Bachelor Thesis in Media Technology, Blekinge Institute of Technology, spring 2017

Follow the Raven - A Study of Audio Diegesis within a Game's Narrative

Joakim Rosén | Ted Anderberg

Tutor: Annika Olofsdotter Bergström

Examinator: Pirjo Elovaara

“Reality is that which, when you stop believing in it, doesn't go away”

Philip K. Dick

Abstract

Virtual Reality is one of the next big things in gaming, more and more games delivering an immersive VR-experience are popping up. Words such as *immersion* and *presence* has quickly become buzzwords that's often used to describe a VR-game or experience. This interactive simulation of reality is literally turning people's heads. The crowd pleaser, the ability to look around in 360-degrees, is however casting a shadow on the aural aspect. This study focused on this problem in relation to *audio narrative*. We examined which differences we could identify between a purely diegetic audio narrative and one utilizing a mix between *diegetic* and non-diegetic sound. How to grab the player's attention and guide them to places in order for them to progress in the story. By *spatializing audio* using HRTF, we tested this dilemma through a game comparison with the help of *soundscapes* by R. Murray Schafer and *auditory hierarchy* by David Sonnenschein, as well as inspiration from Actor Network Theory. In our game comparison we found that while the synthesized sound, non-diegetic, ensured that the sound grabs the player's attention, the risk of breaking the player's immersion also increases.

Key Words:

Soundscape, immersion, presence, diegetic sounds, audio narrative, spatial audio, auditory hierarchy

Abstrakt

Virtual Reality är en av de nästa stora sakerna inom spelbranschen, då fler och fler spel som levererar en immersive VR-upplevelse utvecklas. Ord såsom *immersion* och *presence* har snabbt blivit buzzwords som ofta används för att beskriva ett VR-spel eller en upplevelse. Denna interaktiva simulering av verkligheten får bokstavligen folk att vrida på sina huvuden. Möjligheten att titta runt i 360 grader, kastar emellertid en skugga på den audiella aspekten. Denna studie fokuserade på detta problem i relation till att de flesta VR-spel inte berättar en lång historia. Vi undersökte vilka skillnader vi kunde identifiera mellan ett helt diegetisk *ljudberättande* och ett som utnyttjade en blandning mellan *diegetiskt* och icke-diegetiskt ljud. Hur en får tag i spelarens uppmärksamhet och vägleder dem till platser för att de ska kunna utveckla spelets berättelse. Genom att konvertera ljud till *spatialt ljud* genom HRTF, prövade vi detta genom en spel-jämförelse med hjälp av *soundscapes* av R. Murray Schafer och *auditory hierarchy* av David Sonnenschein, samt med hjälp av inspiration från Actor Network Theory. I vår spel-jämförelse fann vi att medan det syntetiserade ljudet, icke-diegetiskt, garanterade att ljudet fick spelarens uppmärksamhet, ökade också risken att bryta spelarens immersion.

Nyckelord:

Soundscape, immersion, presence, diegetiskt ljud, ljudberättande, spatialt ljud, auditory hierarchy

Table of Contents

1. Background

This chapter brings to light the background of our study as well as the general theme, the problem surrounding said theme, the reason why it was conceived and our research question.

In recent years there have been major developments in the field of Virtual Reality. It is becoming more and more available for average consumers and in 2016 the VR-headset was even voted Christmas gift of the year in Sweden (HUI Research, 2016). Frederick P. Brooks Jr stated in 1999 that VR was real (Brooks, 1999). But what is the status of VR now 18 years later? The main appeal of VR is the visual aspect, the ability to enter a new reality and look around in 360-degrees. This focal point on the visual aspect is casting a shadow on the aural aspect. But audio in VR is by no means lagging behind. There have been major developments, specifically within simulating binaural audio, from companies such as Oculus and Two Big Ears.

However, the audio is still not getting the same attention as the visual aspect. Amidst the uncertainty on deciding what we wanted to explore in our bachelor thesis it became clear that audio narrative in virtual reality seemed to be somewhat uncharted territory. Cohen, Villegas & Barfield mentions the lack of literature about spatial audio in the magazine Virtual Reality. Although they don't specifically mention audio narrative, it is a part of the theme.

> The rapid advances in the technology and science of presenting spatial sound in virtual, augmented, and mixed reality environments seem to be underrepresented in recent literature. (Cohen, Villegas & Barfield, 2015, p. 1)

The general theme for our paper is audio for virtual reality, a highly wide-ranging theme. We wanted to explore these, if you will, uncharted waters. Explore the concept of diegetic sound vs non-diegetic sound through a completely diegetic audio narrative in contrast to a traditional audio narrative. What sets them apart?

Audio narrative can focus on the game's storyline or, in this case, information the game conveys for the player in the form of objects or events outside the player's field of vision such as dangers, characters or signs for guidance. The way these sounds are presented to the player affect *immersion*. Immersion being one of the most important things in VR, there needs to be an extra emphasis on these presentations.

1.1 Aim

The aim of this study was to create a deeper understanding of the role sound has in virtual reality by exploring diegetic audio narration. Diegetic refers to diegesis which is a collective word for the world the game takes place in, diegetic sounds in turn means sounds the characters in the game hears. Examine the difference between fully diegetic audio narration, through the game environment, compared to traditional audio narration and the impact this choice between the two has on the player. Shine a light at this underrepresented field of virtual reality and examine concepts such as *immersion* and *presence* from an audiological perspective. Try to understand how the audio ended up second, in terms of importance, to the visual aspect and how it has a significant role in these buzzwords. We did this by creating a virtual reality game together with other students. In a seven-man team, where we were responsible for the audio, we created a game called Norn which portrays the story of two sisters in a fantasy Nordic setting. We created two versions of this game, where one version of the game uses a fully diegetic audio narrative the other draws a distinct line between the two.

1.2 Writing Process

Our writing process has been very iterative and sporadic. We both have a big passion for sound and when working on a big project like this it’s easy to get lost in the sound design and forget that the bachelor thesis needs writing. This was both for good and for bad seeing that we got a lot done for the game during our sound-phases. While having phases where we focused on sound designing certainly allowed for us to make the best out of our sound designing skills, the writing process didn’t get the attention it needed. Yes, even if the times when we only focused on writing gave us polished pieces, it resulted in the thesis lacking coherence and in turn resulted in time spending correcting and rearranging the thesis. On the

positive side, this “bouncing around” approach also enabled us to reflect on our practical work and vice versa during these phases, so it has been an incredibly reflective process. In the future we will strive for a more structured workflow in the writing process. Our process enabled us to reflect on our practical work which in turn brought to life new sound design approaches and ideas. By writing workbooks we could also go back and visit ideas we might have forgotten.

1.3 Problem

One of the bigger problems with creating a fully diegetic audio narrative arises with guidance, in our case more specifically guidance through sound cues. In order to deliver a fully diegetic game experience the clues for direction must also be diegetic, that includes sound cues, the question is: how do we make the player aware of them? That these diegetic sound cues are not simply part of the encompassing auditory world but that they rather highlight something important in this world. Or alternatively: how do we make the player realize that these diegetic sound cues act as the game's way of telling you where to go?

On top of that is the problem with incorporating fully diegetic music or soundtrack. In a traditional sense the soundtrack of a game is played outside of the games world, usually to make the player aware of what kind of mood the game is orchestrating. To make the music diegetic it would be required to emanate from an object, such as a character singing or playing an instrument. How would one solve the problem of music performed by a symphony orchestra?

