed by a playtest and qualitative tests of the effects of these implementations, with regards to performance in a competitive game. Appended to the thesis are videos showing some of the implementations and sounds, as well as various documentation, such as the interviews conducted, and illustrations of some of the audio systems inside of the game and middleware engines.

1 Footsteps and gameplay

1.1 Technical limitations and the evolution of character sounds

As briefly described in the introduction, footsteps and their associated sounds have developed quite a lot throughout videogame history.
Ever since PlayStation popularized using a compact disc (CD) in the fifth generation of video game consoles, which allowed for a lot more storage space than the typical cartridge, the possibilities of higher complexities in games have increased exponentially. Perhaps audio has been one of the biggest benefactors,

transitioning from simple synthesized tone generated waveforms or very short audio samples, to more expensive and complex audio, procedurally generated or in the form of high-resolution recordings.
This turning point of using CDs in the 90s started the rapid growth of increasing allocation of memory and processing power to audio in games. With more memory came higher bit depths and sample rates, as well as the addition of digital signal processors (DSPs) such as reverb and other effects manipulating the audio signal in real-time. An example of DSPs importance with regards to footsteps is when you walk into a cave as a player in a game, and your footsteps change to match the boomy echo of the cave environment. Another example is when your character goes underwater and sounds are filtered the second you immerse yourself in the water. The benefit of manipulating the audio in real-time is the fact that the sound can be processed by the computer at the time of the event, so the sound designers don't necessarily have to make additional versions of each individual sound, but can adjust them with DSPs in the game engine.

1.2 Footsteps as a gameplay mechanic

Gameplay is a somewhat vague term, as it deals with the experience of a game, which can be quite abstract. Kristine Jørgensen, an associate professor at the University of Bergen, describes it as follows:

> "Gameplay is not a feature designed into the game alone, but an emergent aspect of interaction between the game system and the player's strategies and problem solving processes. In short, gameplay is how the game is played, delimited by the game rules, and defined by the dynamic relationship that comes into being when the player interacts with these rules."
> (Jørgensen, 2008)

Interaction and feedback are at the core of gameplay in games and as such, sound shares the same priorities. Sound is a great sensory stimulus to the player's consciousness and even to the subconsciousness, affecting the mental processes without the player even noticing. (Peerdeman, 2010) Also, as Patrick Ng, Keith Nesbitt, and Karen Blackmore's research from their studies in player performance shows, sound is often the most effective medium to convey information, as reaction times to auditory stimuli is often quicker than visual stimulus. An example of a real-world situation, in which auditory signals are essential, are ambulances that have a siren in addition to flashing lights to warn others of their approach. (Ng, Nesbitt, Blackmore, 2015) Therefore, due to the aforementioned limitations in older hardware, the sounds of footsteps in old games had very little association with actual footsteps, and most of the time their only role was to inform the player that they were successfully moving their character. (Kastbauer, 2011, 00:24)

Compared to now, footsteps can give the player feedback of varying degrees of importance, such as what type of movement (walking, running, sneaking, crouching, jumping, etc) they are attempting. Also, foley sounds, such as the clothing the player is wearing, give an additional layer of feedback and adds to the personalities and believability of the different characters in modern games. The use of foley, reproducing "fake" sound effects, comes from the film industry, and leaves the sound director with a lot of aesthetic choices as opposed to the original recorded sounds, which often don't have as much impact on the listener. (Peerdeman, 2010) Footstep and foley sounds in conjunction are an almost unavoidable part of most video games' sound design. As Damian Kastbauer, software product manager at Audiokinetic, puts it:

> "If you're working in games today, chances are good that you've recorded, implemented, or designed systems for the playback of character footsteps

> and Foley at some point during the course of your career. It's even more likely that you've played a game where, at some point during your experience, footstep sound wrestled your focus away from the task at hand and demanded your listening attention." (Kastbauer, 2010)

While there are numerous examples of audio affecting gameplay on a macro level, it has special relevance for choices taken on a micro level, since game audio is in particular influential for the player's individual choices and behavior. (Jørgensen, 2008)
Games such as the *Thief* series (Zevik, 2010, 5:37) and *Splinter Cell* series (Fallen Foggy, 2019, 7:05) emphasized footsteps as a mechanic, and as an integral part of the gameplay. Both games are generally about sneaking around, and paying attention to the sound the player's character is emitting. Splinter Cell even includes an indicator of how much noise the player is making, which is an interesting example of visuals supporting audio. With footsteps being the most prevalent sound in most games, it makes sense to think of its usage in a broader perspective;

> "(...) for a sound type that may be heard for countless hours across every level in a game, surely they deserve more than a passing thought." (Kastbauer, 2010)

Since our auditory perception has evolved to deal with sounds in a context in natural environments, our auditory system is attuned to interpreting and filtering sounds in relation to specific contexts. (Jørgensen, 2008) As such, forcing the player to use their auditory perception along with their visual perception increases the potential of gameplay and immersion greatly.
The music researcher Sander Huiberts says, in investigating the properties of the auditory domain and the visual domain from a usability-related perspective:

> “(...) sound exists in time and over space and vision exists in space over time. Consequently, audio is very suitable for presenting time- based information with the advantage that the information is perceived even though the player is not immediately next to the source. With communication based on the visual domain, there is a chance of missing this information because another object on the screen is attracting his attention.” (Huiberts, 2010)

One of the main challenges in integrating audio as part of the gameplay is if the audio does not respond well to the gameplay and therefore disturbs the player's flow.
As memory and voice count has increased, the ability to layer sounds in numerous ways, as well as recombine them at runtime by the audio engine, has become a reality that many sound designers have been quick to adopt. Layering sounds and adding randomization keeps content size down by having elements shuffled on top of each other, with an exponentially increasing amount of combinations allowing for endless variation. In parallel to these audio aspects, animation and depth of player control have also seen a lot of development, which altogether has helped bring footsteps and foley closer to reality. Examples include additional step types beyond walking and running, such as crouching, as well as clothing types and surface materials upon which the player steps. (Kastbauer, 2010) Interestingly, a lot of these early attempts to infuse the sound with variation and detail have been with immersion and believability in mind. Online shooters such as Rainbow Six Siege use these highly granular audio systems as important information to the players in relation to each other.

One of the key differences between footstep sounds today as compared to the past, is its role as less than a notifier of active actions of the player and more as a

notifier of the other characters around the player. This is especially prevalent in multiplayer games.

1.2.1 Multiplayer games

In most online multiplayer games, the primary role of sound is helping each player with localizing each other, and providing important feedback about where the player is within the game world. Developing map awareness takes time, and different audio cues will help the player in distinguishing between different variables such as zones on a map, weapon projectiles, or types of enemies. Compared to single-player games, where storytelling and immersion are vital, multiplayer games are required to prioritize in a much stricter manner, focusing on feedback and information above everything else. This is especially true for sound, which can prove to be ineffective in outputting multiple auditory messages simultaneously because users can have difficulties focusing on many sources of information at the same time. However, one of its biggest strengths is sound's capability of positioning an object that is not visible on screen in the game world. (Huiberts, 2010)

Walter Murch, one of the pioneers of modern sound design, has a famous ruleset called the "law of two-and-a-half" from his lecture "Clear Density, Dense Clarity". This "law" attempts to answer some of the problems that sound designers face when balancing a mix.
Roughly speaking, based on the left-right duality of the human brain, Murch explains that sounds, in general, can be spread out across a spectrum of color, one side of the brain being violet, representing Encoded sound (dialogue) and the other being red, representing Embodied sound (music). In between these two poles exist

linguistic effects, such as footsteps, and “musical” effects, such as atmospheres and room tones (see appendix A).[1] Murch argues that:

> “(...) generally speaking, the two departments [of the brain] seem to be able to operate simultaneously without getting in each other’s way. What this means is that by dividing up the work they can deal with a total number of layers that would be impossible for either side individually.” (Murch, 2005)

Therefore, attempting to blend more than two instances of the same sound, i.e. footsteps, can be redundant, due to our brain’s incapability of processing more information of the same source at once. However, blending different parts of the spectrum of sounds, such as music, dialogue, and sound effects, can work together quite nicely, up to a maximum of five layers:

