IMMERSIVE AUDIO IN VIRTUAL REALITY

Using spatial audio to enhance presence in VR training simulations

Graduation report

Luca Malte Frösler

Immersive Audio in Virtual Reality

Student: Luca Malte Frösler

Student number: 427683

E-Mail: 427683@student.saxion.nl

Graduation coach: Yiwei Jiang

Client Company: Strukton Rail

Company supervisor: Pieter Cornelissen

Project coach: Alejandro Moreno Celleri

Preface

I want to preface this graduation report by thanking my friends and family. I want to especially thank my girlfriend for her soothing support during the time of writing this thesis.

Additionally, I want to thank Yiwei Jiang for coaching me throughout the process of this thesis, as well as Alejandro Moreno Celleri for project guidance as well as additional support.

Moreover, I want to express my gratitude for Pieter Cornelissen from Strukton Rail, who has been invested in this project from the beginning and always provided good feedback as well as project guidance.

Back when I started my study Art & Technology at Saxion University in 2015, I did not really know what I want to do, but intuitively it felt like the right path. During my multimedia study I was introduced to virtual reality (VR), while curating other interests such as the creation of digital music as well as web, video- and audio design.

My journey into the world of VR development accelerated with the Thales VR simulation project two years ago. After developing another VR project about playing a piano with haptic gloves and fulfilling my internship in the VR/AR field half a year later, it was now time to graduate with another VR project.

I chose the Strukton project, again intuitively, since it is a meaningful VR use case in an interesting field.

Looking back at my study, everything feels how it should be. This project merges my two passions of VR development and audio together.

Abstract

Strukton Rail is interested in facilitating training and increasing employee safety awareness through a VR simulation.

In this paper it is discussed in how far audio can be used in order to increase the immersion of the simulation. Besides the well-known means of spatial audio, there are other means that can be used to increase user presence in a VR simulation. In this thesis, theoretical principles of achieving realistic audio are gathered and implemented in the VR simulation. In order to evaluate to what extend spatial audio helps to immerse the user into a simulation, a set of audio evaluation attributes are gathered. The attributes are used to develop a research method in which two versions of the VR simulation are tested by Strukton employees and evaluated comparatively. The results indicate that spatial audio improves user presence as well as perceived realism in a VR simulation, leading to a higher level of immersion.

Table of contents

1. Introduction

Today, riding trains as a means of transportation is strongly integrated into our cultures and livelihoods. Trains are supposed to be on time precisely to the minute and naturally we expect them to do so. When problems occur, they need to be resolved as quickly as possible, otherwise trains along the respective railways may be delayed or unable to ride at all. This can cause problems for travellers.

To prevent these incidents, companies like Strukton Rail provide maintenance solutions to ensure a smooth flow of the train system (Strukton Rail, n.d.). Strukton Rail's workforce tackles all kinds of problems that relate the rail infrastructure. The employees go through extensive theoretical trainings, eventually becoming electrical engineers. While they have a great repertoire of expertise, when they start, they often lack the practical experience of working in unfamiliar environments, since the scope of traditional training is limited.

Therefore, Strukton Rail is interested in exploring the immersive technology of VR as a means of simulating training experiences.

1.1. Client Objectives

The client is interested in a VR simulation that facilitates training and increases safety awareness for their railway maintenance workers. These workers are usually in an age group between 40 and 55.

At the current stage the client is focusing on developing the VR simulation for three different purposes:

1. Training for the railway maintenance workers
2. Demo for showcasing the possibilities and immersion of VR training for Strukton Rail
3. Give office employees of Strukton Rail an impression of on-site maintenance work

For the training the client intends a scenario that is common for their workforce: Working at heights. Using a KROL ("Kraan op lorries", meaning "crane on rail tracks"), different kinds of maintenance checks and repair works are done at heights.

In order to develop the product and come to a consensual vision, Strukton Rail initially put forward the following requirements:

Figure 1. Employees on a KROL working on catenary system (Strukton Rail, 2019)

- Realistic VR simulation
- Simulating dangerous work in a safe, non-harmful environment
- Training at night and day
- Different weather conditions such as fog, rain and snow
- Several levels of difficulty with less user guidance at higher difficulties
- Working under time pressure

While putting forward these requirements, Strukton is open for ideas and input for the product, giving the project group freedom to develop the simulation to their liking and trusting their growing expertise in the field of VR development. If the prototype serves to be useful, Strukton can use the created product as a basis for training and may continue the development of the simulation further.

1.2. Problem analysis

Strukton regularly faces the problem of absenteeism, which is caused by work accidents on site. To diminish this problem, Strukton is interested to make use of a VR safety training simulation, which raises safety awareness and simulates dangerous scenarios in a non-harmful environment, enhancing employee work preparation. Through this, employees do not only have the theoretical knowledge of their traditional training, but also gain practical knowledge in VR. In the simulation workers are provided with solutions for problems, as well as guidance that ensures safe working conditions. Strukton's use case of a VR simulation would be a reasonable solution for increasing safety awareness of their employees, hopefully leading to more safety on-site and therefore in the long-term, less absenteeism.

The problem that arises is that the VR simulation might not immerse the participant enough. A lack of immersion could distract the participant and would lead to the user to not feel present in the virtual environment (VE). This could lead to a lack in engagement and ultimately less effective training.

2. Problem definition

In order for the VR simulation to be a success, it requires the user to feel present in the VE. It is to be investigated how spatial audio can be utilized in order to reach a high level of immersion. Only if the VR experience is immersive, the participant will feel like being really "there". This is required, in order to have the trainee learn about safety awareness and possible dangers on railway sites. If the VR simulation serves to be immersive the trainee will feel present in the VE and be able to do training effectively, leading to higher awareness about safety and dangers.

3. Scope

The focus in this graduation report will focus on how far audio within a VR simulation can be used to increase the level of presence for the user. Reaching a high level of immersion in VR is a complex task and is mainly dependent on hardware-related specifics (field of view, display resolution, frame rate, etc.) as well as visual factors, such as the realism of the 3D models. In this graduation report I will not dive into these topics but focus on the audio part of a VR simulation, with the goal of solving the problem of a lack of immersion for the participant. Since it needs to be tested, how far immersive audio enhances the subject's presence in the VE, I cannot take all factors that contribute to presence (such as in-game user stimulation) into account. Instead I will keep the focus on leveraging the audio to increase user presence.

In addition to that, I will not dive into all audio-related software that is out there, but toolsets that can be integrated with the Unity game engine. Next to that, I will focus on the middleware FMOD, which is used for audio implementation.

Parallel to the work of this graduation report a full-time VR group project is conducted, as part of the minor Immersive Media at Saxion University. In a scope of 20 weeks a VR safety trainer prototype is to be created for Strukton Rail. The product is created within a project group

consisting of five graduates from Creative Media & Game Technologies, including me, as well as three ICT Software Engineering students who are doing this project as their minor.

My role in the project is VR development using the Unity game engine and C# scripting, as well as the curation and implementation of audio. Through this, I can write my thesis and conduct research about the named topic, while at the same time utilizing and iterating my researched approaches and solutions throughout the development of the product. In addition, I am the project leader of the project group, which includes client- and teacher communication and project planning.

4. Main and Sub Questions

4.1. Main Question

To what extend can immersion in a VR training simulation be increased through the use of spatial audio?

4.2. Sub Questions

What other auditory factors contribute to immersion in VR?

I ask this question since there are other factors besides spatial audio that contribute to immersion.

What toolsets are suitable for the design process of spatial audio in VR training simulations?

I have added this question, since it is not apparent which toolsets are the most suitable for this VR training simulation.

How can immersion through spatial audio be tested?

I ask this question because the level of immersion needs to be evaluated.

Are there possible ethical issues involved in the design process?

A question that discusses if the implementation could arise possible ethical issues such as copyright infringements.

5. Approach

In order to answer the question of which audio factors contribute to immersion in VR, desk research is necessary. It needs to be investigated, what factors in the area of audio make up immersion and increase the feeling of presence for a user in VR. Existing case studies that involve the issue of immersion in VR, through means of audio, should be reviewed. The results of these studies will be analysed and integrated into this report.

Furthermore, theoretical research should be conducted on how to emulate spatial audio and how humans perceive sounds in real-life. In addition to that, existing SDK's that provide spatial audio solutions should be researched and analysed based on their suitability for this case. The SDK that fits the case the best should be selected, utilized and experimented with throughout the scope of the product development. In addition to that, other means of achieving a realistic audio representation should not be left out. If adding additional value, they should also be incorporated and experimented with throughout the development of the VR simulation.

Moreover, user tests should be conducted. If possible, there should be a few instances of user tests throughout the development of the product. In each instance a number of users should experience the respective current state of the VR simulation for a short amount of time. Afterwards, they are given a survey in which they rate the perceived audio based on several criteria. The criteria need to be researched and may include the level of perceived realism, as well as the perception of presence in the virtual scene.

Concerning the research method, two versions of the VR simulation shall be created, which differ in their audio settings respectively. One version will have crude (non-spatial) audio, while the other should contain spatialized audio. After this, the users will answer a questionnaire to give feedback about different audio properties for each version of the VR. A pilot test shall be conducted in this manner, which can be improved if needed. At last, a final test with two employees from Strukton will be conducted. With the acquired theoretical knowledge and results from this test, conclusions will be drawn, and the main and sub questions will be answered respectively.

6. Theoretical framework

6.1. The potential of immersive virtual reality

When one is experiencing VR through a head-mounted display (HMD), they know that it is not the same kind of reality that they experience on a day-to-day basis, but that the world that they are in is artificial. Although the participant is aware of this, the experience is still *real*. VR enthusiast J. Lanier stated, as early as 1989, that the virtual world is exactly as real as the physical one while at the same time it "has this infinity of possibility that you don't have in the physical world" (Conn, Lanier, Minsky, Fisher, & Druin, 1989). VR enables unique experiences, such as allowing us to experience familiar things from different perspectives as well as being able to experience infeasible or normally impossible things. It is a new dimension creating a link between our perception and emotions and the infinite possibilities of our imagination. Compared to traditional computing technology, VR is different in the way that it focuses on immersion (Mestre R., n.d., p. 1). "In this sense, VR (and more generally computerized devices) really acts as a problem-solving device, transforming enormous quantities of mind-breaking data into "graspable illusions"" (Mestre R., n.d., p. 1). Due to this potential, VR has been utilized in many different fields, such as exposure therapy for various anxieties, education, health care, mental health, manufacturing, space, museums, as well as industrial training. Studies have shown, that practical construction training in VR is more engaging for the participant and increases the learning effect, opposed to traditional training (Sacks, Perlman, & Barak, 2013, p. 1016). Furthermore, the trainee's ability of being alert for longer periods of time is higher on average when experiencing training in VR (Sacks, Perlman, & Barak, 2013, p. 1016).

6.2. Short history of spatial audio

In 1933, Alan Blumlein criticized that sound from a single sound source lacks realism. Subsequently, he created a system where audio is recorded and played through two channels instead of one, inventing stereo sound as we know it today (Shankleman, 2008). Blumlein's innovation of spatial audio established the foundation for immersive sound experiences and brought mankind one step closer to emulating human auditory perception.

Taking a leap to the 21ˢᵗ century, the theory of spatial audio has been put into practical use and has been integrated in a variety of technological solutions. Not only in cinema, where now Dolby Atmos is the new standard, which places the viewer in midst of a scene, with up to 128 audio channels (Dolby Atmos Cinema Sound, n.d.). But also, in virtual reality, the potential of spatial audio has been explored and became a requirement to increase the participant's presence. Game developers have seen the need of utilizing life-like audio in their VR applications to create experiences that stretch the boundaries of immersion.

6.3. Means of achieving immersive audio

Audio is necessary to increase user presence, as well as creating an emotional impact. By adding accurate sounds to the things that the participant is seeing, the experience feels wholesome. When looking at traditional media, such as movies, audio is a very important aspect that immerses the viewer more into the depicted story. Classic sound effects, like the flying of a "Tie Fighter" spaceship in the "Star Wars" movies or the scream of the "T-Rex" in "Jurassic Park" will never be forgotten. They are iconic because the viewer does not question their authenticity. They make the audio-visual experience wholesome and believable. It is long known that the visual and auditory sense are strongly connected, letting audio-visual events appear "as one" (Spence, 2007, p. 66). Therefore, if one can see an event, one should also be able to hear it.