As mentioned before another problem is the fact that the aural aspect is underrepresented in VR and literature. Games are an audiovisual creation which should include a symbiotic relationship between the two. They both need equal priority.

> In R&D of virtual and mixed-reality systems, the audio modality has been somewhat underemphasized, partly because the ‘‘original field’’ of virtual reality focused on the dominant (at least for some tasks) visual modality.
> (Cohen, Villegas & Barfield, 2015, p. 1)

Mark Nazemi and Diane Gromala also brought this dilemma fourth at the 2012 Audio Mostly, stating that the auditory cues are often secondary in terms of importance:

> For historical and perceptual reasons, the designs of Virtual Environments generally focus on the visual; sound, it may be argued, habitually is of secondary importance. Such neglect of sound, however, can have a negative impact when critical information has to be exchanged amongst the participant and the navigable environment.
> (Nazemi & Gromala, 2012, p. 16)

Mark Grimshaw broached the role of audio in a virtual environment in his study *Sound and Immersion in the First-Person Shooter*. Even though Grimshaw focused on first-person shooters the concept of immersion applies to both. A first-person shooter is, in one respect, essentially a virtual reality without having a VR-headset. Both are a simulation of reality since both have the point of view from a person's eyes. Grimshaw states that “if the diegetic sonic world of the film exists solely for the characters on screen, then the diegetic sonic world of the FPS game extends from the screen to physically encapsulate the player in the acoustic ecology's real resonating space” (Grimshaw, 2008, p. 122). With these studies in mind it may be stated that diegetic audio does have a considerable effect on immersion within the virtual environment. If so, is there a way to apply it in order to maximize the player’s immersion? And how? This becomes even more relevant in VR.

1.4 Research Question

Which differences can we identify in a VR-game using a fully diegetic audio narrative versus one utilizing a mix between diegetic and non-diegetic sound?

1.5 Limitations

We put our study to the test in a VR-game that we collaborated on with other students and there is one matter that we discovered regarding audio in VR-games. The fact that the player can decide how fast or slow something should be swung, how hard or light something should be thrown to the ground. This is not the case in non-VR games since the developers create animations for the players' different actions, such as punching or throwing. It is almost exclusively either 0 or 1 and nothing in between. The sound designers can in turn just create sounds based on these animations. In VR everything is dynamic, the player can solely decide how hard or light he or she wants to throw objects and the sound needs to match this choice.

What comes closest in comparison is velocity, which is used in synthesizers to control how hard or light a note is to be played. Similar to velocity in synthesizers, we can solve this issue by linking the speed of the action in-game to the volume of one audio sample so that the volume of the sample decreases based on the velocity of said action. However, in the real world, when certain objects are thrown at different velocities unique tonal qualities are revealed (or concealed) at the point of impact with another object. That object in turn shapes the resulting sound. On top of that the angle of which object number one collides with object number two helps shape the sound. All in all: every throw is unlike the other. For this to be represented in VR there needs to be multiple audio samples recorded at different velocity so that if the player throws a stone on the ground with force a different sample plays as opposed from one that plays when the player throws the stone with less force.

Naturally this added a significant amount of work for us responsible for the sound design. We were aware of that as soon as something is apparently out of place in VR there is a risk of breaking immersion. We made the choice to not focus on this. It was a question of priority, where to allocate our time and resources. We focused on the topic of our study and worked on the comprehensive sound design of the game.

Summary
VR development is growing more and more and becoming more available for the consumers. The main attraction, the visuals, is casting a shadow on the aural aspect which is why we conducted this study and aimed to explore a purely diegetic audio narrative within the VR medium. Examine how to grab the player's attention and review terms such as immersion and presence.

2. Previous and Current Research

In this chapter we present the previous and current research relevant to our study and showcase pertinent topics that allows us to conduct this study. We delve into what is crucial for VR, such as spatial audio, immersion and presence, and describe diegesis and soundscapes.

2.1 Virtual Reality

While the recent surge of virtual reality games can make it seem like this is new technology, this is not the case. Ivan Sutherland, computer scientist and regarded as the father of computer graphics, introduced his vision of VR at the IFIP Congress in 1965. Sutherland's conference paper, *The Ultimate Display*, brings up ideas like eye-tracking, haptics and head-mounted displays. As Sutherland puts it: "It is a looking glass into a mathematical wonderland." (Sutherland, 1965, p. 506)

Suffice it to say that Sutherland's vision was ahead of its time. This was even the case when Frederick P. Brooks Jr, software engineer, computer scientist and recipient of the Turing award, surveyed the field of VR in 1994 and stated that VR almost works. It wasn't until the dawn of the 21st century that this status changed. Brooks said: "I think our technology has crossed over the pass - VR that used to almost work now barely works. VR is now really real." (Brooks, 1999, p. 16)

Brooks (1999) brings up 4 technologies that are crucial for VR:

- the visual (and aural and haptic) displays that immerse the user in the virtual world and that block out contradictory sensory impressions from the real world;
- the graphics rendering system that generates, at 20 to 30 frames per second*, the ever-changing images;
- the tracking system that continually reports the position and orientation of the user's head and limbs;
- and the database construction and maintenance system for building and maintaining detailed and realistic models of the virtual world.

Brooks also presents 4 important, but not nearly as crucial, auxiliary technologies:

- synthesized sound, displayed to the ears, including directional sound and simulated sound fields;
- display of synthesized forces and other haptic sensations to the kinesthetic senses;
- devices, such as tracked gloves with pushbuttons, by which the user specifies interactions with virtual objects; and
- interaction techniques that substitute for the real interactions possible with the physical world.

Since we focused on the aural side of VR in this study we concentrated on the technologies concerning audio, such as the aural displays and synthesized sound.

* Today at least 60 frames per seconds is sought (Oculus, 2015)

2.2 Diegesis and Soundscapes

> Every film still creates its own filmic world, its diegesis. Everything which happens inside this world is called diegetic and what happens (in the movie) outside this world is called non-diegetic. (Dykhoff, 2012, p. 169)

Klas Dykhoff explores diegetic and non-diegetic sounds in his article *Non-diegetic sound effects*. Typical examples of non-diegetic sounds are voice-narration and original music score since the characters in the film are unaware of them because they don't exist in the same world (Dykhoff, 2012). Non-diegetic sounds are common in films today and are even more common in video games because of the interaction between the player and the user interface or game menu. A traditional audio narrative consists of both these types of sounds.

To understand diegetic and non-diegetic clearer we had to explore something that Michel Chion calls the audio-visual contract. Dykhoff (2012) explains this contract as an understanding between the audience and director about what rules applies to sound and image in the film. Dykhoff states that "Chion's argument is that the audience accepts that what's happening on screen is 'reality' for the duration of the film and that everything outside the film, is irrelevant" (Dykhoff, 2012, p. 170). Synchresis is a part of this contract. Chion writes:

> Synchresis (a word I have forged by combining synchronism and synthesis) is the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time. (Chion, 1994, p. 63)

The viewers themselves make the connection between the visual event and the sound that is played with it and the sounds are no longer separate events, but something tied to the image (Dykhoff, 2012). In regards to our study, this becomes relevant in terms of audio narrative, guiding the player forward in the story.