> “(...) at any one moment (for practical purposes, let’s say that a ‘moment’ is any five-second section of film), five layers is the maximum that can be tolerated by an audience if you also want them to maintain a clear sense of the individual elements that are contributing to the mix. In other words, if you want the experience to be simultaneously Dense and Clear.” (Murch, 2005)

While Murch comes from the world of film, his points can apply just as much to videogames. For these reasons, multiplayer games, and the audio middleware that they use, often have implemented several systems that help players get important feedback during the game. The degree to which footstep sounds can help a player is limited since they sit quite low in the mix due to their repeating nature, but some games have implemented some subtle uses of footsteps as a gameplay tool, such as the Call of Duty games. In Modern Warfare 2, one of the perks (in-game

[1] Appendix A is a visualization of Murch’s aural spectrum

modifiers) players can unlock is “Ninja Pro”, which reduces the player's own movement noise volume by 75%. Opposite the “Ninja Pro” perk, there is the “SitRep Pro” which allows the player to hear other players’ footsteps louder. (DefaultDadPlays, 2010, 1:08, 2:31) These are examples of simple volume adjustments with influence on gameplay and loadout capabilities, which makes you wonder what possibilities more advanced changes in sound can lead to. In integrating game audio, these systems can help support the game, not only as an information system and support for gameplay, but also by providing an understanding of how the game should be played, and how to behave in a specific in-game context. It is this context that is essential when deciding which role is the most prominent. In *World of Warcraft (WoW)*, Kristine Jørgensen analyses the game mode Battlegrounds, where she explains why a certain sound signal cannot be categorically associated with one specific function. As an example, she uses the situation in which a player casts a friendly spell on someone. The spell has several functions: It may be a responsive signal that confirms a player casting the spell; it may identify a change in player state, and it may orient that player with respect to other players. Which function is most important depends on the context the player is in. Therefore, particularly in multiplayer games, it is important to understand whether a sound relates to an event that directly influences one's player or not.

> “This demonstrates the importance of understanding the context of a specific sound before being able to understand which of the different informative functions is most relevant at a specific point in time.” (Jørgensen 2008)

Context, therefore, dictates priority and the hierarchy of importance in sounds. Jørgensen further explains figure-ground segregation as an important feature

when in chaotic situations. For instance, enemy generated sounds distinguish themselves functionally from ally generated sounds in terms of usability, in the sense that enemy generated sounds tend to provide a higher degree of urgency and negative information to the player than ally generated sounds. The reason is that all actions that the enemy takes are potential threats towards the player and his team. (Jørgensen 2008) The ability of the human brain to group and organize perception into background and foreground information is a natural occurrence, and so as sound designers, we can help the player along by implementing simple dynamic changes, such as increasing or decreasing the volume of sounds equal to their importance to the player within the given context. Speaking of dynamics, an important, yet often overlooked design feature of sound in game worlds is silence, which when used correctly can improve focus and concentration, as a sudden silence can make our hearing very alert. (Huiberts, 2010)

Another important tool used in a lot of multiplayer games is sound obstruction and occlusion, which relates to the point about map awareness and information. In competitive games where players' main objectives are eliminating each other, each player's awareness of their surroundings and ability of positioning other players is essential to success. Say for example a player is listening actively for players around a corner they aren't able to see. If a sound is only affected by the distance between itself and the player, there would be no way for the player to know whether another player was behind a wall or directly next to them. This is where obstruction and occlusion can help.
We can break positioning down into three factors: Volume, time & anatomy (our ears). Volume and timing of sounds can easily be adjusted by the game engine, but emulating our ears' anatomy is a bit more tricky. Since game audio isn't in an actual physical world, sounds won't be affected by their surroundings like in the real world. Therefore, sound obstruction and occlusion are an attempt to emulate

sound physics in “closed” environments by changing the sounds dynamically in-game. Various parameters, such as the wall types and differences in distance and elevation between the sound source and the listener, act as barriers that help calculate the type of occlusion happening to the sound. Roughly speaking, obstruction has to do with the direct path between a source and a listener only, whereas occlusion also affects the indirect paths (reflections) to the listener. A great example of occlusion in competitive games is *Counter-Strike*, where every bit of information is crucial to the players’ survival. (3kliksphilip, 2017, 00:19 - 00:31). In conclusion, context-specific clues and dynamic changes in audio can elevate the player’s awareness when their eyes might be busy focusing on more urgent tasks, such as other players, which eventually leads to better performance. In addition, reflection adds additional depth in awareness, by dynamically rendering reverb based on room size or other aesthetic choices made by the developers. It can be quite expensive, especially when rendering early reflections.

2 Research

2.1 Case studies

To get a better understanding of the different approaches to footsteps in games, as well as their usage in a competitive regard, the author has conducted interviews with different sound designers and programmers from three different games from the industry. Two of the cases studied are competitive multiplayer games, that examine footsteps in an online competitive context, while the last case, a narrative-driven single-player game, examines footsteps in direct integration with visuals and gameplay.

2.1.1 Rainbow Six Siege

Rainbow Six Siege (*R6*) is an online tactical FPS video game developed by Ubisoft Montreal, in which footsteps have arguably the most important role in a game with relation to gameplay. *R6*, being a first-person game, is a great example of audio being essential to providing feedback on screen, as the player rarely sees their own feet, and is constantly in danger of shootouts with other players just around the corner. Nicolas Teplitzky, technical audio director on *R6*, explains:

> "(...) we knew that foley would have a huge importance in this game. First because we wanted to be as realistic as we could to support the tactical aspect of the game. Then we knew that in a game that is extremely lethal, foley would be one of the only things you would hear until you die. (...) In terms of resources, we recorded pretty much everything with our foley artist. The scope of the game is quite small. We didn't have a gazillion of movements or materials to cover so [we] had the opportunity to focus on details."[2]

Antoine Laloy, an audio designer on *R6*, explains that the biggest priority for sounds in regards to competitive gameplay were the dynamics in the mix and the sound propagation (simulating sound physics, such as reflections, obstruction and occlusion). (Ubisoft North America, 2015, 1:22 - 2:34) A focus on dynamics helps the high intensity of the game, which is one of the core aspects of *R6*, and sound propagation aims to give proper feedback as well as serving the tactical gameplay.[3] Teplitzky adds on to Laloy's point about propagation:

[2] Appendix B - Interview Nicolas Teplitzky
[3] Appendix C - Interview Antoine Laloy

> "We're (...) using a propagation system that is repositioning sound sources based on the geometry configuration. This was a bold decision for a competitive player and even though the system is not as precise as we would like it to be, it was one of the condition[s] to stay[ing] realistic."[4]

Special attention was put to the realism and amount of details in *R6's* navigation sounds such as footsteps, in order to help the player gain more information, simply out of listening to other players characteristics, as they navigate through the map: their weight, armor, and speed. (Dion, 2017) Similar to the "Ninja Pro" perk from Call of Duty as mentioned earlier, *R6* features certain characters with skill-sets that are similar, reducing footstep noise and allowing for easier stealth tactics.