In order to create a wholesome aural experience, soundscapes are created that immerse the user into the story. In traditional film there is a distinction of two types of soundscapes, one being diegetic sound and the other being non-diegetic sound (Bordwell & Thompson, 2001).

"Diegetic sound is a form of sound that originates from the game environment the game simulates, or the environment the film represents. Diegetic sound can consist of urban commotion or of birds singing. Nondiegetic sound equals the musical soundtrack that usually changes according to the events in the film, or game (Järvinen, 2002, p. 119)."

An immersive and wholesome story profits from both diegetic and non-diegetic soundscapes. In a similar manner, audio for VR experiences should be wholesome. For the specific case of this VR simulation I would argue that the focus lays on realistic diegetic sounds. Since the nature of the simulation is rather a serious game and not a game that incorporates storytelling or tension, non-diegetic sounds such as used in movies will only be used to a minimal amount,

if used at all. Diegetic sounds have the purpose to immerse the user into the story. The participant should not feel that something is off, at any point in the VR simulation. When it is raining, or a train is passing by, it should be heard in VR exactly how one hears it in real life. Moreover, diegetic sounds that we take for granted in our everyday lives, such as the sound of dropping a hammer onto concrete, will be implemented respectively.

Additionally, for the purpose of raising safety awareness for the user, non-diegetic sounds may be incorporated. A use case could be dynamic feedback for the participant. If there is e.g. a danger icon popping up in the user interface (UI), a fitting sound should go with it. A good sound effect is one, where its sonic representation itself conveys all the information that is needed and aims at the user's emotions. In that way it ensures for a smooth user experience, and other means of conveying information, such as text, can be narrowed down to a minimal amount. This leads to less user distraction and a higher level of presence.

Non-diegetic sounds could be in the form of notifications or prompts, that indicate oncoming danger. These sounds could also notify the user when he/she does a mistake or takes a risk, as well as if he/she is working in an unsafe manner. Therefore, it is crucial that these sounds are designed appropriately and convey the respective urgency, depending on the type of notification. By using sound design means such as alterations in pitch, frequency, loudness and playback speed, the desired importance of the sound can be incorporated. Intuitively we know that sounds that are high in volume and pitch are very alarming, for example the sound of a police siren or an alarm clock. These sounds are deliberately constructed in this way. Science has shown that at above moderate levels of volume (more than 40 dB) response time of a recipient decreases and "perceived urgency increases as fundamental frequency increases […]" (Haas & Edworthy, 1996, p. 196), meaning that a higher pitch frequency increases the general attentiveness of the recipient.

Subsequently, depending on the danger warning or notification in the VR simulation, its auditory representation should be designed in a way that the safety and danger awareness for the participant will be increased and the response time of the user will be decreased respectively. A crucial difference between VR and traditional media is the presence of a third dimension. In VR, the participant can stand inside the scene of

action and e.g. face an enemy. Therefore audio, should emphasize this presence, by displaying the aural information correctly in the virtual world. Sounds that are appearing around the participant should be heard from the direction of where they are being emitted from, like in real-life, enabling the user to localize audio-visual events. Furthermore, it is important that audio levels are mixed together in a way that they are balanced and sound natural. The absolute volume of a sound is not as important, since the user can adjust the overall loudness, but the proportions between all sounds need to be adjusted carefully in a sound mixer. In the implementation stage, a list of noise levels in respect to distance can be useful to design sound volumes realistically (Figure 2).

Sound sources (noise) Examples with distance	Sound pressure Level L_p dB SPL
Jet aircraft, 50 m away	140
Threshold of pain	130
Threshold of discomfort	120
Chainsaw, 1 m distance	110
Disco, 1 m from speaker	100
Diesel truck, 10 m away	90
Kerbside of busy road, 5 m	80
Vacuum cleaner, distance 1 m	70
Conversational speech, 1 m	60
Average home	50
Quiet library	40
Quiet bedroom at night	30
Background in TV studio	20
Rustling leaves in the distance	10
Hearing threshold	0

Figure 2. Sound source examples with corresponding sound pressure level (Table of Sound Pressure Levels, n.d.)

In addition to mixing the volumes of sounds, frequencies need to be mixed correctly as well. Both of these means can be achieved by using a common audio editing program, also called Digital Audio Workstation (DAW) such as Ableton, Logic or FL Studio.

Moreover, sound samples that are used should be of high quality, possibly recorded with a spatialized microphone such as Zoom H2n. Sounds can be sampled from sources on the web, although it is important that acquired samples have a royalty-free license, to avoid copyright infringement.

Another factor to get closer real-life acoustics, is to emulate the reflections of an acoustic room, by making use of reverberation. In the physical world reverberation is ubiquitous and human hearing has adapted to it, e.g. to sense spatial information. "Simulating the acoustic room response is essential for a realistic sound environment. The level, time, and density of early reflections and the reverb tail provide an impression of the size, form and materials of the surrounding environment (Naef, Staadt, &

Gross, 2002, p. 66)." Therefore, this could be another element than enhances realism through audio.

In addition to that, another important aspect of realistic audio is sound occlusion. Occlusion is the (partial) dampening of frequencies by physical objects which are placed between the sound source and the listener. Usually high frequencies are dampened more. One knows this phenomena for example when standing outside of a nightclub and mid- and high frequencies of the music are dampened, while low frequencies are transmitted to a greater extend. To emulate occlusion on a simple level, low-pass filters can be used (Ibánez, Alvarez, & Peinado, n.d., p. 2). Occlusion is a crucial factor of acoustics in real-life and should therefore be implemented into the VR simulation.

By using techniques like these, it is hypothesized, that a natural and realistic sounding representation of the environment can be achieved. The specific means that can be used by developers to enhance the realism of perceived audio in VR, vary and are limited by the toolsets used, which is discussed in chapter 6.6.

6.4. Spatial Audio

Although we only have two ears, we can localize sounds in three dimensions. When hearing, our brain constantly determines the origins of sounds in our environment. The brain receives sonic information from the outer and inner ear and analyses the subtle difference between the intensity and time it takes for a sound to reach both ears. It also considers how sound waves are reflected around our head and inside our ear. This human skill of precise sound localization sets a high bar for creating immersive sound experiences.

Scientists have researched this phenomenon of sound localization thoroughly and characterized it as a head-related transfer function (HRTF). This function describes the modifications to a sound from being emitted at a source to arriving inside the eardrum. The specific modifications which include an approximation of the individual form of the head and outer ear of the listener provide information to render a sound precisely (HRTF, 2016).

HRTF enables the rendering of common audio into different types of spatial audio. On one hand there is common stereo, on the other hand, there is transaural audio. Transaural audio provides "precise and easy localization" (Guastavino, Larcher, Catusseau, & Boussard, 2007, p. 58), but lacks reaching a high level of immersion (Guastavino, Larcher, Catusseau, &

Boussard, 2007, p. 58). Since transaural audio is emitted via loudspeakers, it can lead to crosstalk, meaning that "[…] the signal reproduced by one loudspeaker arrives at both ears" (Kaiser, 2011, p. 2). This is something that needs to be avoided at all costs, in order to not break immersion. Therefore, binaural audio should be used. Binaural audio is a type of surround audio that sends signals to the left and right ear individually, requiring the use of headphones (Transaural Rendering, 2006). A common sound format and technique of rendering binaural audio is *Ambisonics*. Ambisonics is unique, because it can cover sound sources in all three dimensions, including sound sources which are above and below the listener, which encapsulates the listener into a 360° auditory environment, like in real-life.

Research has shown, that Ambisonics provide the highest level of immersion (Guastavino, Larcher, Catusseau, & Boussard, 2007, p. 58). Thus, to create a highly immersive VR experience, Ambisonics should be used.

6.5. Spatial Audio Quality Evaluation Principles

According to Berg and Rumsey (n.d.), who did extensive research on systematic evaluation of spatial audio quality, the attributes listed in Figure 3 are the most important when it comes to assessing the quality of spatial audio. It should be noted that their research methods consisted of listeners assessing speech and instrument recordings which have been spatialized to different degrees. Therefore, not all listed attributes may be used for the audio evaluation of the VR simulation. Interestingly, their research has shown, "that an enveloping sound gave rise to the most positive descriptors and that the perception of different aspects of the room was most important for the feeling of presence. (Berg & Rumsey, n.d., p. 7)" This indicates that the source envelopment attribute plays a major role in establishing presence for the user. This is why the qualitative research will inter alia put a high emphasis on this attribute.

Attribute	Description
Naturalness	How similar to a natural (i.e. not reproduced through e g loudspeakers) listening experience the sound as a whole sounds.
Presence	The experience of being in the same acoustical environment as the sound source, e g to be in the same room.

Preference	If the sound as a whole pleases you. If you think the sound as a whole sounds good. Try to disregard the content of the programme, i.e. do not assess genre of music or content of speech.
Low frequency content	The level of low frequencies (the bass register).
Ensemble width	The perceived width/broadness of the ensemble, from its left flank to its right flank. The angle occupied by the ensemble. The meaning of "the ensemble" is all of the individual sound sources considered together. Does not necessarily indicate the known size of the source, e.g. one knows the size of a string quartet in reality, but the task to assess is how wide the sound from the string quartet is perceived. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the width of the sound source.
Individual source width	The perceived width of an individual sound source (an instrument or a voice). The angle occupied by this source. Does not necessarily indicate the known size of such a source, e g one knows the size of a piano in reality, but the task is to assess how wide the sound from the piano is perceived. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the width of the sound source
Localisation	How easy it is to perceive a distinct location of the source – how easy it is to pinpoint the direction of the sound source. Its opposite is when the source's position is hard to determine – a blurred position.
Source distance	The perceived distance from the listener to the sound source.
Source envelopment	The extent to which the sound source envelops/surrounds/exists around you. The feeling of being surrounded by the sound source. If several sound sources occur in the sound excerpt: assess the sound source perceived to be the most enveloping. Disregard sounds coming from the sound source's environment, e g reverberation – only assess the sound source.

Room width	The width/angle occupied by the sounds coming from the sound source's reflections in the room (the reverberation). Disregard the direct sound from the sound source.
Room size	In cases where you perceive a room/hall, this denotes the relative size of that room.
Room sound level	The level of sounds generated in the room as a result of the sound source's action, e g reverberation – i e not extraneous disturbing sounds. Disregard the direct sound from the sound source.
Room envelopment	The extent to which the sound coming from the sound source's reflections in the room (the reverberation) envelops/surrounds/exists around you – i.e. not the sound source itself. The feeling of being surrounded by the reflected sound.

Figure 3. Attributes of Spatial Audio Evaluation (Berg & Rumsey, n.d., p. 12)

6.6. Software Development Kits

As a developer one can choose from a few different software development kits (SDK's) that enable the integration of HRTF-based spatial audio. Due to the fact that the SDK's differ in their functionalities, as well as offer exclusive features only supported by certain SDK's, it is to be investigated, which SDK is most suitable for creating realistic audio in a VR simulation. Current state-of-the-art SDK's that offer spatialized audio are inter alia:

- Steam Audio by Valve
- Resonance audio by Google
- Oculus Spatializer by Oculus

It should be noted that all of these SDK's can be used within Unity (or Unreal), usually without the need of much additional scripting. Additionally, these plugins also work through the middleware software FMOD. With FMOD, developers have greater control over their audio design in either Unity or Unreal and are also able to implement spatial audio by using plugins of the above-mentioned SDK's. All of these SDK's include basic functionalities such as equalizing, filters and (physic-based) distance and near-field attenuation, meaning the decrease of sound volume when the listener is moving away from a sound source. However, the SDK's differ in their unique features. When looking at attenuation, Oculus Spatializer offers a unique feature called the volumetric radius. "This [feature] lets you set a radius around the

point source where the sound should feel enveloping and omni-present (Gould, 2018)."
Google Resonance has a similar feature and calls it spread. On the other hand, Steam Audio
offers air absorption, which absorbs higher frequency levels more, when moving away from
sound sources. What's interesting about Google Resonance, is that it offers the most flexibility
when it comes to sound directivity, enabling for creative use cases (Gould, 2018). Another
major feature is sound occlusion. This feature is supported by Steam Audio as well as
Resonance Audio, but not by Oculus Spatializer. It should be noted that, although Google
Resonance offers sound occlusion, when using middleware such as FMOD, out-of-the-box
real-time sound occlusion is not possible and would require workarounds in the game engine.
Steam audio offers the biggest flexibility here, as well as multiple options of processing the
occlusion. Since it uses sound occlusion based on physical rendering, the developer is required
tag geometry and specify materials of objects in the game environment. If implemented
correctly, this may yield a more realistic sound representation (Gould, 2018).