With synchresis in mind one can say that sound designers/editors are lying to the audience/player. Actors in movies don't actually hit each other but the sound of punches dramatizes the punch. Through synchresis the audience perceive these sounds as coming from the events on screen (Dykhoff, 2012). Sounds can also be non-diegetic when they are unrealistic compared to the location, for example if gunshots are present in a film or game taking place in ancient Greece. (Dykhoff, 2012)

Dykhoff also utilizes Johnny Wingstedt's narrative functions for music and argues that diegetic and non-diegetic is relevant to more than just music. Wingstedt (2010) divides them into:

- emotive function (to convey emotions);
- informative function (including information about era, cultural space/environment, and social status);
- descriptive function (describing movement and size);
- guiding function (leading the eye or the thought);
- temporal function (creating continuity, define structure and form in a scene, or in a film);
- rhetorical function (contrast the story with well-known pieces of music)

The same categorization can also be used for sounds that aren't music, and even for dialogue (Dykhoff, 2012). These categories will serve as great guidelines for our study, especially the guiding function for leading the thought. Dykhoff finishes his article:

> By whatever expression one chooses to call it, I am convinced that the question of diegetic or non-diegetic is as relevant as ever. It has implications for all auditory aspects of film narrative and it can be used in conjunction with several other analytical tools for film sound. (Dykhoff, 2012, p. 178)

Raymond Murray Schafer coined the term *soundscape* in 1977 in order to describe the sonic characteristics of an environment, the collections of noises and sounds that surrounds us on a

day to day basis. Schafer (1977) argues that in order to analyze the soundscape, one must first determine the significant features of the soundscape i.e. what sounds are the most prominent, whether by individuality, numerousness or domination. Additionally, some system(s) of generic classification must be constructed. He presents *keynote sounds, signals* and *soundmarks. Keynote sounds* consist of sounds created by the geography and climate of the soundscape e.g. water, wind, birds, *signals* are foreground sounds and are listened to consciously e.g. bells, whistles, horns and *soundmarks* (derived from landmarks) which refers to community sounds whose uniqueness results in the sounds being specially noticed by the people in the community.

2.3 Spatial Audio

> The aim of spatial sound rendering is to create an impression of a sound environment surrounding a listener in 3D space, thus simulating auditory reality.
> (Väljamäe, 2005, p. 4)

Spatial sound technology refers to modeling the propagation of sound within an environment while accounting for the human listener (Kapralos, Kanev, Collins, Hogue & Jenkin, 2015). It does this through the use of Head Related Transfer Functions, abbreviated as HRTF. A non-spatialized sound is convolved with transfer functions corresponding to the desired spatial position of the source (Väljamäe, 2005). It is similar to how a convolution reverb works where one records the pop of a balloon in a desired space and loads this sample recording into a convolution reverb-plugin. This plugin then recreates the acoustic space through the sample, essentially simulating the reverb from the recording.

In a study conducted in 2009, focusing on auditory localization in a virtual reality through the audio-visual relationship, Nguyen, Suied, Viaud-Delmon & Warusfel (2009) concluded that spatially different visual and audio cues influence the auditory localization and that it systematically shifted toward the visual stimulus. Nguyen et al. (2009, Figure 1) deducted that vision strengthens the ability to locate sound sources when both auditory and visual stimuli were spatially aligned as opposed to an auditory alone condition.

In addition, they found that a slight underestimation of the stimulus angle occurred in the visual only condition, as well in the spatially aligned condition, but not in the auditory alone condition. They hypothesized that this underestimation could strictly be a visual phenomenon or that it could have been hidden in the larger deviations observed in the auditory only condition.

Besides technologies there is another matter which is of the utmost importance in VR that, if broken, will pull the player out of the virtual world and destroy the experience as a whole.

2.4 Immersion

> The goal of immersive virtual environments (VEs) was to let the user experience a computer-generated world as if it were real-producing a sense of presence, or "being there," in the user's mind. (Bowman & McMahan, 2007, p. 1)

One word that pops up more often than others when talking about Virtual Reality is *Immersion.* Grimshaw (2008) states that immersion derives from the player 'becoming' the game character and that the goal may be defined as the player's perception that he is within the game environment. In addition to immersion Doug A. Bowman and Ryan P. McMahan (2007) presents *presence* in their article *Virtual Reality: How Much Immersion Is Enough?* They claim that presence refers to a user's subjective psychological response to a VR system, how the player feels himself inside the virtual world giving the experience of "being there".

Likewise, in their book *Understanding Virtual Reality*, William R. Sherman and Alan B. Craig (2003) argues that the use of sound in VR vastly improves the player's ability to become immersed in the virtual world. Continuing on perceived perception they claim that one aspect of our perception is the ability to generalize, while still emphasizing that there still are upsides and downsides to this aspect. However, one particularly interesting example they give of how our ability to generalize can be beneficial is regarding *symbols*. They argue that it is desirable to use the most generic instance of the object available, since it will likely have

the least cultural bias and chance for misinterpretation. In our case this applies to sound and is something we will discuss later in this study.

Laura Ermi and Frans Mäyrä (2005, Figure 2) presents three forms of immersion. *Sensory immersion* relating to the audiovisual execution of games. Powerful sounds and visuals easily overpower the sensory information coming from the real world making the player entirely focus on the game world. *Challenge-based immersion*, requiring motor and mental skills and relates to the feeling of immersion when the player achieves a challenge regarding strategic thinking or problem solving. Lastly, *imaginative immersion*. The dimension of game experience in which one becomes absorbed with the stories and the world, or begins to feel for or identify with a game character.

Summary

Brooks presents 8 different technologies crucial for VR and since we focused on the aural side of VR we concentrated on the technologies concerning audio. Dykhoff explains diegesis which is the world in the film takes place in. Everything that happens inside this world is called diegetic and what happens (in the movie) outside this world is called non-diegetic. Brooks expresses how the question of diegetic and non-diegetic sounds is as relevant as ever. Sherman & Craig argues that the use of sound in VR improves the player's ability to become immersed in the virtual world.

Spatial sound technology refers to the simulation of our hearing and Nguyen et al. states that depending on whether spatial sound occurs alone or together with visual feedback impacts our ability of auditory localization. Bowman & McMahan established the differences between immersion, the VR-system's ability to trick the player into feeling that they're somewhere else and presence, being the experience of "being there". Ermi & Märyä continues and presents three forms of immersion: *Sensory immersion, Challenge-based immersion* and *imaginative immersion*.

This research is the foundation for how we went about to fulfill our study, which methods we chose, and took inspiration from, for the practical and technical aspect for this thesis.

3. Methods

In this chapter we present our methods to provide a basis for our study and to discover solutions to the problems we perceive. We have chosen productions methods, such as *soundscapes*, to help us with the practical side of the study, to create and implement sounds in such a way it fits in VR. We have also chosen analytical methods which will aid us with our study's research question such as listening modes and auditory hierarchy. We've also taken inspiration from Actor Network Theory, which also helped us with the creation of sounds.

A part of our bachelor's thesis consists of a representation. This representation is in the form of a VR-game in two versions with different audio narratives for the comparison and examination of our issue. We worked together with five other students who are studying game development and we were responsible for the audio. The game is set in a Nordic fantasy setting centered around two sisters, the player being one of the sisters. Given the time constraint the game is fairly short but it still contains a story to be told. The player will carry out small varying tasks or events around the map and it is here we implemented our diegetic audio narration, how to guide the player to these different events in the game only using diegetic sounds.