Matching the tactical aspects while staying true to the goal of realism was one of the biggest challenges the audio team faced. *R6* allows for procedural destruction of the environment, and while this opens up a lot of tactical opportunities, it also comes with several challenges in terms of performance:

> "Running at 60fps with a procedural destruction system was what shaped how we approach sounds in the game. (...) The compromises we had to make were in regards to our propagation system because this is what drives the polyphony in the game for the most part."[5]

Dynamically changing the way a sound can be propagated, having obstruction, occlusion and reflections responding to the increasing destruction of walls and ceilings, can be quite expensive with the constant movement of players. On top of this, in order to ensure that all players are in complete synchronization with each other, everything between the servers and players have to be replicated, which is

[4] Appendix B - Interview Nicolas Teplitzky
[5] Appendix B - Interview Nicolas Teplitzky

also a difficult task. Due to these limitations, while wanting to run at a high frame rate, the team had to make some compromises. Teplitzky explains that for sounds other than weapons, they couldn't afford having two propagation paths being computed. Another limitation was no multi positioning of sounds, i.e. using one instance of sound at two positions. The only reverb that is played dynamically is the one from the player, and this was an issue with regards to the goal of rendering a sound in as realistically an environment as possible. A workaround the team used was simulating reverb, by baking the reverb into the weapon sounds at the gun position, which helped a lot. However, no early reflections were used, as they couldn't afford it.
Still, despite these limitations, the team managed to stay true to the high level of intensity and tactical gameplay, and implementing dynamic mixing helped tremendously:

> "The whole concept for the mix was to manage the awareness of the player in those close quarters environments. [The] Player had to be conscious of his own level of noise and he should be able to estimate how other players would detect him. Also, obviously, we had to make him able to detect, and if possible locate threats."[6]

Following this concept, setting priorities and establishing some form of hierarchy within the sounds allows the audio middleware to prioritize in situations where sounds are fighting with each other in the mix:

> "In terms of priorities, weapons are king (...) Then comes everything that is a potential threat (enemy's movements and gadgets). Teammates sound is what we are discarding the most easily when we reach max polyphony."[7]

[6] Appendix B - Interview Nicolas Teplitzky
[7] Appendix B - Interview Nicolas Teplitzky

Max polyphony (amount of sounds playing simultaneously) is used for performance reasons, but it can be argued that it also represents Walter Murch's "law of two-and-a-half", in that it helps the player focus, as having both enemies and allies have the same volume in the mix would only lead to confusion for the player.

Louis-Philippe Dion, former audio director for *R6*, explains in an article diving deep into the sound of the game:

> "the best element that we had was the sound of the other players who you could not see. So we removed all "imposed" emotion-giving sounds to focus on what really mattered: the sound created by players. (...) Keeping the experience void of artifacts, to me, gives Siege a sound print that is not only fun to listen to, but also that influences the game greatly." (Dion, 2017)

Watch a video showcasing the sounds of footsteps in *R6*. (W. K., 2017, 0:22 - 0:40)

2.1.2 Overwatch

Another online competitive game that deals with priority of sounds is the team-based FPS *Overwatch* developed by Blizzard. Among other audio systems, *Overwatch* prioritizes based on threat levels that dynamically change depending on the context the player is in. In their 2016 GDC talk, Blizzard's Scott Lawlor, audio director, and Tomas Neumann, senior software engineer, explain their thoughts behind how they achieved *Overwatch's* dynamic audio landscape. Compared to *R6*, *Overwatch* isn't aiming for high levels of realism, and so their approach to variation in sound and VO (voice over) is much more minimal.
Lawlor explains they draw inspiration from the Pavlovian response, in other words classical conditioning. Classical conditioning is when a conditioned stimulus (for

example a sound effect) becomes associated with an unconditioned stimulus (some sort of reflex) in order to produce a behavioral response, the conditioned response. In other words, we can condition the players, and elicit certain responses and reactions from them by playing specific sounds. (GDC, 2016, 38:17)

In attempting to get a clear mix, the fundamental question they asked themselves was; “Who is your greatest threat?”. The sound designers at *Overwatch* were directly inspired by Murch’s theory, and this resulted in, among other things, the importance system. Simply explained, the importance system tells the audio middleware engine how to prioritize each player in relation to each other, based on different parameters in-game. For example, if there’s one enemy near the player, that enemy will have a “threat score” of 40. If we introduce another enemy next to the player, who is aiming directly at them, perhaps dealing damage to them, they will have a score of 120 (see appendix D).[8] These threat values put the players in respective buckets[9], that then get sent to the middleware, *Wwise*, where they can be easily prioritized in volume, pitch, filtering and voice count[10]. In a game where there are six players on each team, sounds can get drowned in the mix quite easily. As such, the importance system helps the player with getting clarity in high-intensity situations. (GDC, 2016, 06:25 - 14:37) A video example shows the effects of the importance in-game. (GDC, 2016, 16:08)

Another way of giving the player more clarity was giving each character a unique sound, such as clothing, jewelry or other personal items distinct to that character. These sounds play on top of the footsteps and help players plan out what kind of tactical approach they should take before visually knowing who they’re next to. A

[8] Appendix D - Overwatch screenshots from GDC: Threat level
[9] Appendix D - Overwatch screenshots from GDC: Bucket index
[10] Appendix D - Overwatch screenshots from GDC: Importance RTPC’s

video example shows the effects of unique footstep sounds depending on the character. (GDC, 2016, 32:54 - 33:54)

2.1.3 INSIDE

INSIDE is a puzzle-platformer adventure game developed by Playdead. In contrast to *R6* and *Overwatch*, *INSIDE* is a single-player game, with a heavy emphasis on narrative and adventure, and the game's audio is closely integrated with gameplay, with some of the puzzles in the game being directly reliant on aural cues from footsteps, such as the march scene (see appendix E[11] and a video demonstrating the aural cues (Prokrast Nation, 2018, 1:54 - 2:04)) Jacob Schmid, audio programmer at Playdead explains;

> "The most important goals in making the footstep system in INSIDE were to have context sensitivity and integration with visual effects, simultaneously making sure that the system was user friendly for the sound designers and easy to troubleshoot."[12]

Footstep and foley sounds depend on the context in which the main protagonist, the boy, is in. Sounds are chosen based on an analysis of the current animation, the preceding sound (in order to support more complex actions) and the boy's game-logical condition, i.e. whether he is in contact with the ground, or has been in water recently. With visual elements being tied to the audio, which is influenced by the emotional state and location in the game, the audio team had to make the process of adding sound events seamless, as well as easy to debug, in order to work as efficiently as possible. Schmid describes how analysing game logic and animation states, along with having a playback component that uses these data, is

[11] Appendix E - INSIDE screenshot from GDC: Audio driven gameplay (march scene)
[12] Appendix F - Interview Jacob Schmid [translated from Danish]

a usable solution in simplifying complex game systems. An example of the audio system can be found here (Audiokinetic, 2016, 15:39 - 17:10).

In summary, from the research gathered from the case studies, there are multiple ways of approaching footstep sounds in video games, and each approach requires a different way of thinking in terms of what's important to the specific game. One common trait in all the case studies is the fact that footsteps and their role in the respective games have been considered and iterated upon more than one would think, and it can therefore be beneficial to think about them as they can lead to great results.

2.2 First Class Trouble

To help answer the problem statement and its related sub-questions, several footstep sounds and systems were designed and implemented by the author, based on the research and analysis conducted previously. The system was integrated into the co-operative online multiplayer game *First Class Trouble* (*FCT*), developed by Invisible Walls. Afterwards, in order to test the results of these updated systems, a playtest along with a survey was conducted among a selection of the game's playerbase.

2.2.1 What is First Class Trouble?

FCT is a game about deception, where two out of six players are personoids, who have to deceive their way to victory. This can be done in multiple ways, either by killing off the other players or by avoiding detection of their identity. The other four players are called residents, and they have to cooperate and find out who the personoids are. While not an FPS like *R6* and *Overwatch*, *FCT* does have some

gameplay elements in common with these titles. For example, the competitive aspects of either avoiding threats, or being the threat and attempting to hunt other players. Players have different ways of figuring out who the personoids are, either by witnessing them killing other players, or by deducing through conversations and other events happening in-game. One tool all players have available is the ability to personally vote every other player either up or down, i.e. signalling whether or not they think someone is trustworthy.
Games vary in length, and the further the game goes, the more the intensity and tension increases. This is why auditory information can be an important factor in deciding a player's outcome in surviving.