7. Implementation

7.1. Scene setup

The scene consists of a KROL on a rail track which is situated on a bridge in a city environment.
There are catenary systems along the bridge, as well as a table with a hardhat that can be
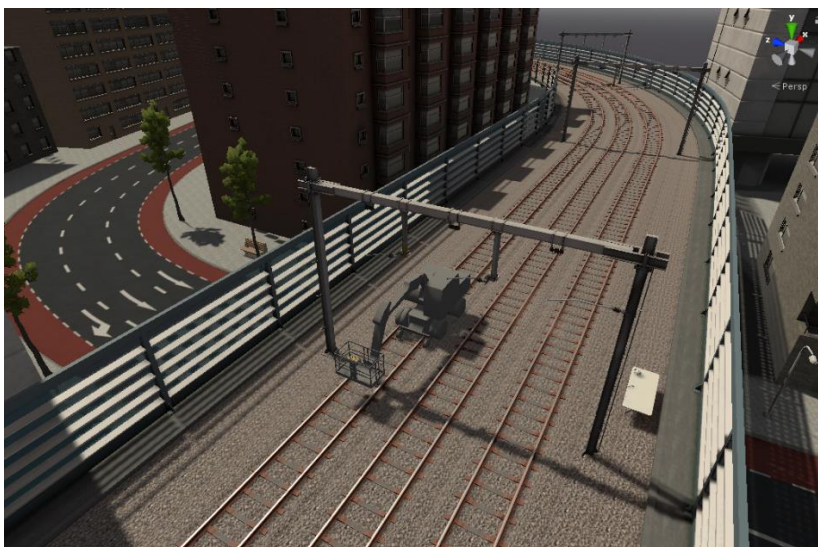worn, and an adjustable wrench which is used to tighten bolts on the catenary systems.

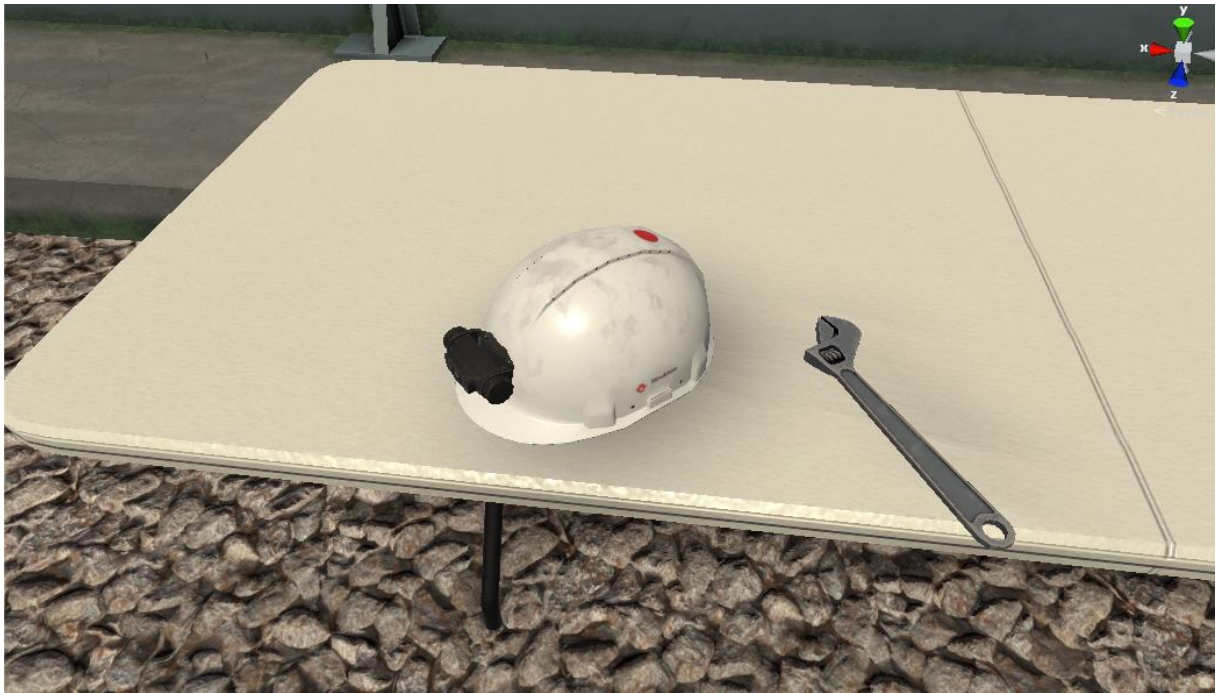*Figure 4. Main scene with rail track bridge, KROL (grey box) and catenary system*
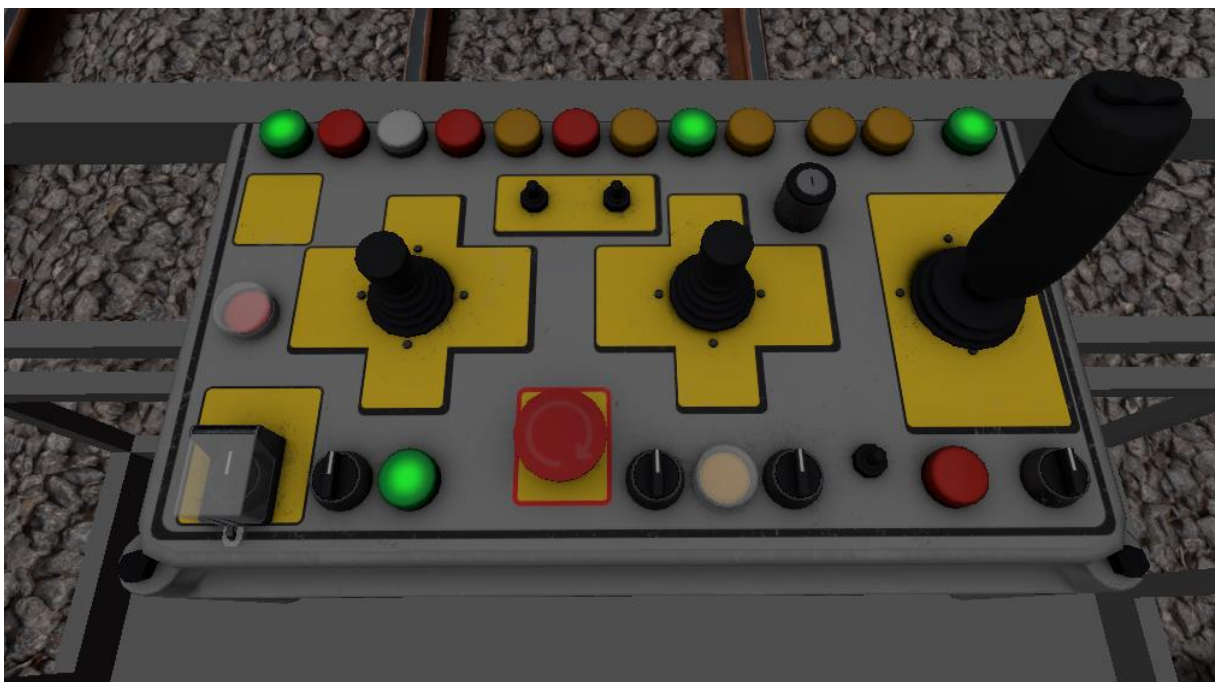
Figure 5. Interactable objects on table

Figure 6. KROL control panel inside the KROL basket

7.2. Project Setup and FMOD Integration

To have more control about audio events, and also to learn about game audio software, I used FMOD to handle audio played in Unity. FMOD has the advantage that it can be easily integrated with a Unity project. In addition to that, all major spatial audio SDK's are supported by FMOD. Another advantage of FMOD is that after an FMOD project has been built and connected to the Unity engine, the FMOD master sound bank can be easily updated at any time. This is handy, because a lot of development is done in different scenes simultaneously, which

Figure 7. FMOD build integrated with Unity

all use the same sound bank. After adding the "FMOD for Unity" package into the Unity development scene, I was able to link the FMOD build to the Unity project (see Figure 7).
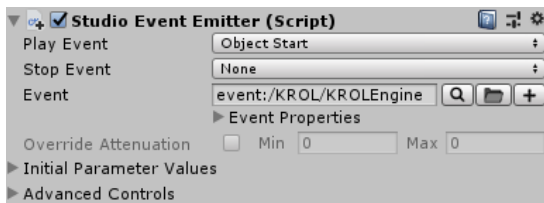
7.3. Workflow

Figure 8. Studio event emitter component to play sound events

After linking the FMOD build file with the Unity project, my workflow consisted of creating 2D or 3D sound events in FMOD, which are send to FMOD's mixer. After building the FMOD project file, Unity has access to these sound events. In order to trigger a sound event in Unity, the script component "Studio Event Emitter" by FMOD is used. Within the component one can set triggers to play or stop the event. Additionally, one is able to set initial parameter values, in case parameters have been added to the event in FMOD. This method is fine for playing sound events that should be
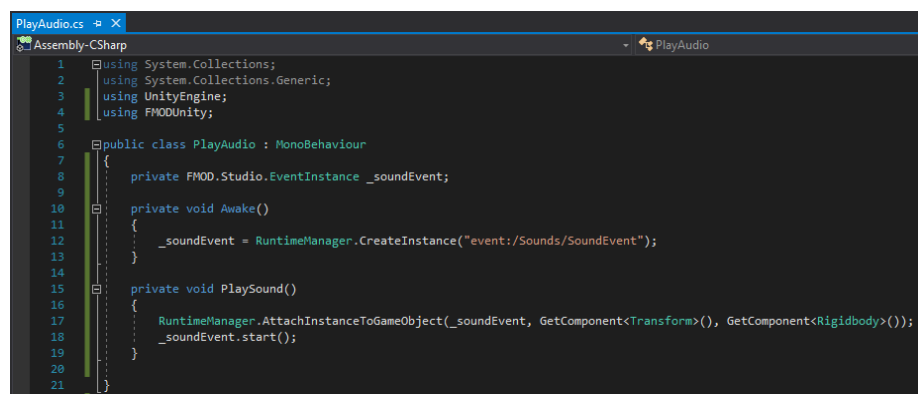
Figure 9. Example code of initializing and playing an FMOD sound event

played continuously but lacks flexibility when triggering sounds at more specific events. To solve this problem, sound events are triggered through code (Figure 9).

Before creating sound events I sample and edit the audio. You can read more about my sampling workflow in chapter 13.8.