3.1 Analytical Methods

3.1.1 Actor Network Theory

Seeing as we were looking at the connection between the player and an audio cue, or a human and non-human entity, we felt that Actor Network Theory was a good inspiration for our study. John Law, Bruno Latour and Michel Callon developed ANT in the 1980s and can be described as method to explore the relational character of materials (Law, 2012), or as Latour states (Cited by Mark Cypher and Ingrid Richardson, 2006)

> Bruno Latour conceptualizes the human-technology relation as a heterogeneous network of human and non-humans (technologies/machines/materials) that work together to make things possible that neither could achieve without the other. (Cypher & Richardson, 2006, p. 4)

Granting every actor equal amounts of agency and how different entities, or actors, interact and develop into new actors. In this case, how the player (Actor 1) and the audio cue (Actor 2) forms the third actor, the guiding element. Cypher & Richardson (2006) states that there is a level of influence which alters each of their associated goals and objectives. An example Cypher and Richardson brings up is: "The human's goals are bound up with all manner of social and psychological narratives that aid in the personal belief that a gun is required for protection*"* (2006, p. 4). In our case the player is constrained with social and psychological aspects that aid, or prevent for that matter, in the personal belief that the audio cue is a hint for guidance.

Taking inspiration from this method helped bridge the gap between the player and the audio cue and was a useful tool for us to look at how they interact, as well as the creation of sounds. It generated thoughts surrounding the workings of our diegetic audio narrative which we go into further in the discussion-chapter.

3.1.2 Listening Modes & Auditory Hierarchy

Sonnenschein states that "hearing is passive and listening is active. Our ability to listen is multi focused, which means we can glean information through several different psychological and perceptual perspectives" (Sonnenschein, 2001, p. 77). Sonnenschein presents Chion's three listening modes but also adds a fourth. 1) *Reduced* listening pertains to the observation of the sound itself and not the source or its meaning; 2) *Causal* listening consists of listening to a sound to be able to gather information about its cause; 3) *Semantic* listening pertains to the spoken language and other code systems that symbolize ideas, action, and things; 4) *Referential* listening consists of being aware of or affected by the context of the sound,

linking not only to the source but principally to the emotional and dramatic meaning. (Sonnenschein, 2001)

These listening modes gave us an idea of how the player could hear and react to our audio narration cues, as well as to be used as a tool to analyze our sound design.

Sonnenschein (2001) presents auditory hierarchy and introduces three levels of how the way sound can command our attention:

1. The immediate presence of something we are consciously listening to
2. The support sounds for the event or environment that are heard but do not command direct attention
3. Background sounds that we don't notice but that constitute the created reality and can influence our subconscious

Sonnenschein also states that "moments of exaggeration of the sound object can be effective either explicitly to generate comedy or irony, or implicitly to focus our attention on an otherwise subtle event" (Sonnenschein, 2001, p. 194). The sacrifice of realism and logic for the sake of enjoyment, or in this case guidance.

3.2 Production Methods

3.2.1 Our Sound Designing Approach

In relevance to the study we conducted, we used Schafer's term *soundscape* to help us categorize the themes of our fictional world through the difference between *keynote sounds, signals* and *soundmarks*. With the core established and since the game we created is based around Norse mythology, we added *archetypal* sounds. *Archetypal sounds* refer to ancient sounds, who often possess symbolic qualities that we have inherited from remote antiquity or prehistory (Schafer, 1977).
We started by identifying similarities between the sounds we use and Schafer's terms. Our *keynote sounds* are the sounds that create the ambience of our game e.g. the birds, wind and

waves rolling onto the beach. *Signals* are sounds that call attention to our player such as hammer impacts when smithing, firing your bow or skipping stones. We classify our use of seagulls as our *soundmarks*. Our *archetypal* sounds are essentially the raven we use as the guiding actor. However, since the raven is being used as a guidance actor, it is also a *signal* sound.

With this in mind, to help us determine what sound had higher position in the diegetic audio hierarchy we decided to deconstruct what sounds we planned to use in the game and the importance of said sound. Importance being the core sound design philosophy behind the game and subsequently the study. We arrived at this conclusion:

1. *Guiding sounds* (Signals)
2. *Character sounds* (Signals)
3. *Feedback sounds* (Signals)
4. *Ambiences* (Keynote Sounds)

So, the sounds with the highest position in the hierarchy is dialogue and the bird calls we use as directional guidance. In order to achieve maximum clarity we have to make sure the player can hear the bird cues wherever they are in the level. While these sounds still have unique tonal characteristics, the added reach emphasizes the sounds as highly important. The following hierarchical position consists of footsteps, considering footsteps innately describe the notion that one is moving and conveys what surface one is travelling on. Following footsteps comes feedback sound such as when the player picks up an item or similar actions. While this at first encounter might not sound like it has importance in terms of guiding the player, the fact that these (for example) items still have a set location in the world informs the player of their directional properties, and calls attention from the player as to where the sound source came from. Lastly, we have ambient sounds. These sounds encompass the world the player moves around in, sets the tone and gives the player a world to relate to. The idea is that player relates to his/her experiences in the real world to get a sense of location on where the game might play out.

Lastly, when designing the guide for the naval area we decided not to use our raven but rather seagulls. This pertains to the general perception that seagulls flock near harbors and coastal areas. They are, in a way, a *symbol*. As Sherman & Craig (2003) claims, we use these in order to avoid confusion and hopefully have the least cultural bias.

3.2.2 Creating the Diegetic World

One effective method for designing sound in general is to break down the object needing sound into lesser parts. Deconstructing the object reveals the working parts. Each individual part generates a sound and together with the others aurally completes the object. An example is the sound for the teleport-function in the game. With our study in mind we wanted a diegetic sound for this. Instead of seeing it as a magic phenomenon we're seeing it as we're simply skipping the transportation from A to B. How is the player going to know this? With the use of sound.

We deconstructed the sound of a person taking one or two steps. In this case the sound effect consists of footsteps, clothes, grass, small metallic noise imitating keys or harness, and arrows moving about in the quiver. All of these sound effects depend on where in the world you are moving to, we made unique instances of sound effects playing footsteps on different surfaces that directly corresponds to what material you are moving to. Hearing this gives the player information that they are traveling and where in the world they are traveling to.

Another example is our arrow impacts in our game. We have impacts for different materials, rock, wood, grass etc. We wanted a sound for when only hitting the leaves of a tree. We recorded this by hitting leaves with a stick and were happy with the results. When processing the recordings we felt that only hearing leaves rustling wasn't enough. So, what happens when an arrow hits the leaves? The arrow flies through the leaves or maybe the arrow hits a tree branch not visible for us. We incorporated sounds of a stick hitting the ground. We now have a collection of samples hitting the leaves, some with branch-hits and some without. These triggered randomly creates more variety and adds a more realistic feel.

When creating sounds for the archery we had ANT in mind. This eased our creative process and made us look at the bow in different perspectives. With ANT we set our focal point on the

connection that brings the player and the audio cue together, which in turn creates the guiding element that will hopefully result in the player going to where we want him or her to go. In addition to this we also had challenge-based immersion in mind. It will require motor skills to mimic the use of an actual bow. If there is any sound that stands out and is perceived unrealistic it will grab the player's attention and break this challenge-based immersion. All the sounds for these actions needs to match and be as close to real-life as possible. This is the reason we chose to record these sounds at the local archery club, to get to know the process and know how a bow and arrow sounds and feels.