2.2.2 Design and implementation of footstep systems

Before implementing the gameplay responsible systems, the old audio solution had to be re-done in order to match the requirements in terms of variations of footsteps and foley sounds (see Appendix J for a video example of the old footstep sound system[13]). After purchasing a bundle of footstep sounds in numerous variations on different surfaces, the sounds were imported into *Wwise*, the middleware solution for *FCT*, edited and placed into different hierarchies of switch containers. Switch containers allow you to group objects according to different alternatives that exist within a game, such as surface types. The hierarchy of switch containers was organized based on what changed the most often in-game: shoe type, movement mode and finally surface type.
Notifiers within the animations used in the game were also fine tuned, so their sync points lined up perfectly with playing back the sound. Additional notifiers were also added, expanding on the amount of movement modes from one movement mode to walk, run and sneak, as well as jump, land and stop. In addition to having the

[13] Appendix J - FCT_Footstep_old_VideoExample

movement mode setting a switch, a game parameter (RTPC) was attached to the switch, which allowed for smoothly changing different parameters outside of switch containers. This RTPC, representing movement speed, as an example increasingly changes the pitch of jumping and landing sounds according to the current speed or movement mode of the character (see appendix G).[14]

In addition, another RTPC was added based on the character's gender, adding an EQ with a small boost in the lower frequencies, symbolizing the slightly heavier male character. (see appendix G).[15]

Finally, new surface types, or physical materials as they are called in Unreal Editor, were added, making the total surfaces in the game consist of carpet, concrete, glass, gravel, laminate, leather, metal, sand, tile and wood. Sub-materials have been added as well, including broken glass and liquids, which can be layered on top of the primary surface materials for increased variation.

An RTPC called wetness was also added, taking inspiration from *INSIDE*, which increases when a player has either been pushed into water or stepped in a puddle of liquid. This wetness causes the player's shoes to sound wet, and this effect reduces over time.

Additionally, characters can customize their outfits in the game, and, taking inspiration from *Overwatch's* goal of giving each character in the level a unique defining sound, this allowed for even more variation, with different clothing categories layering on top of the footsteps. For footwear, ankle boots, regular boots, dress shoes, high heels and sneakers were added. For the lower body, cargo pants, jeans, jogging pants, skirt and suit pants were added. And finally, for the

[14] Appendix G - FCT Wwise screenshots: Character movement mode switch + movement RTPC
[15] Appendix G - FCT Wwise screenshots: Weight RTPC EQ

upper body, denim jacket, dress, hoodie, regular jacket, leather jacket, shirt, suit jacket and t-shirt were added. All individual clothing had a distinct material or weight, and this should be reflected in the character's sound.
While cosmetic improvements are very subtle with regards to competitive gameplay, the one thing they can add is a better sense of feedback , which can improve the players reaction within certain contexts in addition to improving their deduction capabilities, such as being able to distinguish between running and walking (see Appendix J for video examples of the new footstep audio system[16] and Appendix K for standalone sound examples[17]).

Taking inspiration from *R6* and *Overwatch's* distinction between allies and enemies with regards to importance of sound, one way of granting players competitive advantages in a less subtle manner, would be to let the player's vote change their individual mix of every other player's individual priority, such as filtering and changing volume. This was implemented through an RTPC called "threat level" which is attached to every other player (see appendix G).[18]
Implementation wise, a blueprint (gameplay scripting system inside Unreal) was created, which sends information to Wwise, whenever a delegate is called, which basically means that it listens anytime a player uses the voting mechanic (see appendix H).[19] (see Appendix J for a video example[20])

Another example of dynamic mixing on the player's part connected to gameplay choices, is a "killer" state. When playing as the personoid, if the player picks up the murder weapon and equips it, their own footsteps will reduce in volume, and others

[16] Appendix J - FCT_Footstep_new_VideoExample + FCT_Stop_VideoExample
[17] Appendix K - FCT Sound files
[18] Appendix G - FCT Wwise screenshots: Threat level RTPC's
[19] Appendix H - FCT Unreal Editor screenshots: Voting
[20] Appendix J - FCT_VotingUpDown_VideoExample

will increase. This helps the player with focusing on listening to other players (see appendix H).[21]

2.2.3 Survey results

A preliminary survey of five people followed by a playtest and a post-game survey were conducted on volunteering subjects from the *FCT* community. The main purpose of the test was to verify whether or not the system had any effects on improving the players' performance. The subjects were not aware of the purpose of the test, except that they would be testing some new features. Some of the questions in the preliminary survey might've indicated some of the systems they would be testing, as it included questions related to footsteps.
In the preliminary survey before the playtests, subjects were asked the following question; "How often do you take notice of footsteps in video games?" to which all respondents answered somewhat similarly, between "A decent amount" to "Always". Subjects were also in agreement when answering a related question; "How often do you use the sound of footsteps to gain a competitive advantage? (locating others)", with one subject saying: "Most games don't have the best directional/range hearing, but I do take note if there are multiple sets of footsteps or from where they are coming from." This supports the argument of footsteps importance in multiplayer games, and their tactical use by players.
When asked about difficulties faced when attempting to focus on an individual player in climactic situations, subjects had mixed responses, from "Not very difficult." to "Difficult if everyone is trying to get their info out at once." Subjects had varying degrees of difficulty with regards to locating others, with 32% of the respondents' total answers responding as such. One reason why there's a discrepancy in difficulty between each player could be the varying knowledge of

[21] Appendix H - FCT Unreal Editor screenshots: Killer state

the game's levels and systems.

Finally, subjects had mixed opinions with regards to difficulty in doing tasks stealthily as a personoid, with one respondent saying: "Easy and difficult depending on the map ([If] It is a larger map it is easier, but when the map is smaller and everyone sticks together, it is difficult)". Due to there only being a 33% chance of playing as a personoid in every game, players might not have as much experience in that role as opposed to playing as a resident.

After playing three full games, the subjects were asked questions relating to their experiences compared to when they had originally played *FCT*. Due to time and resource constraints, playing the game without the implementations wasn't possible, but each subject had played the game multiple times before, and had a good understanding of the game and it's sounds.

The first question subjects were asked after playtesting was: "Did you notice a difference in footsteps from what you're used to hearing in First Class Trouble?", to which 40% of subjects responded positively, which one subject saying the footsteps were louder as compared to before. However, only 20% felt like they used footstep sounds strategically more than usual, with one respondent saying: "Footstep sounds weren't loud enough for me to notice over other in game sounds." This is a result of the nature of footstep sounds being very low in the mix, and thus making answers related to them in an actual game context quite difficult to measure. Over half of the subjects felt locating other players easier, however, with 60% responding positively. The same percentage of subjects responded positively with regards to high intensity situations. This is most likely due to the voting system, as it allowed for up to a 12 dB difference in volume between two other players, depending on the votes they were given by the player. While quite substantial, this has shown to help the players during the course of the game.

In the first playtest, subjects were instructed to play as they normally would. Before the second and third playtests, subjects were asked to use the voting system deliberately, and this may have helped the subjects noticing a difference. Finally, 60% of players played as a personoid, and among these, 66,6% felt an improvement in being stealthy.
Overall, every subject noticed a difference in the experience and performance. There was a difference in the amount noticed by each respondent, however, with approximately 44% of the respondents' total answers noticing a difference.
Due to time constraints, the sample size was quite small, which increases the margin of error in the study. Additionally, previous studies on player performance have had mixed results as well, as outcomes and performances depend on multiple variables (see appendix I).[22]

In their study of player performance in MOBA games, Patrick Ng, Keith Nesbitt, and Karen Blackmore tested the effects of sound on player performance in the game *Dota 2.* They tested both individual and team performance levels in two different sound conditions (on and off) using seven individual and four team measures. Out of the four team measures, three showed improvements in performance when sound was present, while only one measure recorded a reduction in performance. They conclude that the mixed results in this experiment may be a result of the difficulty in creating a controlled environment in *Dota 2*. This can be applied to other competitive games, where so many variables change from game to game. While similar patterns occur during games, no two games are exactly the same. Outcomes and performance depends on team strategy, individual hero choices and how the two teams engage and interact on the map.

[22] Appendix I - FCT Survey Responses

> “More work needs to be conducted on the experimental design to better measure the effects of sound on player and team performance in Dota 2 and other similar team based video games.” (Ng, Nesbitt, Blackmore, 2015)

Discussion

While it seems like there is a correlation between footstep sounds and performance, there is not enough data to support this conclusion. As such, more testing is required, and the conclusions by Ng, Nesbitt and Blackmore, in studying *Dota 2*, along with Jørgensen, in her study of *WoW’s* Battlegrounds, can also be applied to *FCT*, which relies heavily on player interaction and team strategy. Only the current version of the game’s footstep system was tested, with the subjects relying on their memory of previous systems for comparison. Also, with a very low sample pool of five people, along with the subtle and less noticeable nature of the sounds and implementations, the overall margin of error was quite large. In order to get data less prone to error, further testing will require a much bigger sample size, with more types of respondents, including people who have less experience with the game.