7.4. SDK Implementation

From the beginning of the project I implemented audio through FMOD. Based on my research I discovered that the highest level of immersion can be yielded by making use of Ambisonics. Another motivation to implement spatialized audio was the fact that localization of objects can be improved, which, according to my research, is another factor of real-life acoustics. Therefore, during week 12 of the development process, I implemented the Google Resonance plugin into FMOD to spatialize the audio. To fully spatialize the scene, I had to copy all existing
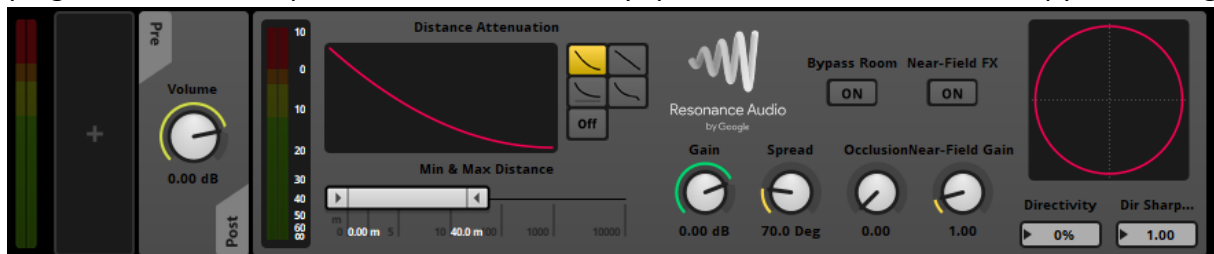

Figure 10. Resonance audio source in FMOD

FMOD sound events and replace their audio source component with the Google Resonance audio source. This component enables one to set different parameters concerning spatial audio. The implementation within Unity was straightforward and required an import of the "Google Resonance for FMOD" package. With this the spatialization of an audio source worked out of the box and results were impressive. A downside when it came to sound occlusion was that I can only manually set the occlusion parameter to a value between 0 and 1, which "muffles" the sound more or less. After some research I learned that without extensive workarounds it is not possible to have dynamic sound occlusion, e.g. when an object obstructs the line of sight between the user in VR and a sound emitting object, the sound should occlude respectively. This was a problem, since according to my research, sound occlusion is a major part when it comes to creating a realistic sound representation. Because of this, I switched from Google Resonance to the Steam Audio SDK, since Steam Audio supports occlusion out-of-the-box.

Figure 12. Steam Audio Spatializer

The Steam Audio Source (see Figure 12) is similar to that of Goole Resonance but has some differences. The HRTF Interpolation can be set to either "Nearest" and "Bilinear". With "Nearest" abrupt changes in frequency can be heard when rotating around a sound source (Gould, 2018). This is due to the HRTF being swapped. With the "Bilinear" setting there is interpolation and the audio sounds smooth. It is noticeable that custom attenuation can be set and only physics-based-attenuation can be enabled.

Figure 11. Steam Audio Geometry and Material components

This requires the geometry inside the Unity scene to be tagged, otherwise spatial audio will sound faulty. This can be done using the Steam Audio geometry component (Figure 11). Additionally, to achieve realistic occlusion, materials of objects in the scene can be set with the Steam Audio material component. Steam offers eleven different material presets, but the user is also able to use their own custom material and set all values of frequency transmission, scattering etc. themselves. While experimenting with the different materials, I noticed, that the cars that drive under the bridge are occluded so much that they are barely audible, which felt a bit unrealistic. Due to this, I created a custom material, applied it to the glass walls of the bridge and changed the values until I perceived the occlusion as realistic.
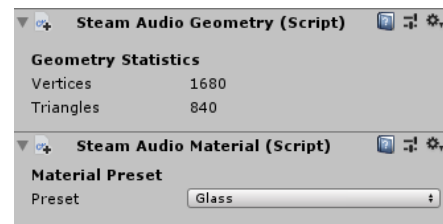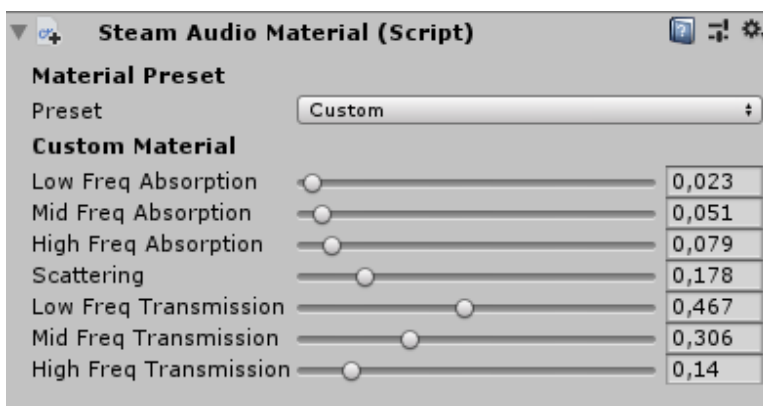

Figure 13. Custom material to achieve desired occlusion

7.5. Sound Features

7.5.1. KROL

The KROL is the main interactable object in the scene. Based on my research, it is evident that perceived visuals are strongly connected to their sound. Therefore, the KROL requires a realistic auditory representation. The following sounds have been added to the KROL game object in the scene:

Figure 14. FMOD speed parameter with pitch modulation

- Engine sound (always looping)
- Sound when lifting the basket of the KROL
- Sound when rotating the basket
- Back up beep sound (when driving backwards)

When one is driving the KROL forward or backward with the right joystick on the control panel, the engine sound goes up in pitch, mimicking acceleration. This is done in FMOD, by using a parameter (in this case "speed"). I linked the value of the speed parameter to the pitch of the engine loop (see Figure 14). Then in Unity, through code I map the actual moving speed of the KROL to the value of the parameter (0 = actual pitch, 1 = maximum value of increased pitch) to change the engine pitch in real-time. In a similar fashion, I created a sound event when moving the basket of the KROL up and down. The sound event consists of a start sample (basket lifting off), a looped section (while moving), and an end event (basket stopping). Again, I am using a parameter in FMOD which I control from Unity. Through changing the value of the parameter, I can trigger different time marks of the sound event. Moreover, I have done something similar for when rotating the KROL basket with the left joystick. On a side note, to convey the heavy mass and significance of the KROL I increased the loudness of the engine quite a bit before importing it into FMOD, since according to my research higher volumes is one of the factors to increase alertness of the participant.

7.5.2. Feedback of interactable objects

To increase realism of the simulation I coded a system that plays sounds when objects are dropped onto the ground. The loudness of the sounds is in relation to the force with which the objects are thrown. A different ground hit sound event is played, depending on the material of the ground on which the object is dropped onto (which can be set via tag). To increase the feedback for the user I also added sound effects when the user is picking up either the hardhat or the wrench. For both the hardhat and wrench sound event I incorporated a multi instrument in FMOD, which plays a different sample out of an array of samples, each time the sound event is triggered. The idea is that the samples themselves convey the feeling, weight and material of the object that is picked up.

7.5.3. Environment ambience

Through my literature research, I learned that an immersive story profits from diegetic sounds. This includes sounds that are not emitted from points in the field of view, but which appear to be emitted from the environment. Thus, in order to enhance the immersion, I have added a sound ambience to the scene. I have created a 2D sound event loop which consists of several layered samples such as city noise (heavy in low frequencies), traffic in the distance, wind, and birds or crickets (depending on if the scene is currently set to day- or night-time). For the traffic, I recorded a clip of cars passing under a bridge in Enschede. The ambience sounds are kept 2D and non-spatial, since they are not coming from any particular direction but are ubiquitous.

7.5.4. Traffic System

As mentioned in the previous chapter, part of the ambience has been a sound of cars passing under a bridge. This was okay, but I wanted to have actual moving cars around the scene that emit sounds from their positions, to make the scene more alive. I created a traffic system, where cars spawn with randomized times and follow predetermined routes, going along the streets and under the bridge in the VE. Each car has the same sound event of a car engine driving. For each instance of the sound event, the car sound is played from a randomized time mark in FMOD, so two cars don't play the sound synchronously. The implementation of the traffic system added a lot of the, otherwise static scene.

7.5.5. Room Acoustics

According to my research, simulating the properties of an acoustic environment through reverberation is an important aspect to increase realism. To achieve this, I have created a Steam Audio probe box in Unity (Figure 15), which takes all tagged geometry into account. To create the "reverb zone", the probe box needs to be baked. On the FMOD audio event which should have reverb I can then click "Indirect" to allow for the reverberated signal to become audible. I also enable "Indirect Binaural" for the signal to become spatialized, which according to Steam Audio "gives a better sense of directionality to indirect sound and improves immersion" (Valve Corporation, 2017).

Figure 15. Probe box (pink) to compute reverb

7.6. Ethics

In order to populate the scene with audio, I had to find ways to acquire fitting sounds. For most of the sounds, it was not applicable for me to recreate the sounds on my own, because of limited time and resources. Therefore, most sounds were sampled from sources online or from my private sample collection. In order to avoid any copyright infringements, I made sure that all rights of the used sound samples were reserved. Such "royalty-free" sound files can be found on various websites such as freesounds.org, or websites that list sound samples which are in the public domain.

8. Research methodology

8.1. Preliminary research method

Around week 12 during the development process of the simulation, two Strukton employees visited the project group for a Q&A and also to provide feedback on certain aspects of the VR simulation. A general survey has been created to ask the two employees about their work experience. A second survey was created about certain audio aspects which was filled in by the two Strukton employees. You can find this survey in the annex (chapter 13.5). Before answering this survey, both subjects tested the VR simulation in the current state of week 12 for about 10 minutes each. At this point in the simulation a first iteration of the audio had been implemented. This included environmental sounds, feedback sounds of the KROL, such as moving the platform up and driving back and forth, as well as object drop sounds. It was stereo audio meaning it used basic panning, enabling the user to localize objects, but spatial audio had not been implemented yet.
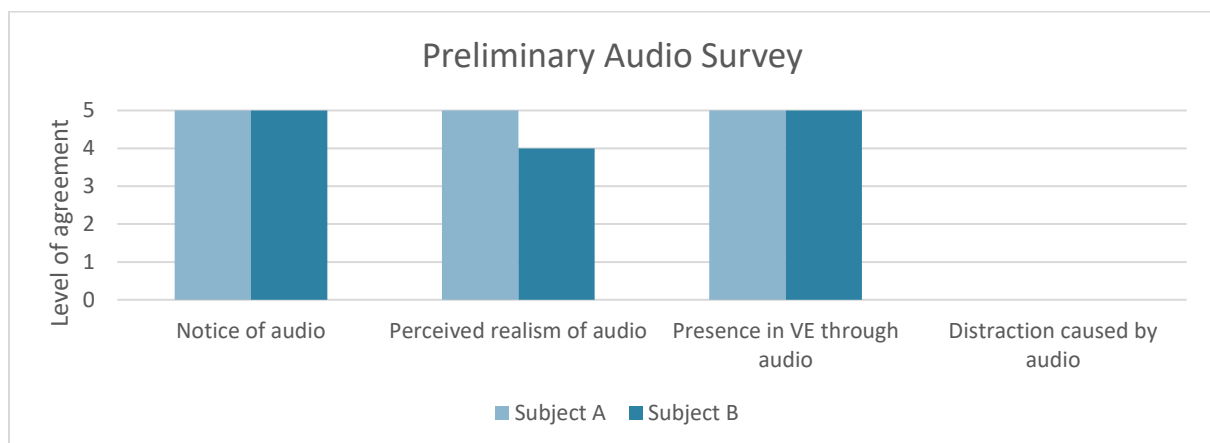
8.1.1. Preliminary results

Figure 16. Results from the preliminary research method with two Strukton employees (Luca Frösler, 2019)

Both subjects indicated that they perceived the audio as realistic and that the audio made them feel present in the VE. This assured me that I have been on the right track when it comes to audio, but I knew there was still room for improvement and increasing the potential of the simulation by utilizing spatialized audio.

8.2. Final research method

In order to assess the spatial audio in comparison to usual audio, a method has been developed in which two versions of the VR simulation are tested by the subject. Both versions provide the exact same visual and interactive experience, but the audio differs between them. In the first version (A) HRTF audio is disabled, and all audio is put to mono. There is still distance attenuation of sounds, but there is no panning of sound sources and therefore it is not possible to localize the origins of sound sources. Sound occlusion is completely disabled, therefore the sounds of cars going under the bridge could be perceived as loud or unrealistic. Moreover, a crude version of reverb is applied to the scene. I have created this crude audio version of the simulation after creating the spatialized version, which has been developed throughout the project. In the second version (B) HRTF-audio is enabled. The listener is able to localize sounds. Additionally, sound occlusion is enabled. Cars that are going under the bridge are muffled and appear quieter. Furthermore HRTF-based reverberation is enabled in the scene. Environmental sounds such as city ambience sounds, crickets and wind, are played as 2D sound sources in both versions of the simulation. In order to keep the conditions of every user test the same, a protocol has been created. The subject is testing both versions of the simulation for approximately 2-5 minutes. The subject is asked to focus on the audio and provided suggestions for what to do in the simulation, such as moving around the bridge, picking up and dropping the wrench and the hardhat, and paying attention to the sound of the KROL. You can find the complete protocol in the annex (chapter 13.4). After that, the subject is asked to fill out a questionnaire, which I have created based on the theory about spatial audio evaluation principles. In the questionnaire I put emphasis on the attribute of presence since, according to my research, this is the highest indicator for user immersion. Some attributes of the evaluation principles concerning room width, room size, room level and room envelopment have been merged into one attribute, since the VR simulation takes place outside, and reverberation therefore does not play that big of a role. Additionally, I did not incorporate the attribute of low frequency levels, because the HTC Vive Pro headphones do not provide most accurate levels of bass and sub bass frequencies (except when deliberately pushing the headphones harder onto the ears). After doing a pilot user test with five subjects, I made several improvements to the survey based on feedback from the subjects. I articulated the questions more clearly, so that they are easier to understand. In addition to that, I added a last question in which the subject is asked which version he/she thinks the

mono version is and which version the spatialized version is. You can find the final survey as part of the annex (chapter 13.6).