We wanted to create a rich Nordic environment that felt alive and therefore we wanted a lot of different birds singing. From the start of our study we had set the goal to record as much of the sound effects we could ourselves. However, recording birds in the spring for a game that takes place during the summer is not a very good approach. Therefore, we got our birds from audio libraries, although we made sure the birds we used can be found in Scandinavia in an attempt to emphasize the Norse mythology narrative.

3.2.3 Attenuation

One of the biggest criteria for achieving immersion in VR is by using believable attenuation curves, however this can prove to be difficult, especially considering the line between realism and art. VR as a medium requires everything to behave like they would in the real world or else you risk breaking the immersion, when you are put in a VR-world you expect it to act naturally and frankly the only reference we have is the world we live in. As such, if these two worlds somehow don’t match the chance of you feeling like a spectator rather than a participator increases.

Under the conference Oculus Connect Tom Smurdon (2015, Figure 3) argued about the importance of correct attenuation curves. He found that the more he used inverse square law curve the better things sounded. However, when we want to get the player’s attention we have to change these curves in order to make the sound audible in game. If we were to use the inverse square law curve on our guidance sounds the curve, while realistic, resulted in the further away players were from the sound source the lesser chance we had of conveying our message with the sound. In some cases we found that the most suitable way was to increase

the max attenuation rather than minimum (Oculus, 2017, Figure 4), unlike Smurdon who believed that the best way to deal with this was to use the minimum value. In the end, it's all a case of reality versus clarity. This is an example of how exaggerating the sound, or in this case the curve, can be an effective tool to command attention as we mentioned earlier.

Summary

We've taken inspiration from ANT since we were exploring the connection between two entities, the player and the audio cue. ANT also inspired us as to how we recorded, created and edited the sounds for the game. Further we utilized Sonnenschein's listening modes and auditory hierarchy and together with a modified version of Schafer's *soundscapes* created a list for which type of sound had highest priority. We also presented production methods which we've learned from our profession, from developers such as Oculus and from the previous research.

4. Discussion and Conclusion

Here we present the results of applying the analytical and production methods we found relevant to our study, such as listening modes, auditory hierarchy and ANT, and subsequently discuss and analyze these results. Lastly we present our conclusion by showing the differences we identified between the two audio narrative forms, diegetic and non-diegetic.

4.1 Discussion

During this study we've always had the quote by Philip K. Dick in our minds. It has served as a driving force and has helped with both writing this thesis and the creation of the aural world within the game. It challenges the core concept of VR and begs the question, will VR ever be so immersive we won't stop believing in it?

4.1.1 Avians vs Synthesis

Since one version of the game utilizes a synthesized sound to act as the player's guide while the other uses a sound that occurs in the real world, this causes issues regarding maintaining the immersion of the player. In our testing we found that while the synthesized sound ensured that the sound caught our attention, we immediately knew that this was not a natural sound, that it wasn't part of the game diegesis, and in turn the sound subsequently broke our immersion. We discussed this further by deconstructing what these sounds consist of. First of all: the raven and seagull guide.

Ravens are a type of bird that has for a long time been regarded as ominous (in Swedish folklore the conception of "nattramn" exists, where ghosts of suicide victims manifest themselves as ravens) yet has for a long time coexisted with humans. Because of the croaking call the raven emits and the fact that we established a soft soundscape with overwhelmingly melodic bird calls, this helps the raven stand out. On top of that most of the frequency information from the raven appears in the mid to high-mid range of the audible frequency range, which also happens to be the area where our ears are most sensitive to according to S. Johansson (personal communication, March 18 2015). As for the seagulls, they also contain roughly the same frequency information as the raven on top of being portrayed as a naval *symbol*. Therefore, we felt that they could substitute for the raven in this area of the game.

Moving on to our synthesized guide. We discovered that one benefit to this method was that since synthesized sound never occurs in the real world it was sure to cut through the established sounds and grab our attention. Although this attention came at the cost of our immersion. Using a synthesized sound as a guide could result in it not being realistic in relation to the setting of the scene. There's also no spatial information as to where this sound is coming from and together with no directional information from the visuals resulted in the player becoming disoriented.

It can be argued that depending on the genre and setting the game takes place in there is some leeway in diegetic and non-diegetic sounds. Science fiction is convenient for example: if we wanted a sound that beeps faster the closer a player gets to the target area, we could just let

the player know, through dialogue or otherwise, that it's a gadget the character carries. In that manner one can create whatever gadget needed to make elements diegetic.

Also, concerning the four listening modes we mentioned earlier. They gave us an idea of how the player could react to the audio cue, however one problem with these modes is that we cannot assume the player listens in any of these modes. These modes were invented by individuals who have a lot of knowledge and experience about sound. Would someone who doesn't have this knowledge and experience listen and think of sound in the same way?

4.1.2 Actors

To get the player to go where one wants him or her to go and to better understand the connection between the player, the audio cue and the guiding element one should perhaps look at each actor individually. Aiming to make sure one know each actor's meaning and to review them.

The player is a human and non-human entity considering the player is a character in the game, which is controlled by a person in real life. This is something to take into consideration in regards to looking at the connection between actors. Does this differ when the game is trying to simulate reality? In this case we saw the player as the third actor since they, in one sense, become one in VR. The character and the person playing becomes the player, but do we view this third actor as a human or non-human entity? The audio cue is by no means human so by process of elimination we can view the player as human.

The audio cue is the non-human entity and is in regards to guidance represented by different bird calls. For example, when the player is prompted to go catch some fish for dinner we wanted to play sounds of seagulls calling to indicate a presence of water nearby. Like mentioned before, we utilized the general perception that seagulls are often found near water. We asked ourselves, does this coupled with the spatial information, where the sound is coming from, give enough information for the player to have an idea of where to go? A solution could be to make use of drama and implement the sound as of two seagulls fighting instead, which could be intriguing for the player.

The player and the audio cue in turn forms this third actor, the seed that springs the thought ‘Ah I need to go there’. The player and the audio cue work together to make the guidance possible. Could they achieve it without the other? The player could probably figure out where to go just by exploring the game and eventually end up at the right place, although in that case we would not have achieved our sound design goal since the game still has a linear story that we want the player to follow.

An interesting side note. During the creation of this study a several VR-games was released and since we had access to both HTC Vive and Oculus Rift it naturally resulted in us testing a lot of them. Interestingly enough one of these games, Wilson’s Heart (Twisted Pixel, 2017) utilized the same idea of diegetic teleport-sounds as we did: diegetic footsteps to portray the transportation to another area.

4.2 Conclusion

In reference to Dykhoff (2012) stating that the question of diegetic or non-diegetic has implications for all auditory aspects of film narrative, we would add game narrative into that question. They are two different mediums but both part of dramaturgy and has the purpose of telling a story. It is also relevant to the guiding function (leading the eye or the thought), part of Wingstedt’s (2010) narrative functions. Our conclusion is that diegesis definitively has a factor in the audio narrative regarding VR-games.

One of the differences we identified between a fully diegetic audio narrative and one utilizing a mix between diegetic and non-diegetic sounds is that VR brings the person playing and the character in the game together, they are in sync. To look in a certain direction the person must turn his or her head, instead of via mouse and keyboard. Referring to how Bowman & McMahan (2007) describes immersion and presence, there is in one sense no need to direct sounds solely at the person playing. The only exception being the music soundtrack.