Due to time constraints, only a couple of interviews managed to be conducted, and this could be expanded upon for further studies. The amount of genres of games analysed could also have been expanded, with social deduction games specifically being noteworthy, especially in regards to the audio systems implemented in *FCT*.

Conclusion

In conclusion, there are many reasons to believe that gameplay and player performance can be improved by foley and footsteps, and with the emergence of competitive games worldwide, optimising and gaining every advantage possible is more relevant than ever before. With regards to how complex a sound designer can make a footstep system, it all depends on the scope of the game and the goals of sound. Kastbauer sums it up perfectly:

> “There is no magic number when it comes to deciding at what point you have enough variation, often it becomes a complicated equation of space vs. perceived need for diversity” (Kastbauer, 2010)

Still, audio is particularly useful in providing information to the player, when their access to visual information is blurry, or restricted by other more important elements. This is important, especially for three-dimensional games, where events often happen offscreen. When a chaotic situation is present, visual information may be hard to grasp, and audio can pick up the missing pieces of the puzzle the player is trying to solve in the given situation. In certain complex situations, audio may also help the player pick up more information than the visual system could do alone. Kristine Jørgensen explains:

> “(...) while visual information can be shut out by closing the eyes, audio has no equivalent shut-down mechanism. Audio is therefore an omnipresent feature which is easy to forget is present at all.” (Jørgensen, 2008)

Peel supports this notion:

> "Sound is one of those things that I think most of us, as humans, don't really pay much attention to (...) When we make a game, we are starting from scratch. We have to manually recreate all of those little sound cues that you might not notice in your daily routine. But they're absolutely critical in giving your brain a complete picture of the world that surrounds you." (Peel, 2020)

Future perspective

When asked about his thoughts on where he sees footstep sounds in games of the future, Jacob Schmid explains:

> "I think footstep sounds are well-functioning in newer games, except from first-person games. These games suffer under some fundamental problems, which can be hard to solve."[23]

Schmid describes 3D placement as the first challenge:

> "a) 3D placement: In first-person games, especially VR games, footsteps should be unambiguously placed at the player's character's feet to avoid any confusion about whether they're coming from the player's character or another source. This requires functioning 3D sound, which in principle requires a model of the player's ears to be emulated with a HRTF [head-related transfer function; a response that characterizes how an ear receives a sound from a point in space]. A fun attempt at a solution is to place the higher frequencies of the footstep sounds in the controller speaker on a PS4 and a PS5, since the controller is usually held at a certain distance beneath the ears of the player."[24]

[23] Appendix F - Interview Jacob Schmid [translated from Danish]
[24] Appendix F - Interview Jacob Schmid [translated from Danish]

Another challenge is cohesion between controller input and the character's animation:

> "b) Cohesion: An even bigger issue is the link between the player's controller input and the character's animation, which in a first-person perspective isn't obvious. In third person games, the player sees an animation, which is an interpretation of the player's intention with the controller input. A footstep sound belongs to this animation, which results in a clear cohesion to the player. In first-person games, the player usually doesn't look at their character's feet, and therefore they have no precise feedback on how their feet are moving in relation to their controller input. Without this cohesion, a footstep sound could be perceived as belonging to a different character or as a completely unrelated sound, which would be a completely failed system.
>
> Even if problem a would be solved with good 3D sound custom tailored to the player's ears, problem b will still exist, as long as we use the simplified intention indicating controller input. Problem b can unfortunately only be solved with a new form of controller input. Until then we must focus on the issue of designing footstep sounds, that don't call too much attention to itself.``[25]

With regards to the future, the footstep audio system in *FCT* will be tested further, expanding upon the possibilities of improving performance and gameplay. Based on player feedback, and seeing the effects of the systems, more will be implemented, and current systems will be iterated upon.

[25] Appendix F - Interview Jacob Schmid [translated from Danish]

Reference list

3kliksphilip (2017, October 24) CS:GO New Audio Occlusion Comparisons [Video file]
Retrieved from https://youtu.be/3DpQjjqSfzE?t=19

Audiokinetic (2016, October) Wwise Tour 2016 - Playdead INSIDE (2 of 3) - Voice [Video file]
Retrieved from https://www.youtube.com/watch?v=gRF8Gt5hys4

DefaultDadPlays (2010, March 18) Reality's MW2 Perk Guide Ninja Pro vs SitRep Pro [Video file]
Retrieved from https://youtu.be/bFB1xj5FwZs?t=68

Dion, L.P. (2017, January 18) Game Design Deep Dive: Dynamic audio in destructible levels in Rainbow Six: Siege
Retrieved from https://www.gamasutra.com/view/news/288565/Game_Design_Deep_Dive_Dynamic_audio_in_destructible_levels_in_Rainbow_Six_Siege.php

Fallen Foggy (2019, January 5) Splinter Cell: Chaos Theory walkthrough 01 (Hard, Mixed style, No commentary ✔) [Video file]
Retrieved from https://youtu.be/uwVPnjOWZWo?t=425

GDC (2016, July) Overwatch - The Elusive Goal: Play by Sound [Video file]
Retrieved from https://www.youtube.com/watch?v=zF_jcrTCMsA

GDC (2016, July) Inside: A Game That Listens [Video file]
Retrieved from https://www.youtube.com/watch?v=Dnd74MQMQ-E

Huiberts, S. (2010, November) Captivating Sound - The Role of Audio for Immersion in Computer Games
Retrieved from https://download.captivatingsound.com/Sander_Huiberts_CaptivatingSound.pdf

Jørgensen, K. (2008, December) Audio and Gameplay: An Analysis of PvP Battlegrounds in World of Warcraft
Retrieved from http://gamestudies.org/0802/articles/Jorgensen

Kastbauer, D. (2010, March 16) Footsteps – Informal Game Sound Study
Retrieved from http://blog.lostchocolatelab.com/2010/03/footsteps-informal-game-sound-study.html

Kastbauer, D. [lost chocolate lab] (2011, February 3) Footsteps Retro Sound Study [Video file]
Retrieved from https://vimeo.com/19525536

Murch, W. (2005, April) Walter Murch
Retrieved from https://transom.org/2005/walter-murch/

Ng, P., Nesbitt, K., Blackmore, K. (2015, February) Sound Improves Player Performance in a Multiplayer Online Battle Arena Game
Retrieved from https://www.researchgate.net/publication/290542348_Sound_Improves_Player_Performance_in_a_Multiplayer_Online_Battle_Arena_Game

Oldenburg, A. (2013, September) Sonic Mechanics: Audio as Gameplay
Retrieved from
http://gamestudies.org/1301/articles/oldenburg_sonic_mechanics

Peel, J. (2020, March 19) How hearing became the most important sense in the modern FPS
Retrieved from
https://www.vg247.com/2020/03/19/rainbow-six-hunt-showdown-sound-design/

Peerdeman, P. (2010, April) Sound and Music in Games
Retrieved from
https://peterpeerdeman.nl/vu/ls/peerdeman_sound_and_music_in_games.pdf

Prokrast Nation (2018, March) Inside, Chapter 19, rhythm and dance [Video file]

Retrieved from https://www.youtube.com/watch?v=uBNzi7p8Y80

Ubisoft North America (2015, November) Tom Clancy's Rainbow Six Siege Official – Sound Propagation – Community Corner #3 [NA] [Video file]
Retrieved from https://www.youtube.com/watch?v=ti6ZWRSiqpQ