8.3. Research results

In the final research method, the protocol (chapter 13.4) has been followed and the survey (chapter 13.6) has been filled out by two test subjects from Strukton Rail.
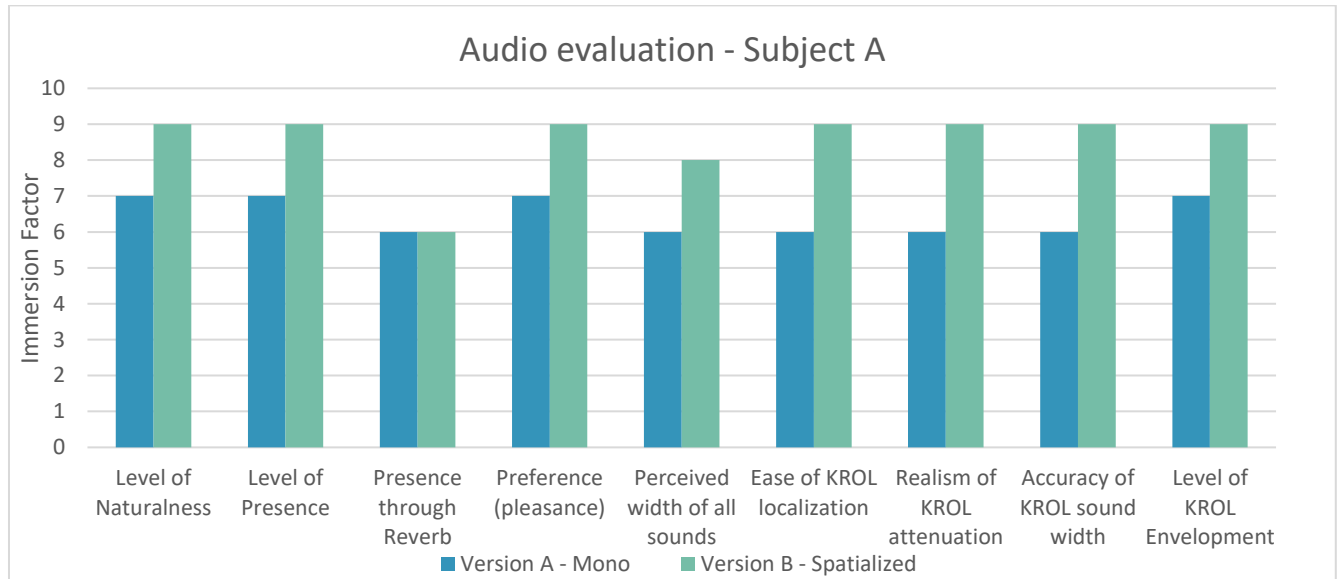

Figure 17. Audio evaluation results based on answers by subject A from Strukton (Luca Frösler, 2019)
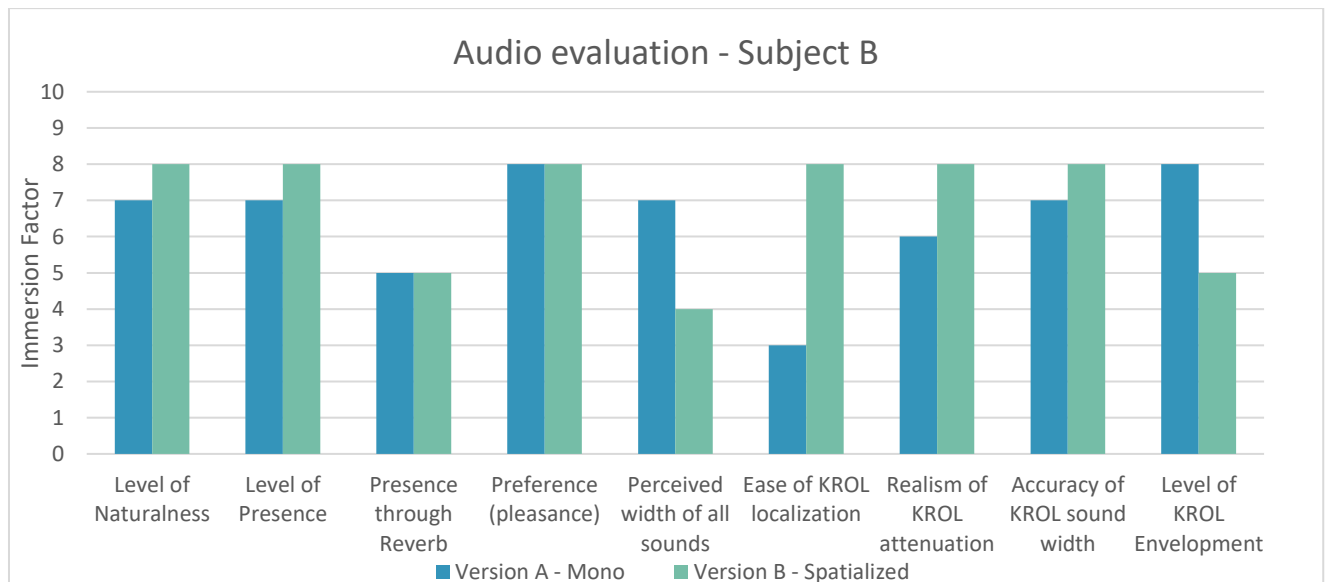

Figure 18. Audio Evaluation Results based on answers by subject B from Strukton (Luca Frösler, 2019)

Additionally, the subjects were asked to indicate whether they noticed environmental ambience audio as well as feedback audio. Both subjects noticed the presence of

environmental ambience audio and indicated that it increased their presence in the scene (Figure 19). Subject A noticed that feedback audio of interactables was there and also indicated that it increased his presence in the scene.
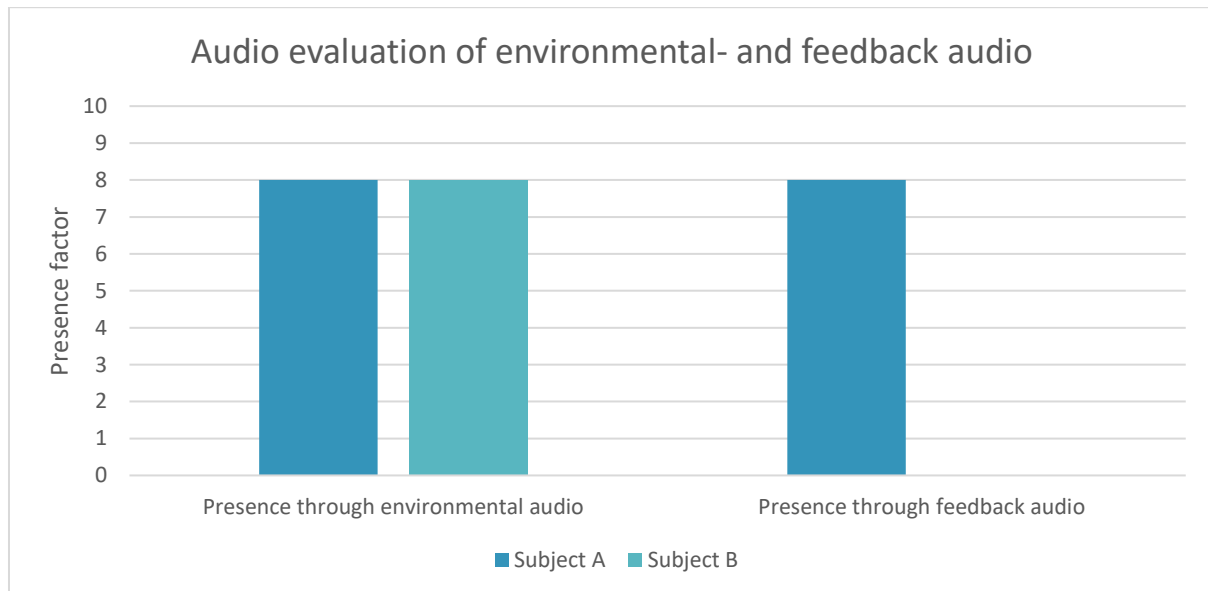

Figure 19. Factor of increased presence through environmental- and feedback audio

8.4. Result analysis

It is significant that both subjects indicated a higher level of audio naturalness, perceived realism and feeling of presence in the spatialized version of the simulation (see Figure 17 and Figure 18). It is also evident that in the spatialized version, localization of a sound source is way easier. Concerning reverberation, both subjects indicated that it has a relatively low impact on the presence in the VE, and that is there is no perceived difference between the two simulations. This may be due to the fact that the reverberation in both versions is barely audible, since it is not a closed room. There is also no comparison to a version without reverberation. When it comes to the width of all sounds as a whole, and the feeling of being surrounded by the KROL, the feedback differs between both subjects. This shows, that the perceived width of all sounds is rather subjective and may be affected by placebo. Technically, the mono version of the simulation would have a bigger perceived width and surrounding, since there is no panning, and the sound of the KROL engine is taking up 360° degrees.

9. Conclusion and discussion

9.1. "To what extent can immersion in a VR training simulation be increased through the use of spatial audio?"

In conclusion, the research has shown, that spatial audio can be designed in a way that the user feels more present in the virtual environment, leading to a higher level of immersion. Moreover, the research proves, that spatial audio is not only preferred over non-spatialized audio but achieves an overall higher level of naturalness and realism. When it comes to individual objects, averaged results indicate, that it is up to 108.34% easier to localize sound sources when using spatialized audio (see Figure 17 and Figure 18). In addition to that, more realistic attenuation can be achieved when using spatial audio, leading to higher immersion of the user. When it comes to sound source width and envelopment, it depends on what outcome the developer wants to achieve. If the sound source should be perceived as wider and rather ubiquitous, the level of spatialization should be decreased. This also counts for sound source envelopment. For specific sound sources, which should be able to be localized easily, spatial audio should be used. Additionally, the final findings indicate, that environmental audio and feedback audio from interactable objects lead to a higher level of user presence.

9.2. "What other auditory factors contribute to immersion in VR?"

I can conclude that besides spatial audio, the biggest factor that contributes to immersion in VR is the presence of environmental audio. Subjects indicated, that on average, their feeling of presence has greatly increased when exposed to environmental audio. Although not recognized by every subject, feedback audio is another possible element to increase immersion.

9.3. "What toolsets are suitable for the design process of spatial audio in VR training simulations?"

Google Resonance or Steam Audio are suitable toolsets for implementing spatial audio in Unity. These toolsets have the advantage of spatializing the audio for the user, without the need of recording or gathering ambisonic sound samples. I would not recommend the Oculus

Spatializer as a suitable toolset, because it lacks sound occlusion. Therefore, a high level of immersion may not be reached.

9.4. "How can immersion through spatial audio be tested?"

One can assess the level of immersion, by collecting data on the subjective perception of the participants, who indicate their feeling of presence comparatively between a spatialized- and non-spatialized version of the VR simulation. In order to test the quality of spatial audio one can refer to use the spatial audio evaluation principles put forward by Berg & Rumsey (Figure 3). The attribute of user presence has the most impact on the level of immersion.

9.5. "Are there possible ethical issues involved in the design process?"