Given that VR simulates reality, any type of reality, it makes sense that the audio narration, any sound for that matter, should be diegetic. This coincides with Sherman & Craig (2003)

who argued that the use of sound in VR vastly improves the player's ability to become immersed in the virtual world and to Grimshaw's (2008) statement that the diegetic sonic world helps to physically encapsulate the player. Based on previous research, if there are sounds not belonging in the reality it can raise the risk of breaking immersion.

Since it's established that immersion and presence is essential for the VR experience, it should not be broken. To maintain immersion one must make a realistic simulation. To make a realistic simulation one must not break the audio-visual contract. We could identify a direct comparison to our own work with our representation. In regards to the game comparison we performed, we identified another difference concerning synthesized sound.

We found that while the synthesized sound ensured that the sound grabs the player's attention, the risk of breaking the player's immersion also increases. The problem with using a synthesized sound is that it contains no directional information. If we would apply it to an object in the game it would become diegetic. A following problem to that is that there isn't any object that would emit a synthesized sound since it takes place in a world where no digital technology exists. This boils down to if the player notices this or not, the player could for example believe that the synthesized sound is a form of magic emanating from an object. As we mentioned earlier, depending on the genre one can make non-diegetic become diegetic.

4.2.1 Concerning Soundscapes

As Schafer (1977) described with *soundscapes*: to analyze a soundscape one must first determine the significant features of the soundscape e.g. "sound cues". Since a soundscape and all of its components generally are naturally incorporated in our auditory field, whether that is in a virtual reality or the reality we live in, we deducted that there are three ways to make sound cues stand out:

1. *Intensity* (how hard the sound cuts through the already established sounds)
2. *Rarity* (how often the sound cue occurs)
3. *Frequency* (how often the sound repeats itself in succession)

Naturally a louder sound will grab more attention than a weaker sound. This is similar to a phenomenon that can occur in the real world known as auditory masking. When the frequency information from one sound interferes with that of another our brain has trouble processing the information coming from both sources (Oxenham, 2013).

This coincides with what Sonnenschein (2001) described in that exaggerated sounds can be effective to focus our attention on an otherwise subtle event. One effective example of this is vehicular horns. Since their inherent loudness and unnatural tone acts as a warning to everyone around it, the sound masks other sounds and therefore it rarely goes unnoticed when heard. Likewise, a louder sound in the game will likely cut through the sounds that the player already established as the soundscape of the game world. However, since we use bird calls as sound cues and these occur naturally in the world we arrive at our second point: *rarity*. It is important to note that the rarity of a sound plays a huge part in how we perceive it. The less often we hear a sound, the less we get accustomed with hearing said sound and in turn the chances are higher that we get alerted by it. Lastly, *frequency*. Repeating sounds have higher chances of grabbing our attention, for example: ringtones, when finished starts over again in an attempt to grab our attention. What originated as a simple sound cue ends up becoming a nuisance that we want removed from our vicinity. It is important to note that this list is not necessarily listed in hierarchical order, that any order could potentially work. However, based on our research and findings, we conclude that the best way to grab the player's attention through diegetic sound is by using these three options simultaneously.

4.3 Final words

Through our research and testing we have developed new knowledge and understanding about the fields we touched upon. Developing for VR is very exciting because it's an evolving technology and encourages creative experimentation. However, this also results in the field lacking documentation of technologies. It is also very challenging since we need to think more than usual about reality and all its infinite components. We always had real-life to compare our work to by simply removing the headset, which made Philip K. Dicks quote

appear in our minds. There were also discussions between us since we both have a different view of the world, our own realities. This resulted in a lot of work on intricate details.

Based on all hours we spent reading, thinking, writing, recording, editing, implementing and testing we can safely say that the question of diegetic and non-diegetic is relevant to the audio narrative in VR. By removing barriers like controllers and keyboards and in turn bring the person playing and the character in the game together, diegetic sounds contributes to an immersive experience.

We would like to stress that we are not saying that every VR-game must have diegetic sound only. It all depends on the genre and what kind of game one wants to make. It could be used as a tool for example, if one wanted to make the player feel distressed or uneasy, a non-diegetic sound could convey these feelings since the player won't know where the sound is coming from. It's also reassuring that audio in VR seems to be getting more attention since multiple new technologies (audio solutions) was released during the development of our game and study.

We would like to emphasize that what we concluded is not necessarily set in stone and that it rather serves as an inspiration to others. By continuing to explore these uncharted waters, we open up for further discussions regarding immersion and presence and their relationship with diegesis and the audio-visual contract. VR is rapidly gaining speed and we will surely see interesting and exciting experiments emerge in the future, expanding the capabilities of VR. In 1999 Frederick P. Brooks Jr made the statement in that VR barely works. Now 20 years later we would like to end this thesis with the following statement: *VR now works*.

5. Dictionary

Attenuation

- The gradual loss of energy through a medium. In our case, the reduction of volume intensity over distance.

Audible frequency range

- Range of which humans can hear sound, 20Hz-20kHz is the generally accepted standard range however personal deviations exists. Environmental factors also influence this range as well as age.

Binaural Audio

- Binaural audio aims to reproduce sound the way our human ear hears it through the use of binaural recordings i.e. a method with the intent of "placing" the listener in the 3-D stereo space with the recorded sound source(s).

Convolving

- An operation in which two functions merges and create in a third function. This third function is often viewed as a modified version of one of the original functions.

DAW

- Short for digital audio workstation, a virtual environment where sound designers create, edit, mix and arrange sound effects to be used.

Diegesis

- Our game creates a fictional world. Everything that occurs in this fictional world is called diegetic and everything that occurs outside (the world) is called non-diegetic.

Frequency

- Number of repeating occurrences per time.

HRTF

- HRTF (Head Related Transfer Function) is simply put, a function that simulates how an ear receives sound from a point in space. This function can essentially be used to simulate binaural recordings and spatialize audio to the listener when played through headphones.

Immersion

- The player ‘becoming’ the game character. The goal being that the player is within the game environment, that he is the character whose hands he sees before him.

Presence

- How the player feels himself inside the virtual world giving the experience of ‘being there’.

R&D

- Research & Development.

Sound cue

- A sound event playing a specific sound at a specific point in time.

Synthesis

- In audio, it refers to a synthesizer generating electric signals that are converted to sound waves with various methods of manipulating said signals.

Traditional audio narrative

- A narrative where sounds are divided into two separate categories, diegetic sound and non-diegetic sound. The most prominent example of this is the distinction between diegetic world sounds such as character sounds, weapons, ambience etc. and non-diegetic sounds such as user interface, loading screens, music etc.

6. List of References

Bowman, D. A., & McMahan, R. P. (2007). Virtual Reality: How much immersion is enough? *Computer*, 40(7), 36-43. doi: 10.1109/MC.2007.257

Brooks Jr, Frederick P. (1999) What's Real About Virtual Reality?. *IEEE Computer Graphics and Applications*, 19(6), 16-16. doi: 10.1109/38.799723

Chion, Michel. (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Christmas gift of the year. HUI Research. (2016). Retrieved 2017 February 28 from http://www.hui.se/nyheter/arets-julklapp-2016

Cypher, M., Richardson, I., (2006) *An actor-network approach to games and virtual environments: Proceedings of the Joint Computer Games and Interactive Entertainment Conference*, 4-6 December, 2006. (p. 254-259) Perth; Australia: Murdoch University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.1857&rep=rep1&type=pdf

Dykhoff, K. (2012). Non-diegetic sound effects. The New Soundtrack, 2 (2), 169-179. doi: 10.3366/sound.2012.0037

Ermi, L., & Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion. *Changing Views: Worlds in Play*, June 16-16, Toronto.