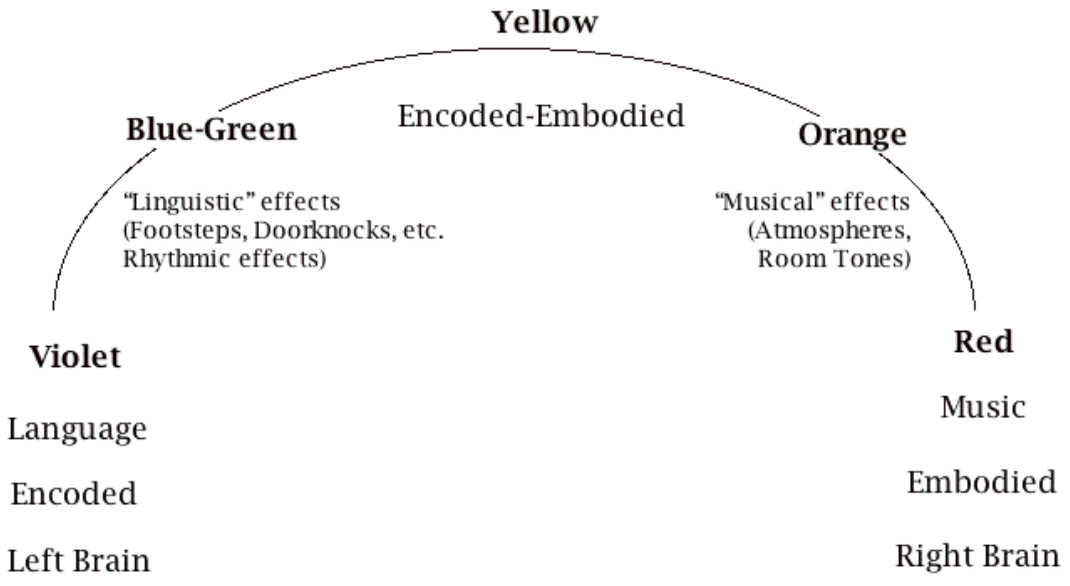
W. K. (2017, September) Tom Clancy's Rainbow Six Siege - footsteps closing in [Video file]
Retrieved from https://www.youtube.com/watch?v=hLoknfeVvBY

Zevik (2010, July 11) "Thief 3: Deadly Shadows", full walkthrough + 100% loot, Mission 2 - End of the Bloodline, Part 1/2 [Video file]
Retrieved from https://youtu.be/REVeD6Hlf_M?t=337

Appendices

Appendices are found in a separate folder along with the thesis.

Appendix A - Aural spectrum



The files used are referenced below, as well as in the reference list:

Murch, W. (2005, April) Walter Murch
Retrieved from https://transom.org/2005/walter-murch/

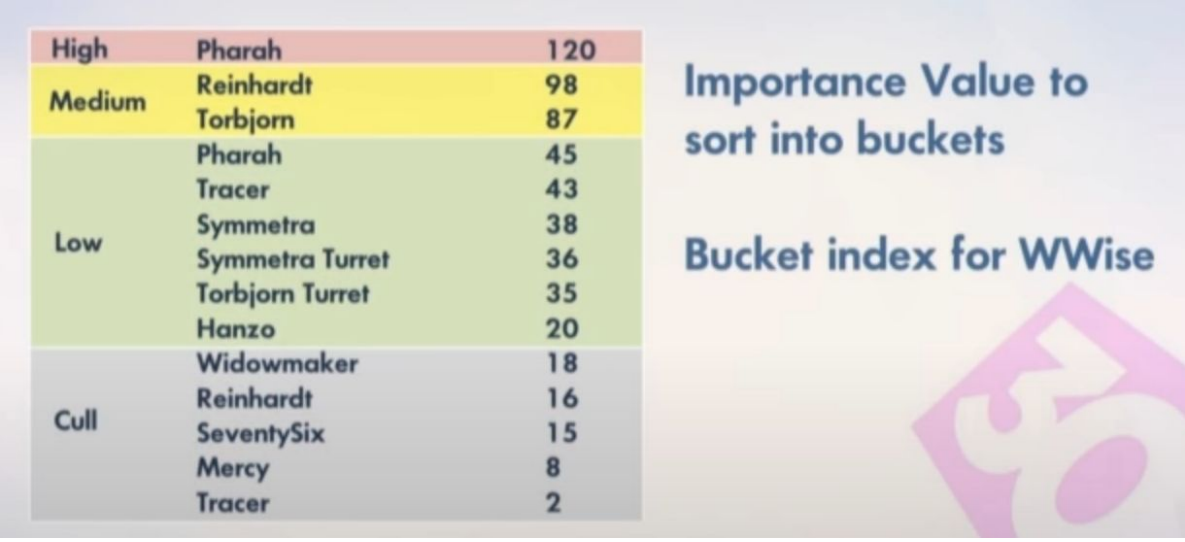
Appendix B - Interview Nicolas Teplitzky

Appendix C - Interview Antoine Laloy

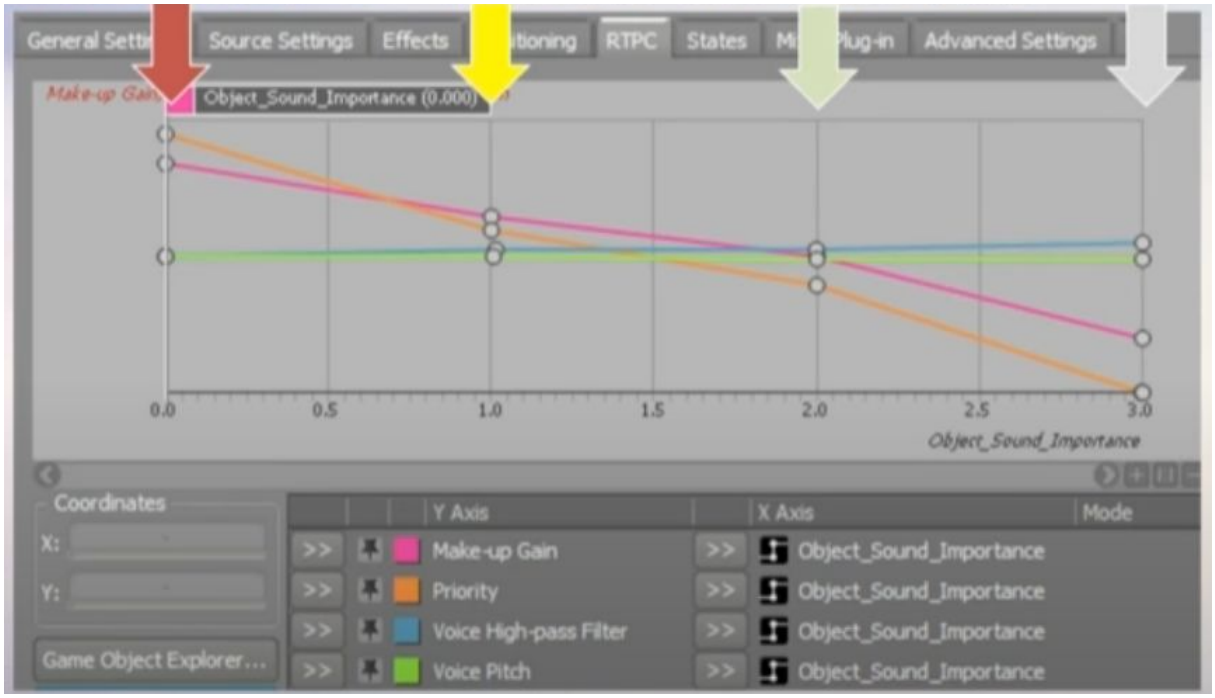
Appendix D - Overwatch screenshots from GDC

Threat level:

Bucket index:



Importance RTPC's:

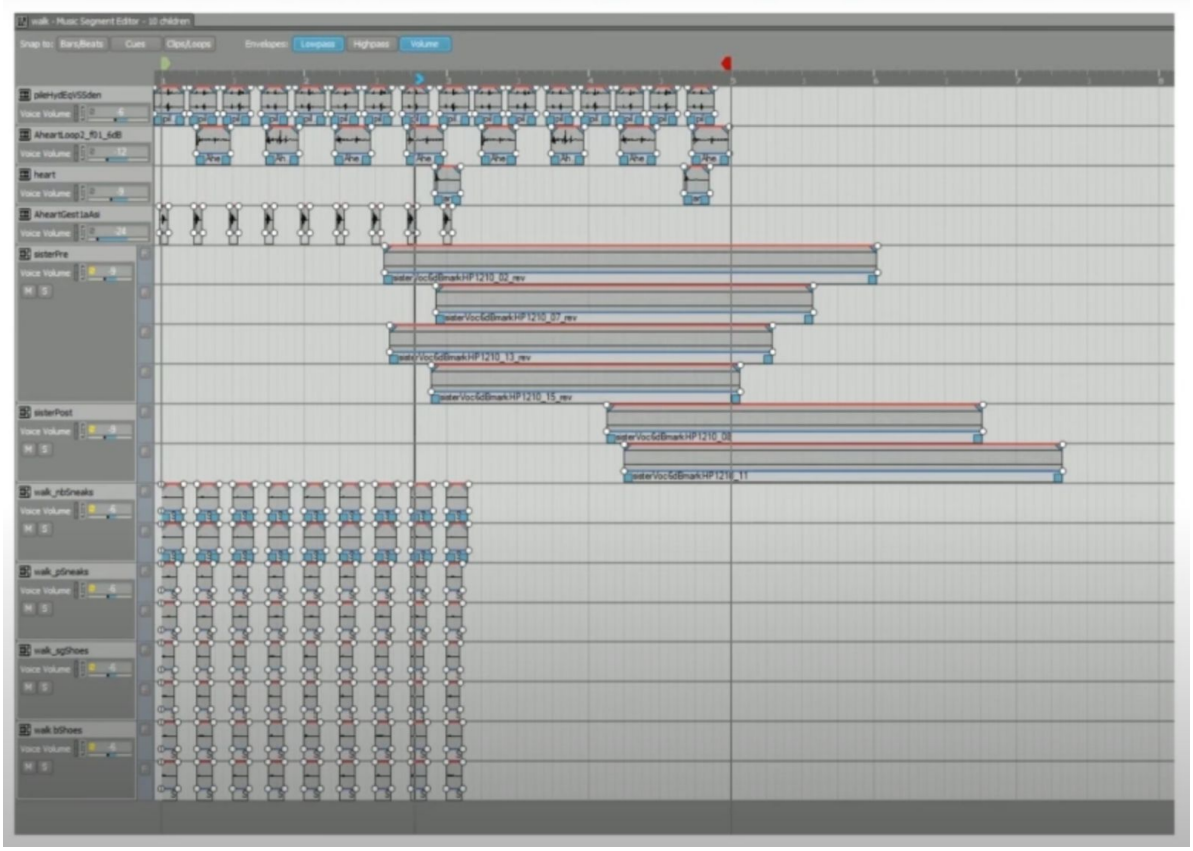


The files used are referenced below, as well as in the reference list:

GDC (2016, July) Overwatch - The Elusive Goal: Play by Sound [Video file]
Retrieved from https://www.youtube.com/watch?v=zF_jcrTCMsA

Appendix E - INSIDE screenshot from GDC

Audio driven gameplay (march scene):



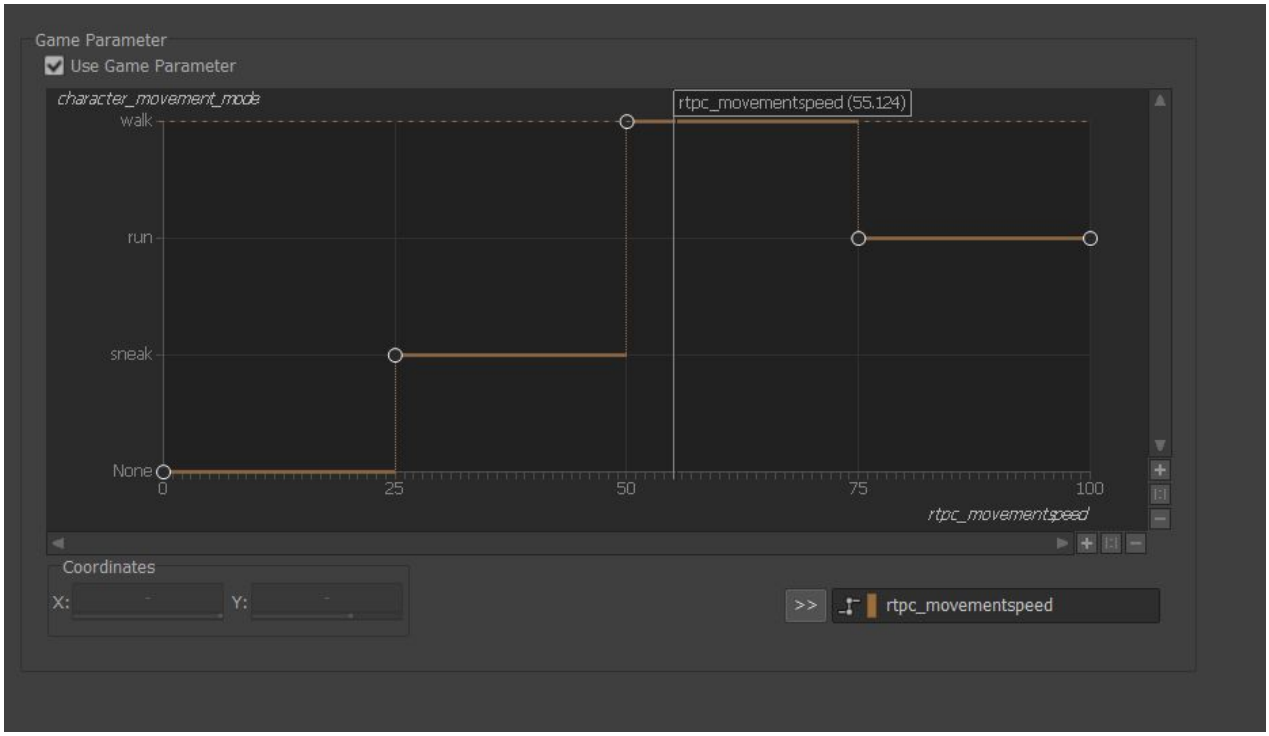
The files used are referenced below, as well as in the reference list:

GDC (2016, July) Inside: A Game That Listens [Video file]
Retrieved from https://www.youtube.com/watch?v=Dnd74MQMQ-E