As mentioned in the ethics chapter (7.6), in order to avoid any copyright infringements, sound samples that are used in the design process, should be copyright free at all costs. On another note, if the product would be of commercial nature, the use of FMOD should be given a second thought, since it is expensive to license FMOD for commercial products.

10.   Recommendations

Concerning developing VR simulations that have a serious game character, I would recommend making use of spatial audio, as well as building an environment soundscape and adding feedback sounds, to enhance user presence. In order to explore the full potential of spatial audio, more experimentation must be made with the mentioned SDK's.

In order to evaluate the effect spatial audio has on the presence of the user, the research method must be improved by further developing attributes that relate to user presence. On the other hand, the attribute of room reverberation may only be of value if the scenery of a VR simulation is set inside.

To Strukton I would definitely recommend making use of spatial audio in their VR simulation. A higher level of immersion will be achieved when doing so, which would lead to more effective training for their employees. Keeping in mind that Strukton will not distribute their product commercially, I would recommend using FMOD as a game audio tool, because it offers flexibility and enables more detailed audio handling.

11. Glossary

Term	Abbreviation	Definition
Ambisonics		Spatialized, binaural, full-sphere surround audio.
Binaural audio		Spatialized audio played through headphones.
Digital Audio Workstation	DAW	Audio editing program for recording, sequencing and composing music.
Decibel	dB	Unit of measurement used for, inter alia, levels of sound volume.
Head-mounted display	HMD	Helmet-like display device worn on the head with either one or two screens inside.
Head-related transfer function	HRTF	A function characterizing sound modifications from where sound is being emitted until hitting the eardrum.
"Kraan op lorries" (crane on tracks)	KROL	Crane that is used by construction workers on rail tracks to perform work at heights.
Software Development Kit	SDK	A collection of software and/or plugins that give developers head starts for developing.
Spatial audio		Audio that is played through at least two audio channels, giving a sense of space.
Transaural audio		Spatial audio played through loudspeakers.
User Interface	UI	The graphical port with which the user interacts with the computer.
Virtual environment	VE	The digital environment that encapsulates the user in a virtual reality experience.
Virtual reality	VR	Computer-generated simulation of a 3D interactive environment using electronic equipment such as a helmet with a screen inside.

12. Bibliography

Berg, J., & Rumsey, F. (n.d.). Systematic Evaluation Of Perceived Spatial Quality. *AES 24th International Conference on Multichannel Audio.*

Bordwell, D., & Thompson, K. (2001). *Film Art. Sixth Edition. An Introduction.* New York: McGraw-Hill.

Conn, C., Lanier, J., Minsky, M., Fisher, S., & Druin, A. (1989). Virtual environments and interactivity: windows to the future. *ACM SIGGRAPH Computer Graphics, 23*(5), 7-18. doi:10.1145/77276.77278

Dolby Atmos Cinema Sound. (n.d.). Retrieved from Dolby: https://www.dolby.com/us/en/technologies/cinema/dolby-atmos.html

Gould, R. (2018, March 29). *Let's Test: 3D Audio Spatialization Plugins*. Retrieved from Designing Sound: http://designingsound.org/2018/03/29/lets-test-3d-audio-spatialization-plugins/

Guastavino, C., Larcher, V., Catusseau, G., & Boussard, P. (2007). Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo. *Proceedings of the 13th International Conference on Auditory Display*, (pp. 53-59). Montreal.

Haas, E., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computing & Control Engineering Journal, 7*(4), 193-198. doi:10.1049/cce:19960407

HRTF. (2016). Retrieved from Articulated Sounds: https://articulatedsounds.com/hrtf

Ibánez, M. L., Alvarez, N., & Peinado, F. (n.d.). A Study On An Efficient Spatialisation Technique For Near-Field Sound in Video Games.

Järvinen, A. (2002). Gran Stylissimo: The Audiovisual Elements and Styles in Computer and Video Games. *Proceedings of Computer Games and Digital Cultures Conference* (pp. 113-128). Tampere: Tampere University Press.

Kaiser, F. (2011). *Transaural Audio - The reproduction of binaural signals over loudspeakers.*

Mestre R., D. (n.d.). Immersion and Presence.

Naef, M., Staadt, O., & Gross, M. (2002). Spatialized Audio Rendering for Immersive Virtual Environments. *VRST '02 Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 65-72). Hong Kong: ACM New York, NY, USA. doi:10.1145/585740.585752

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics, 31*(9), 1005-1017. doi:10.1080/01446193.2013.828844

Shankleman, M. (2008). *Celebrating a stereo pioneer: Alan Blumlein*. Retrieved from BBC News:

https://web.archive.org/web/20080826100857/http:/news.bbc.co.uk/2/hi/technolo gy/7538152.stm

Spence, C. (2007). Audiovisual Integration. *Acoust. Sci. & Tech, 28*(2), 61-70. doi:10.1250/ast.28.61

Strukton Rail. (n.d.). *About Us*. Retrieved from Strukton: https://strukton.com/en/over-ons

Table of Sound Pressure Levels. (n.d.). Retrieved from Sengpielaudio: http://www.sengpielaudio.com/TableOfSoundPressureLevels.htm

Transaural Rendering. (2006). Retrieved from University of Maryland Institute of Advanced Computer Sciences: http://users.umiacs.umd.edu/~ramani/cmsc828d_audio/828d_l19.pdf

Valve Corporation. (2017). *Steam Audio FMOD Studio Plugin 2.0-beta.17.* Retrieved from Valve Software.

W, M., & Normandin, S. (2014). *History and Types of Loudspeakers.* Retrieved from Edison Tech Center: http://edisontechcenter.org/speakers.html

13. Annexes

13.1. Reflection on competences

Technological competences
1. Technical research and analysis *The starting professional has a thorough knowledge of the current digital technologies within that part of the field of work the training course aims at. The starting professional is capable of conducting technical research and analysis.*
In the theoretical framework I have conducted detailed technical research and analysed state-of-the art technology. I have done extensive desk research on existing literature about spatial audio, but also researched what auditory means in general contribute to an immersive experience that comes close to real-life. Moreover, my expertise in the field of Unity development and audio have accumulated in the last years through private interest, work during my internship and study projects. With the use of spatial audio, I have tried to incorporate value into the product that is innovative to the client. This is because Strukton relies on our expertise to create the VR simulation as immersive and realistic as possible.
2. Designing, prototyping and realizing *The starting professional is capable of creating value by iteratively designing and prototyping, based on a (new) technology, creative idea or demand articulation. The starting professional shows an innovating, creative attitude at defining, designing and elaborating a commission in the margin of what is technically and creatively feasible.*
From the problem that was given by the client, I have derived the underlying problem as part of my thesis. From this problem I have created value for the product by improving the product iteratively. Throughout the whole project I have been passionate about designing audio solutions that increase the value of the product by leveraging the immersion. As an example, I have created the train audio system prototype (chapter 13.7) in which I have developed a method to recreate real-life audio perception of a train in Unity. In addition to that I made several iterations of the audio, starting with standard audio in the beginning, moving to spatialized audio using Google Resonance and eventually using Steam Audio to develop spatial audio in greater detail. In addition to that I have developed the "Interactable Audio Handler"; a generic script which can be placed on an interactable object and plays respective sounds when the object is picked up and dropped onto different materials.
3. Testing and rolling out *The starting professional is capable of repeatedly testing the technical results, that come into being during the various stages of the designing process, on their value in behaviour and perception. The starting professional delivers the prototype/product/service within the*

framework of the design, taking the user, the client and the technical context in due consideration.
Throughout the project, at the end of every sprint a playable demo has been created which exhibited the current state of the simulation. The demo was presented to the client in form of a video or sometimes the client visited to try it out himself. Based on feedback, the simulation got iterated further. During the development all parties concerned were taken into account; the client's requirements but also the user (Strukton employee) who will use the simulation in the future. The development phases were planned to create a final working prototype for the end of the project.

Designing competences
4. Investigating and analysing *The starting professional is capable of substantiating a design commission by means of research and analysis. The starting professional, in his/her investigation activities, shows to have a repertoire of relevant research skills at his/her disposal and is able to select from this repertoire the proper method, given the research circumstances. Is capable of developing prototypes as a communication tool within the context of implementation.*
To test my implemented design, I have developed user tests that are based on my research. I have created two prototypes of the simulation to conduct the comparative research method. Based on the results I have drawn conclusions which I have integrated into my recommendations.
5. Conceptualizing *The starting professional proves capable of being able to get to realistic (cross-sectoral) demand articulation and project definition. The starting professional is capable of developing an innovative concept that creates value on the basis of his/her own idea or demand articulation.*
In my opinion I have fully satisfied this competence. Initially, audio was not one of the client's requirements. Based on my experience in VR/AR development that I have gathered in past projects such as the Thales VR project or the OrchestraVR project, as well as personal passion about sound, I know the importance of audio in VR simulations. Therefore, I have conceptualized and implemented ideas for the audio from the beginning till the end of the project. Through my research method and acquired feedback I prove that the conceptualized ideas add value to the product.
6. Designing *The starting professional is capable of shaping concepts and elaborate these in a substantive, graphic and/or aural way.*

Throughout the project I have refined and elaborated the concept of spatial audio. I have applied new knowledge that I have acquired through my research, such as spatialization with Steam Audio, which I have never worked with before, and applied it in the design process. In general, we as a project group have initially had a brainstorm session (chapter 13.9) in which we came up with ideas and concepts for the product. These concepts were iterated adjusted over time. In addition to that together with the project group we created a MoSCoW (chapter 13.10) for prioritization purposes.

Organisational competences

7. Enterprising attitude

The starting professional sees opportunities and possibilities and knows how to translate them from a market-oriented point of view into (new) concepts, products, services, in order to thus get to creating value and new revenue models.

8. Enterprising attitude and skills

The starting professional has enterprising skills in order to be able to function both as an employee and independently. The starting professional is capable of converting commercial skills into innovative products, services or collections; bearing commercial feasibility in mind.

Although the client did not intend the product as a commercial one, requirements and different use cases were put forward. These use cases, such as using the VR simulation as a demo for other office workers, a showcase demo for the higher people at Strukton, as well as using the VR simulation for training purposes were strongly taken into account. Although not proposed by the client, I have added value to the product by utilizing spatial audio in the VR simulation. If the product would be commercial, this would add a lot of value to the product, since it makes the VR simulation more realistic and immersive, improving the training simulation and making it more impressive. Finally, in my thesis, I recommend to the client to definitely use spatial audio for further development of the VR simulation, because it increases the value to a great extent. Concerning commercial feasibility, I made the remark that if the product would be distributed commercially, it would be fine except FMOD should not be used, since it requires expensive licensing.

9. Working in a project-based way

The starting professional shows him/herself capable of being able to accept, set up and carry out projects from an engagement with stakeholders, whether or not in cooperation with others as a team. The starting professional shows that he/she is capable of cooperating with others in a (multidisciplinary) team in a productive way, reaching

a good balance between introducing his/her own expertise and relying on the complementary expertise of others. The starting professional shows him/herself capable of directing team members.

The created product was conducted within a multidisciplinary project group. The requirements and demands by the stakeholder were taken into account and incorporated in the project work. Moreover, there was a good balance between the different skillsets of the project group members. In the beginning we set clear roles for everybody in the project group. I made sure that everybody is content with their role, since I have been motivated to create an awesome product. I was the only one working on audio and I was confident that my skills in it are more than sufficient. That also meant that I let other students of their group have their focus in programming. Due to the fact that I was the leader of the project, I managed team members and made sure that we stay on track when it comes to the group work. When there were tensions in the group I communicated with the students in order to resolve them. In addition to that I communicated with the client and set up regular meetings, in which the client was able to test the product and give feedback. We also regularly updated the list of the requirements together with the client, to ensure that all parties involved are on the same page when it comes to the design process.

10. Communication

The starting professional shows him/herself capable of presenting both his/her person and his/her work professionally and well-groomed to third parties. The starting professional shows him/herself capable of being able to communicate with a client about choices and progress in the design process.