Grimshaw, M. N. (2007). *Sound and immersion in the first-person shooter: Proceedings from The 11th International Computer Games Conference, La Rochelle, France, November 21-23, 2007.* (p. 119-124) La Rochelle; France: University of Wolverhampton, School of Computing and Information Technology.

Kim, S., King, R. & Kamekawa, T. (2015). A cross-cultural comparison of salient perceptual characteristics of height channels for a virtual auditory environment. *Springer, 19*(3), 149-160. doi:10.1007/s10055-015-0269-1

Lam, J., Kapralos, B., Kanev, K., Collins, K., Hogue, A., & Jenkin, M. (2015). Sound localization on a horizontal surface: Virtual and real sound source localization. *Virtual Reality, 19(3-4), 213,222.* doi: 10.1007/s10055-015-0268-2

Law, J. (2012). *"*Notes on fish, ponds and theory*"*. Norsk Antropologisk Tidskrift, 23(3-4), 225-236

Nazemi, M., & Gromala D. (2012). Sound design: A procedural communication model for VE. Paper presented at the Audio Mostly Conference, Corfu, 2012. New York: ACM 16-23. doi: 10.1145/2371456.2371459

Nguyen, K., Suied, C., Viaud-Delmon, I. & Warusfel, O. (2009). Spatial audition in a static virtual environment: the role of auditory-visual interaction. *JVRB - Journal of Virtual Reality and Broadcasting, 6* (5), 1. doi: 10.20385/1860-2037/6.2009.5

Oculus. (2015). Powering the Rift. Retrieved 2017 March 02, from https://www3.oculus.com/en-us/blog/powering-the-rift/

Oxenham, O.J. (2013). Mechanisms and mechanics of auditory masking. *The Journal of Physiology, 591*(10), 2375. doi: 10.1113/jphysiol.2013.254490

Schafer, R. Murray. (1977). *The Soundscape - Our Sonic Environment and the Tuning of the world*. Vermont: Destiny Books.

Sherman, R. W, & Craig, A. B. (2003). *Understanding Virtual Reality: interface, application and design*. San Francisco: Morgan Kaufmann Publishers.

Smurdon, Tom. [Oculus]. (2015, 11 November). *Oculus Connect 2: 3D Audio: Designing Sounds for VR*. [Video file]. Retrieved from: https://www.youtube.com/watch?v=IAwFN9sFcso

Sonnenschein, D. (2001). Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema. Los Angeles: Michael Wiese Productions.

Sutherland, I. (1965) *The Ultimate Display, Proceedings of the IFIP Congress 65.* New York City, Washington. *May 24-29,* (p. 506-508) New York City; Washington: Spartan Books. Retrieved from: http://worrydream.com/refs/Sutherland%20-%20The%20Ultimate%20Display.pdf

Twisted Pixel, (2017) *Wilson's Heart*. Retrieved 2017 May 10 from http://twistedpixelgames.com/wilsonsheart/

Väljamäe, A. (2005). Self-motion and Presence in the Perceptual Optimization of a Multisensory Virtual Reality Environment. (Licentiate Thesis, Chalmers University of Technology, R037) Gothenburg: Chalmers University of Technology. Retrieved from http://www.cyberneum.de/fileadmin/user_upload/files/publications/Aleksander_Valjamae_05_LicentiateThesis_[0].pdf

Wingstedt, J., Brändström, S. & Berg, J. (2010). Narrative Music, Visuals and Meaning in Film. *Visual Communication*. Vol 9, (2), 193 - 210. doi: 10.1177/1470357210369886

6.1 Picture references

Figure 1: Nguyen, K., Suied, C., Viaud-Delmon, I. & Warusfel, O. (2009). Spatial audition in a static virtual environment: the role of auditory-visual interaction. *JVRB - Journal of Virtual Reality and Broadcasting, 6* (5), 1. doi: 10.20385/1860-2037/6.2009.5

Figure 2: Ermi, L., & Mäyrä, F. (2005). Fundamental Components of the Gameplay Experience: Analysing Immersion. *Changing Views: Worlds in Play*, June 16-16, Toronto.

Figure 3: Smurdon, Tom. [Oculus]. (2015, 11 November). *Oculus Connect 2: 3D Audio: Designing Sounds for VR*. [Video file]. Retrieved from: https://www.youtube.com/watch?v=IAwFN9sFcso

Figure 4: Oculus VR, LLC. (2017). Oculus Spatializer Plugin (Version 1.1.4.) [Plugin]. Available from https://developer.oculus.com/documentation/audiosdk/latest/concepts/osp-fmod-notes/

6.2 Technical specification references

Apple Inc. (2017). Logic Pro X (Version 10.0.7) [DAW]. Retrieved from https://www.apple.com/logic-pro/

Epic Games, LTD. (2017). Unreal Engine (Version 4.15) [Game Engine]. Retrieved from https://www.unrealengine.com/unreal-engine-4

Firelight Technologies. (2017). FMOD (Version 1.09.04) [Audio Middleware]. Retrieved from http://www.fmod.com/

Image-Line. (2017). FL Studio (Version 12.4.2) [DAW]. Retrieved from https://www.image-line.com/flstudio/

Oculus VR, LLC. (2017). Oculus Spatializer Plugin (Version 1.1.4.) [Plugin]. Retrieved from https://developer3.oculus.com/documentation/audiosdk/latest/concepts/osp-fmod-overview/

TortoiseSVN. (2016). TortoiseSVN (Version 1.9.5) [Revision Control Software]. Retrieved from https://tortoisesvn.net/about.html

7. Appendix

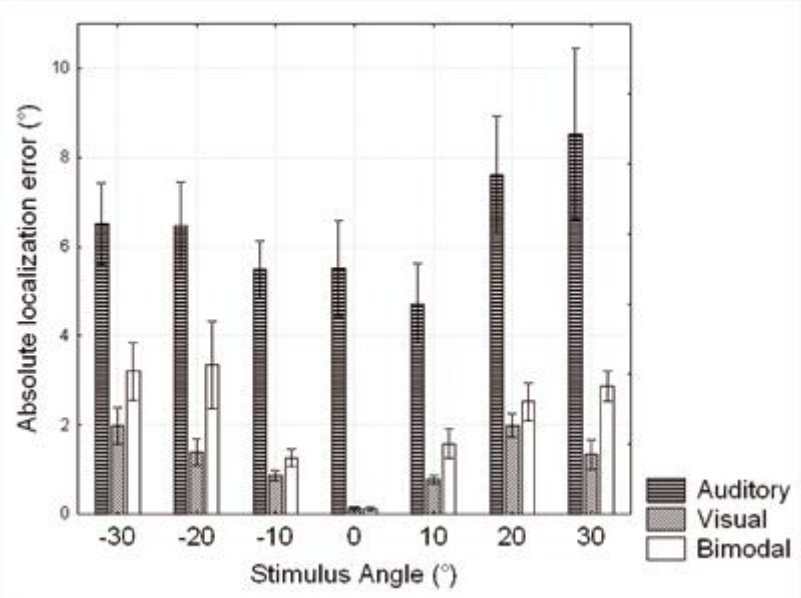

Figure 1. Average absolute localization error for each stimulus angle and in each condition, collapsed across all subjects and repetitions. Error bars represent one standard error of the mean (Nguyen et al., 2009).