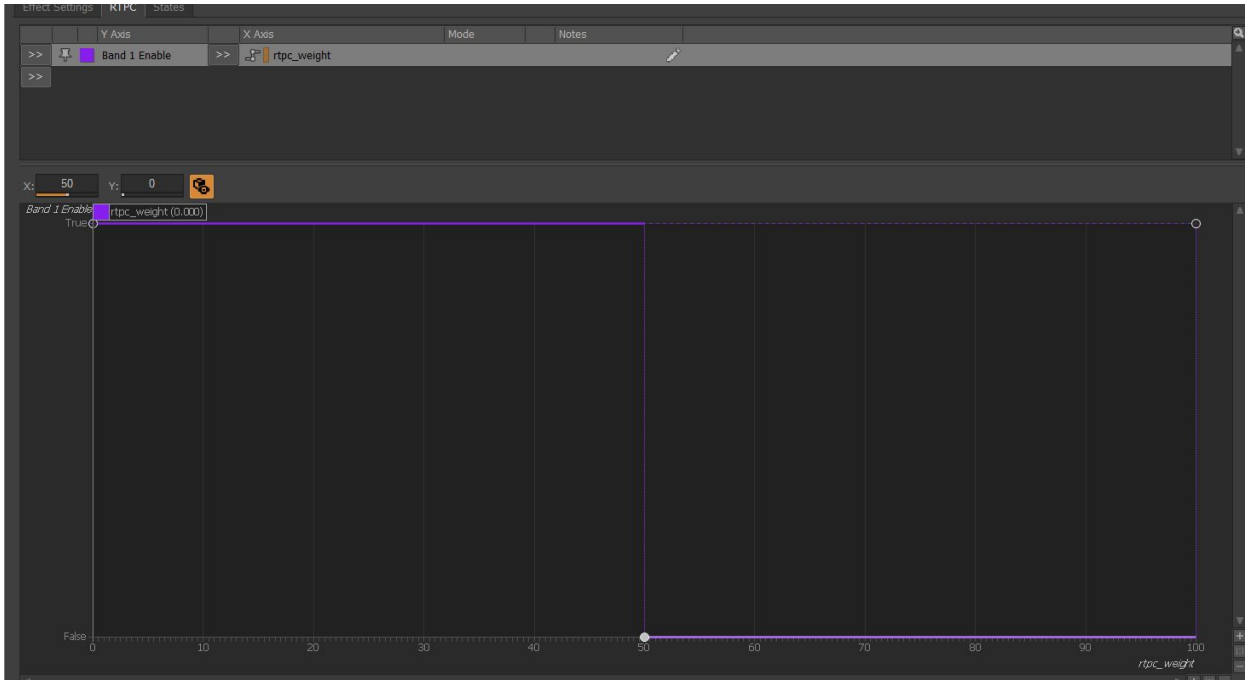
Appendix F - Interview Jacob Schmid

Appendix G - FCT Wwise screenshots

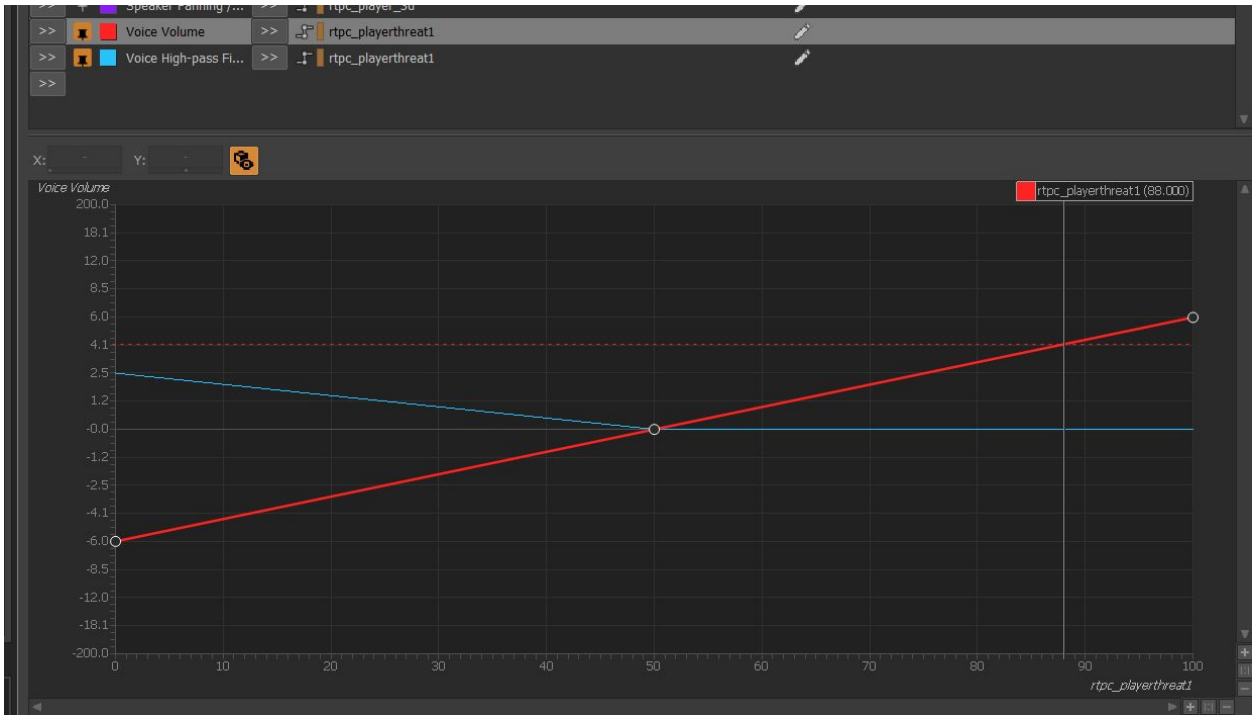
Character movement mode switch + movement RTPC:



Weight RTPC EQ:

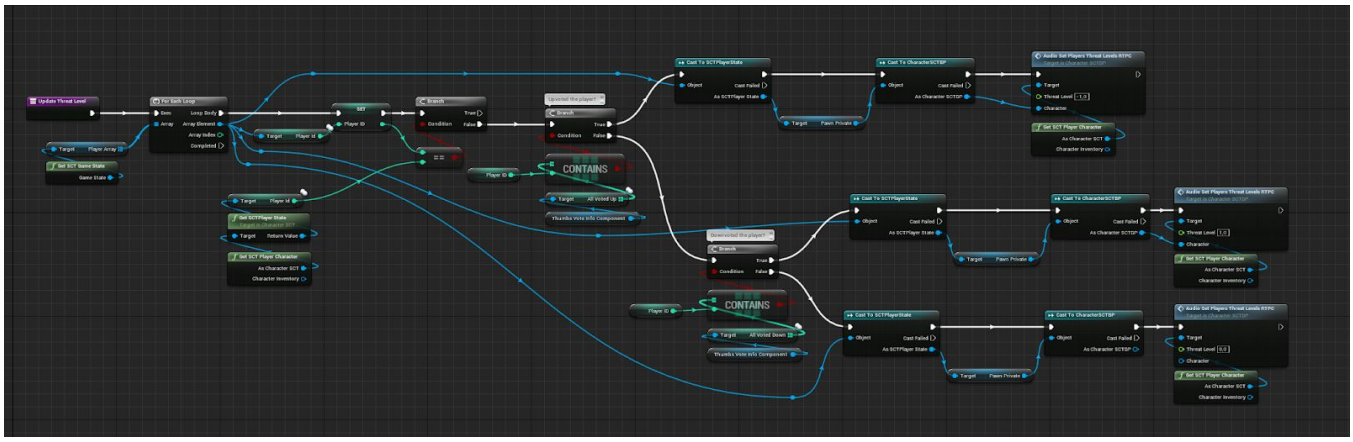


Threat level RTPC's:

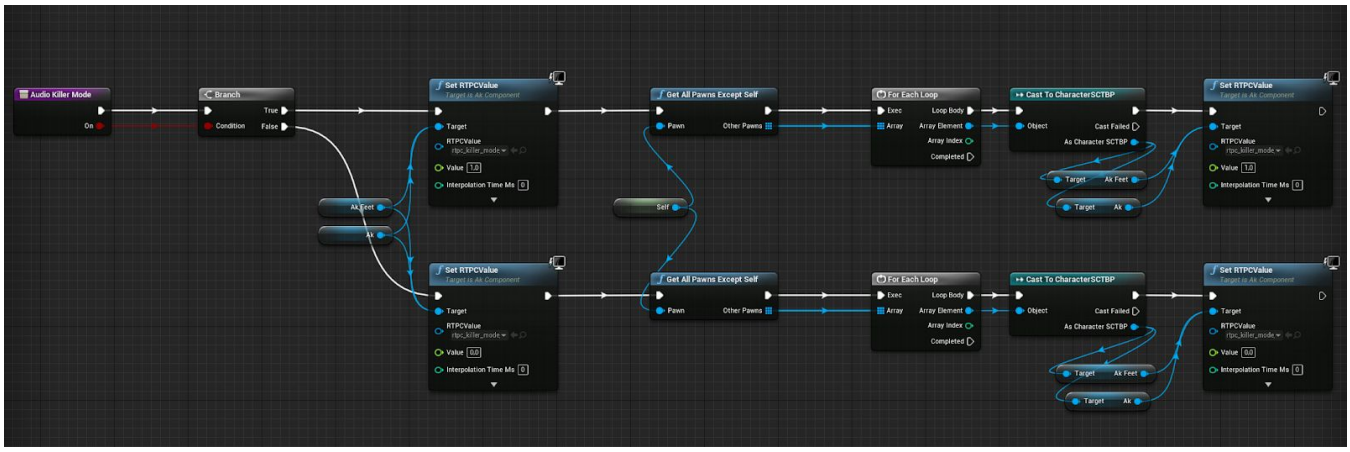


Appendix H - FCT Unreal Editor screenshots

Voting:



Killer state:



Appendix I - FCT Survey Responses

Appendix J - FCT in-game recording examples

Appendix K - FCT Sound files