Since I am the project leader, throughout the project I have been responsible for all communication with the client, as well as communication with the project coach. Communication with the client included progress updates, discussions about the design process of the product as well as planning of client meetings with the project group. I am certain that I have always communicated in a professional manner. Additionally, I am able to communicate myself and my skills professionally to the world. I have created my own portfolio website, which showcases my work and updated my LinkedIn profile. As soon as this project is finished, I will update it on my portfolio website.

| Professional competences |
|---|
| **11. Learning ability and reflectivity**

 The starting professional shows him/herself to be a 'reflective practitioner' by constantly analysing and adjusting his/her own action fostered by feedback of others. The starting professional shows him/herself permanently directed and capable of being able to keep up with relevant developments in the field of expertise. The starting professional is able to further develop and deepen the craftsmanship, the personal substantiation of the professional situation and his/her creativity. |
| Throughout the project I have taken feedback of several parties into account. For one the feedback of the client, when we had play tests, with employees from Strukton, also the feedback of my graduation coach Yiwei, which was very helpful and helped me to adjust and iterate my graduation report. I also got feedback from the students in my project group on aspects of the audio, and improved things based on the feedback. I learned a lot about using spatial audio in FMOD and Unity through the research and practical work in the project, which made me develop my skills in audio further. |
| **12. Responsibility**

 The starting professional has a capacity for empathy with other sectors and shows awareness of ethical issues in his/her role as a designer and is able to explicitly make such considerations in accounting for choices in the design process. |
| I fulfil this competence since I have been seeing the bigger picture in this project from the start. I deliberately avoided to infringe copyrights when it comes to using sound samples, in order for Strukton to have all rights to the content. Furthermore, I was aware of the expensive license that is required to use FMOD commercially, but since the project is only going to be used internally and for demo showcases, this does not apply. |

13.2.   Proof video

Link: https://youtu.be/UgUEcimkOAE
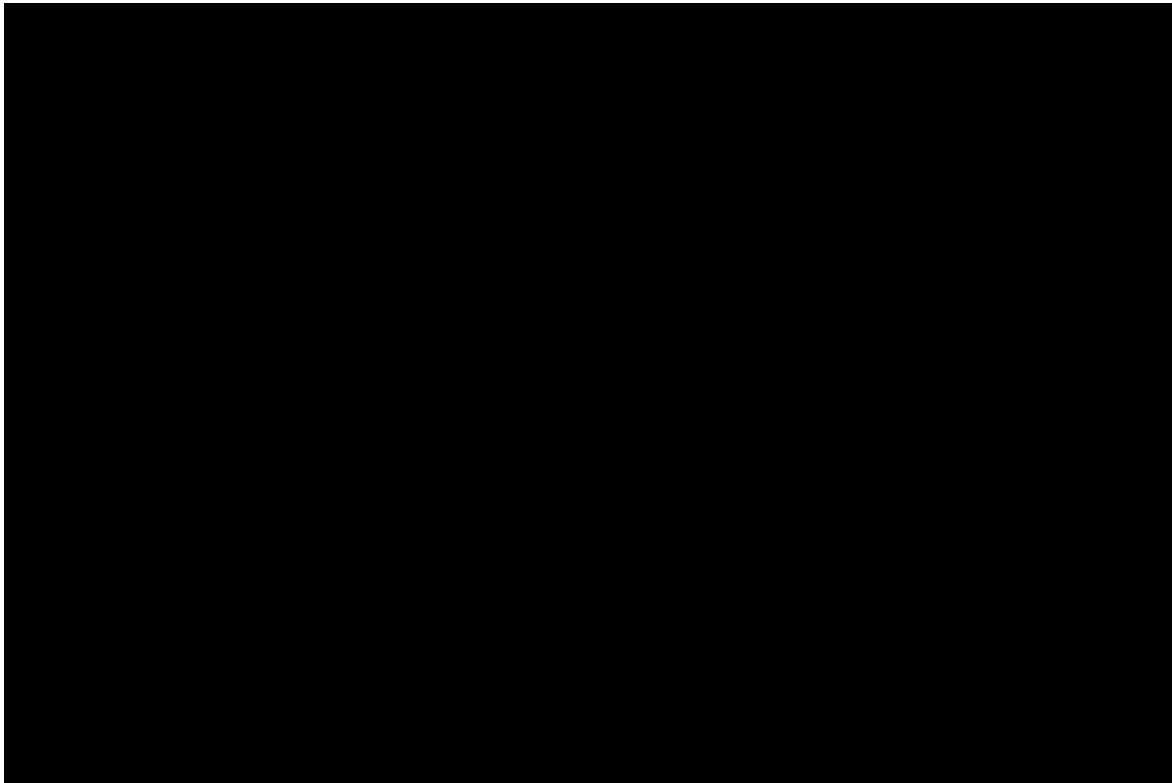
In the proof video you can see what we created for the VR simulation. This video is not the final product video but serves as a proof for the graduation. In all parts of the video you can hear audio that has been created by me.

13.3.   FMOD Projects

You can find the current state of the FMOD projects for the product here:

https://drive.google.com/file/d/1a5TxVRSBqYV6NIoQxfpauPfQ60knCx-q/view?usp=sharing

13.4.   Research method protocol

"STRUKTON VR SAFETY TRAINER – AUDIO EVALUATION PROTOCOL

I will first explain the basic controls of the HTC Vive, so you can interact with the virtual environment.

Then, you are going to test the first version (A) of the VR simulation for approximately 2-5 minutes.

After that you are going to test the second version (B) of the VR simulation for another 2-5 minutes.

While being in the simulation I suggest you do the following things:

- Pay close attention to the audio
- Move around the bridge (teleport) and around the KROL
- Pick up and drop the hardhat and/or the wrench

The two versions of the VR simulation differ in the surround quality of the audio. After testing both versions, I will ask you to fill out a survey and give feedback on different attributes of the audio. The survey will take approximately 5-10 minutes in total and all collected data will be handled anonymously, only being accessible to the researcher and only used for this research. You can quit the survey at any time."

13.5.   Preliminary Survey

These are the preliminary user tests which I created. I asked two Strukton employees for their feedback.

Audio

EN: I noticed that there was audio during the VR simulation.
NL: Ik merkte dat er audio aanwezig was tijdens de simulatie.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
				X

EN: I perceived the audio of the VR simulation as realistic (close to real-life).
NL: De audio in de simulatie klonk realistisch (bijna levensecht).

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
			X	

EN: The audio helped me to feel present in the virtual environment.
NL: Dankzij de audio voelde ik me aanwezig in de virtuele omgeving.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
				X

EN: The audio of the VR simulation distracted me from performing the given tasks.
NL: De audio in de simulatie heeft mij afgeleid van het uitvoeren van de gegeven taken.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
X				

Figure 20. Audio survey answered by subject A

Audio

EN: I noticed that there was audio during the VR simulation.
NL: Ik merkte dat er audio aanwezig was tijdens de simulatie.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
				X

EN: I perceived the audio of the VR simulation as realistic (close to real-life).
NL: De audio in de simulatie klonk realistisch (bijna levensecht).

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
				X

EN: The audio helped me to feel present in the virtual environment.
NL: Dankzij de audio voelde ik me aanwezig in de virtuele omgeving.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
				X

EN: The audio of the VR simulation distracted me from performing the given tasks.
NL: De audio in de simulatie heeft mij afgeleid van het uitvoeren van de gegeven taken.

Volledig mee oneens	Mee Oneens	Neutraal / Ik weet het niet	Mee Eens	Volledig mee Eens
X				

Figure 21. Audio survey answered by subject B

13.6.   Final survey

Strukton VR - Audio Evaluation Survey

In this survey you will provide feedback on audio of the two versions of the VR simulation, which you just tested.

Every question needs to be answered for each version of the VR simulation.

Version A - meaning the first version of the simulation that you tested.

Version B - meaning the second version of the simulation that you tested.

If you have remarks to any of the questions or the simulation itself, there is space at the end to give feedback.

Thank you!

*Required

1. **Your Name** *

Naturalness
Please indicate how natural you perceived the audio as a whole.

2. **Version A: How natural (close to real-life) did you perceive the audio as a whole?** *
   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unnatural	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Natural

3. **Version B: How natural (close to real-life) did you perceive the audio as a whole?** *
   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unnatural	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Natural

Presence
Presence is the feeling of "really being there" in the virtual environment.
Please indicate to what extent the audio made you feel present in the virtual environment.

4. **Version A: How present did you feel in the virtual environment, as a result of the audio?** *
   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Absent	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Present

5. **Version B: How present did you feel in the virtual environment, as a result of the audio?** *
   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Absent	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Present

6. **Version A & B: Did you notice any environmental ambience audio (e.g. city noise, cars, crickets etc.) in the simulation?** *
   Mark only one oval.

   ◯ No   *Skip to question 8.*
   ◯ Yes   *Skip to question 7.*

Presence: Environmental Ambience
Presence is the feeling of "really being there" in the virtual environment.
Please indicate to what extend the environmental audio contributed to your feeling of presence.

7. **In how far did the environmental audio affect your presence in the virtual environment?** *
   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
No difference	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Greatly increased presence

Presence
Please state if you noticed audio getting triggered while interacting with objects.

8. **Version A & B: Did you notice any feedback audio while interacting with objects (picking up/dropping objects) ?** *
   Mark only one oval.

   ◯ Yes   *Skip to question 9.*
   ◯ No   *Skip to question 10.*

Presence: Feedback Audio
Presence is the feeling of "really being there" in the virtual environment.
Please indicate how audio of interactable objects affected your presence.

9. **Version A & B: In how far did feedback audio emitted by interactable objects affect your presence in the virtual environment?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
No difference	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Greatly increased presence

Presence: Room Reflections

Please indicate in how far room reflections (reverb) made you feel more present in the virtual environment.

10. **Version A: To what extend did the room reflections (reverb) make you feel more present in the virtual environment?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
No difference	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Greatly increased presence

11. **Version B: To what extend did the room reflections (reverb) make you feel more present in the virtual environment?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
No difference	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Greatly increased presence

Preference

Please indicate to what extent you like the perceived audio as a whole.

12. **Version A: To what extent did you find all sounds as a whole pleasant?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unpleasant	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Pleasant

13. **Version B: To what extent did you find all sounds as a whole pleasant?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unpleasant	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Pleasant

Ensemble width

All sounds together occupy a maximum width of 360° around your head.
Please indicate how wide you perceived all sounds together (the occupied angle of all sounds)

14. **Version A: How wide (the occupying angle) did you perceive all sounds as a whole?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Narrow	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Wide

15. **Version B: How wide (the occupying angle) did you perceive all sounds as a whole?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Narrow	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Wide

Localization

Please indicate how easy it was for you localize the KROL in the virtual environment based on the audio.
The KROL is the mobile working crane that was next to you during the simulation.

16. **Version A: How easy was it for you to localize the KROL, based on the sound it emitted?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Difficult	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Easy

17. **Version B: How easy was it for you to localize the KROL, based on the sound it emitted?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Difficult	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Easy

Source Distance

Please indicate to what extent the change in loudness of the KROL engine felt realistic.
The KROL is the mobile working crane that was next to you during the simulation.

18. **Version A: When moving closer or further away from the KROL, the loudness of its engine changed. How realistic (close to real-life) did you perceive the change in loudness?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unrealistic	○	○	○	○	○	○	○	○	○	○	○	Realistic

19. **Version B: When moving closer or further away from the KROL, the loudness of its engine changed. How realistic (close to real-life) did you perceive the change in loudness?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Unrealistic	○	○	○	○	○	○	○	○	○	○	○	Realistic

Source Width
In surround sound, a sound source can be placed anywhere within 360° around your head.
Each sound has a certain width, meaning the angle it occupies.
Please indicate how accurate you perceived the width of the KROL engine.

20. **Version A: How would you rate the accuracy of the perceived sound width of the KROL engine (occupying the correct angle of 360°)?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Inaccurate	○	○	○	○	○	○	○	○	○	○	○	Accurate

21. **Version B: How would you rate the accuracy of the perceived sound width of the KROL engine (occupying the correct angle of 360°)?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Inaccurate	○	○	○	○	○	○	○	○	○	○	○	Accurate

Source Envelopment
Please indicate how the sound of the KROL engine, enveloped/surrounded you.
The extent to which the sound source existed around you.