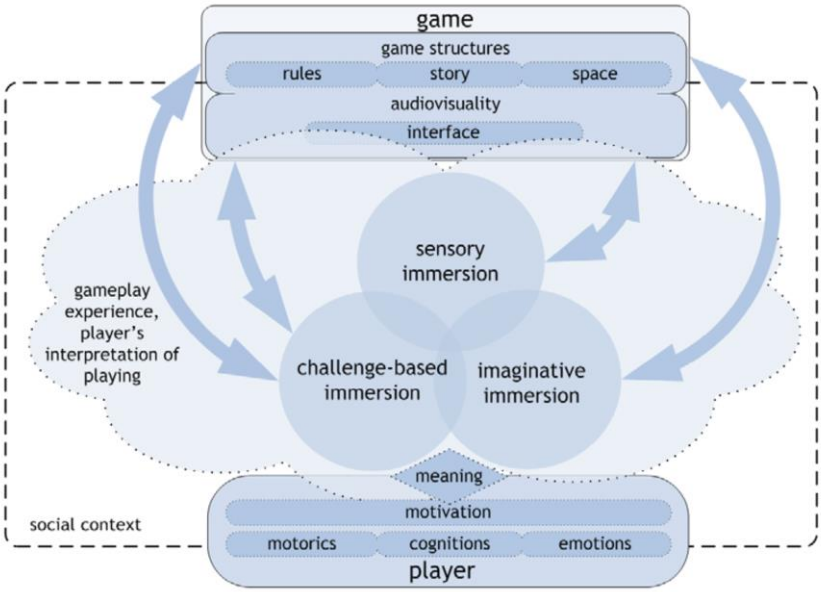

Figure 2: The three key dimensions of immersion by Ermi & Märyä.

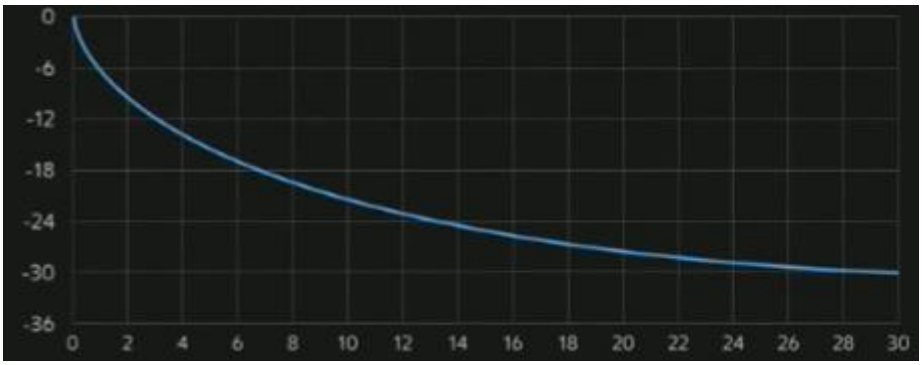

Figure 3. The inverse square law curve Tom Smurdon is referencing in his presentation.

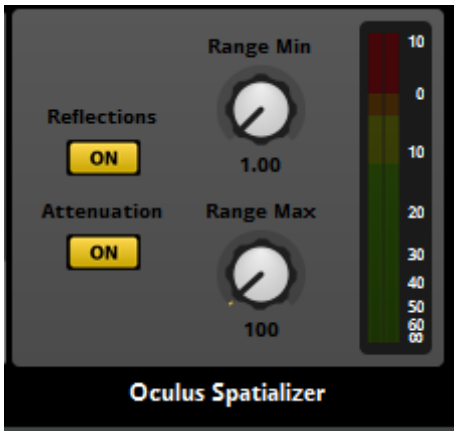

Figure 4. The Oculus plugin in the middleware FMOD studio. To the right is an output meter displaying the outgoing signal to the program's mixer, the two knobs represent minimum and maximum attenuation value.

8. Technical Specification

FL Studio

FL Studio (Image-Line, 2017) is a digital audio workstation (DAW) with its roots from being a MIDI only drum sequencer turned into a fully-fledged virtual studio. However, it's roots do not limit the overall sound designing possibilities. Utilizing a "lifetime free updates" policy, they believe that you should have the functionality you paid for, bug-fixed developed and updated as long as they develop FL Studio. This is a philosophy that Image-Line has maintained since they started developing FL Studio back in December 1997. The most powerful features of FL Studio are: fully customizable track and channel routing, pattern based workflow and automation of almost any parameter. While it lacked behind in technical features compared to its competitors for several years, the most recent update puts it at the same level of technical fidelity as the others.

FMOD Studio

FMOD (Firelight Technologies, 2017) is a middleware and audio content creation tool created by Firelight Technologies. With its visual interface it makes for a more efficient workflow for sound designers and audio programmers. It provides them with the same tools a DAW, like FL Studio or Logic Pro X, offers which makes it possible to edit sound effects in FMOD and by the press of a button get the new changes loaded into Unreal or Unity. It is free for educational purposes and companies with a budget under $500000 which is great for indie developers. By connecting FMOD to Unreal Engine, the game engine we use to develop our game, it provides us with a smoother workflow and allows us to make sound design changes in real time. The use of this middleware was also necessary to generate an easier workflow with the plugin Oculus Spatializer which creates the spatial audio or HRTF.

Logic Pro X

Logic Pro X (Apple, 2017) is a digital audio workstation and was first known as Notator Logic created by C-Lab in the 1990s. In 2002 it was acquired by Apple and their latest version is Logic Pro X. Logic provides recording possibilities, synthesizers, MIDI sequencer, audio samples and audio editing plugins for the creation of sound and music. It is only

available on the Macintosh platform. Similar to other DAWs it is essential for sound designers to be able to create, edit and mix sounds for games or films. The reason we have utilized two DAWs is simply because we are accustomed to them, is a matter of comfort. It does not matter which DAW one works in since all DAWs provide the same tools and possibilities.

Oculus Spatializer Plugin

Oculus Spatializer Plugin is a plugin allowing monophonic sound sources to be spatialized in 3D relative to the user's head location (Oculus, 2017). The spatializer has several parameters to allow for creativity and optimization such as: enabling early reflections, spatial reverb and full room reverb modeling.

TortoiseSVN

TortoiseSVN is an Apache Subversion client, implemented as a Windows shell extension (TortoiseSVN, 2016). It is a revision control/source control open source software allowing users to commit and update version of their work to a project collective directory. It can also revert files to a previous state/version in case of unforeseen consequences or unwanted changes. It allows team members of larger projects to share their work with their colleagues and test changes made with the press of a button. This helps the collective workflow of the project.

Unreal Engine 4

Unreal Engine is a game engine developed by Epic Games with a world-class toolset and accessible workflow allowing developers to quickly iterate on ideas and immediately see results (Epic Games, 2017). One big feature in Unreal is that you don't necessarily need to know how to operate code to create projects. At the same time, full source code access gives the community freedom to modify and extend engine features (Epic Games, 2017). Apart from traditional coding, Unreal offers something called "blueprint visual scripting" which is a more modular approach to programming where users connect nodes to implement gameplay logic. This allows users to create fully fledged projects without ever having to touch a line of code.