22. **Version A: Too what extent did you feel like being surrounded by the sound of the KROL engine?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Excluded	○	○	○	○	○	○	○	○	○	○	○	Surrounded

23. **Version B: Too what extent did you feel like being surrounded by the sound of the KROL engine?** *
    Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Excluded	○	○	○	○	○	○	○	○	○	○	○	Surrounded

Type of audio
Please indicate what type of audio you think you perceived in both versions.

24. **Version A: What type of audio do you think you listened to?** *
    Mark only one oval.
    - ○ Mono
    - ○ Spatialized 3D Audio

25. **Version B: What type of audio do you think you listened to?** *
    Mark only one oval.
    - ○ Mono
    - ○ Spatialized 3D Audio

Remarks

26. **If you have any remarks, criticism or questions about the simulation or this survey, please put them here. Thank you!**

    _____

    _____

    _____

    _____

    _____

Figure 22. Final research method survey

13.7.   Train audio system prototype

In the early stages of the development one requirement of the project was that trains move along train tracks. I made the train move but sound for it was still required. I found a fitting sample of a train engine which was very characteristic with all its screeches and little sounds that you are used to from hearing a train. I was then wondering how to set up the audio source for the train. In real life a driving train emits sounds from hundreds of different areas, especially the wheels which cause friction on the rail track. Nevertheless, in Unity each audio source can only emit sound from one single point. To replicate it lots of audio sources would be needed. To solve this problem, I figured to use one audio source, which moves depending on the position of the player in relation to the train.
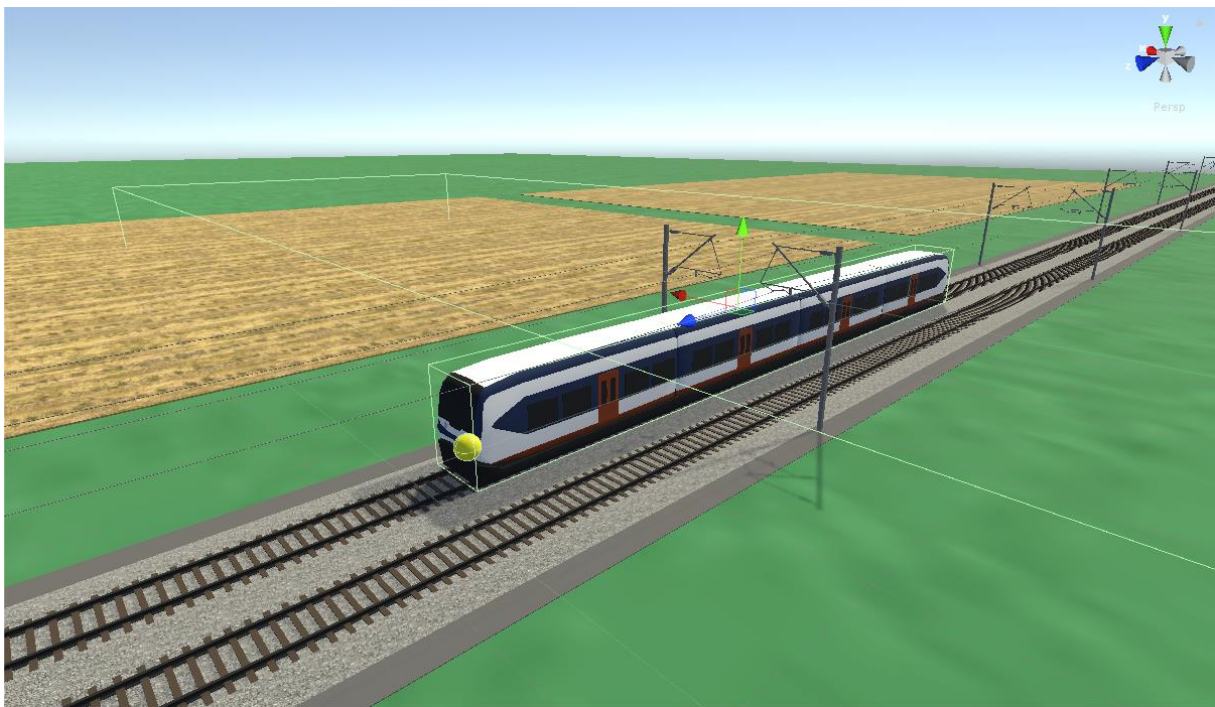

Figure 23. Train with audio source (indicated by yellow sphere)

As seen in Figure 23, the audio source, which is a child of the train game object, is visualized by a yellow sphere. Also notice the huge green collider box around the train. As soon as the player enters this collider (which occurs when the train moves past the player), the audio source moves with the player along the z-axis. The audio source will stay next to the player until the player leaves the big collider again, then the audio source will stay at the end of the train. Through this the perceived audio of the train will behave closely to that of a real one and the audio source will "stick" next to the player.

13.8.   Sampling Workflow

As soon as I had a sound sample that I wanted to implement in FMOD, I first imported it into my preferred DAW, namely FL Studio to cut and process it to my needs. I usually apply an equalizer, to cut away or even out the frequency spectrum of the sample. If needed additional adjustments of the stereo image, volume, and/or compression are made. No reverb or delay has been applied and I kept samples dry, since they would be used in the VR simulation, and reverb would be applied through the game engine at a later stage.
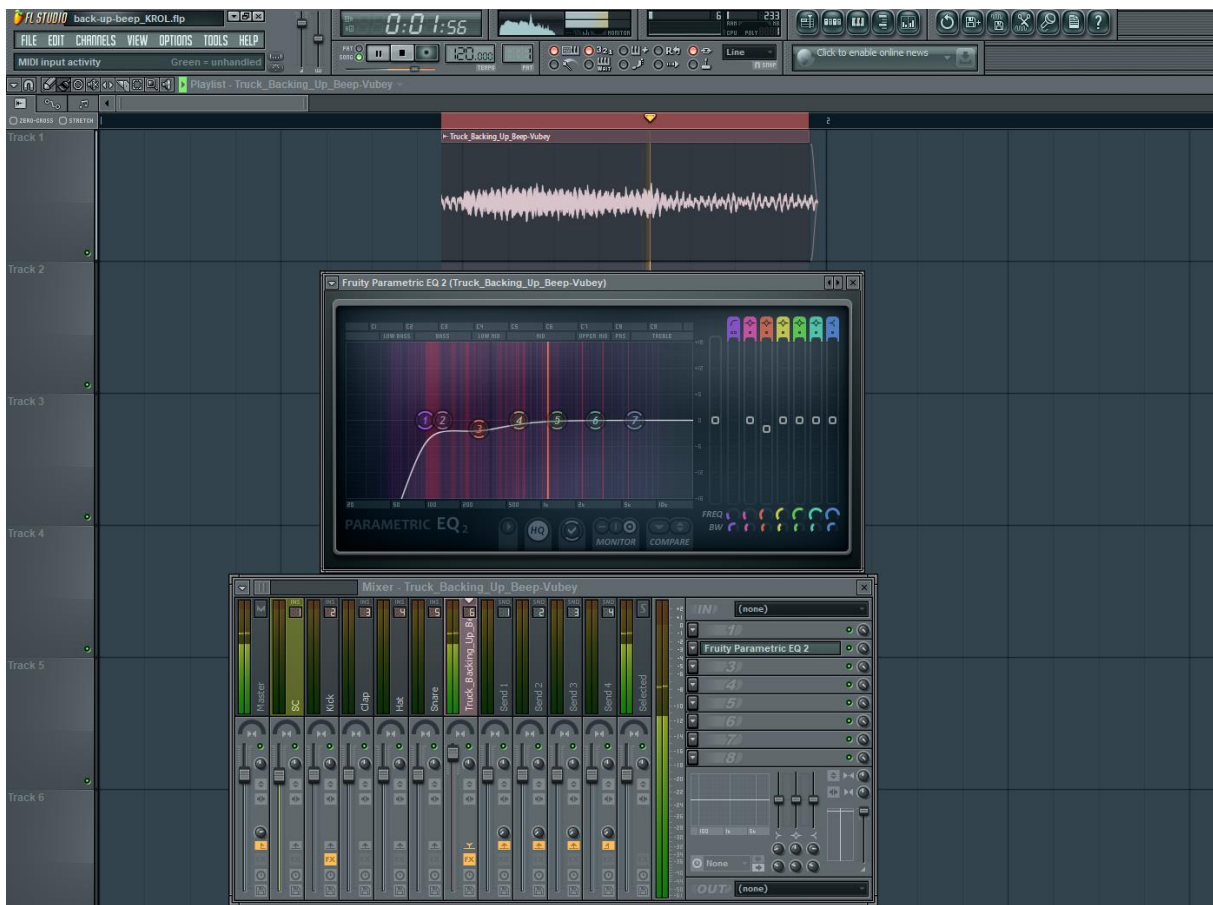
Figure 24. FL Studio with looping sample on top, EQ and mixer at bottom
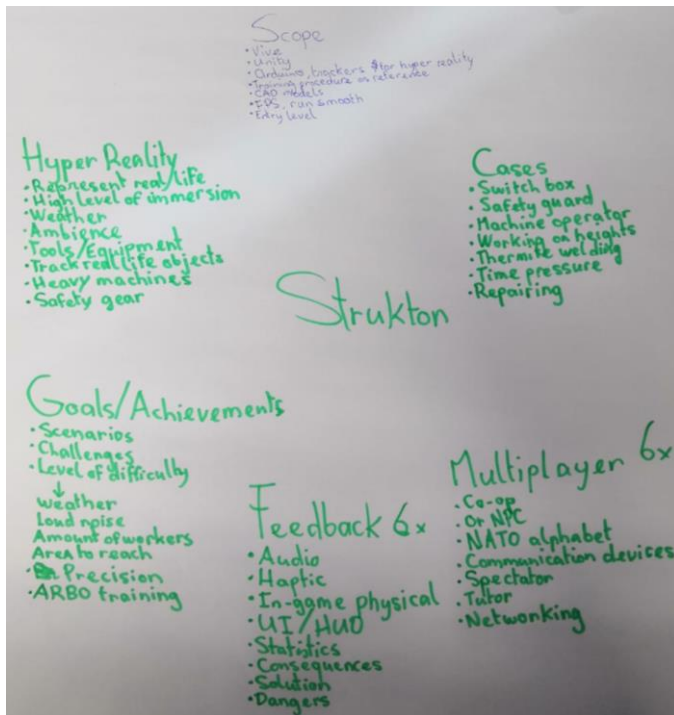
13.9.   Initial Brainstorming

Figure 25. Assembled ideas from a brainwriting session with other students

13.10.   MoSCoW

Must:
- Task mechanics
- KROL mechanics *
- Weather System *
  - Thickness
  - Speed
  - Fog
  - 2nd iteration of random
- Visual Feedback System
  - Outlines *
  - Interaction hints
- Evaluation System
  - Teach mode
  - Practice mode
  - Test mode
- Hand snapping system *
- Environment Assets
  - Convert greybox
- PPEs
  - Wearing functionality *
  - Assets
- Tool assets *
- Tool Belt mechanics
- Carabiner hooking mechanics *

Should:
- Mobile office scene
  - Interactive menu
  - Tutorial
- Movement system
  - Overhaul
- Hands
  - Overhaul
- Decorative environment assets
  - Street lamps, benches, etc..
- Weather System
  - 1st iteration of Shader
    - Puddles
    - Snow sticking
    - Snow footprints

Could:
- Hyper reality elements
- Moving environment assets
  - Cars, trains, birds, etc…
- Clean code for better performance
- Walkie Talkie
  - Hint and guidance system
  - Asset
  - Voice over from Strukton

This is an updated version of the MoSCoW that shows prioritized tasks and epics that were used for the development process.

13.11.   Client meetings

After the project got pitched to us initially, we have arranged several meetings at Saxion with them over the course of the project. We visited the Strukton headquarters in Utrecht, as well as the Strukton Rolling Stock production facility in Enschede. After we got a tour through the production facility and took reference pictures of electrical components, we got a working-in-heights workshop with a safety officer from Strukton. The gained insight into the safety procedures that are required from the workers was useful input for the development of the VR simulation.

Figure 26: Safety workshop with safety officer at Strukton Rolling Stock