Quality Assessment of VR Film

A Study on Spatial Features in VR Concert Experiences

MASTER THESIS

Victor Milesen, Dina Madsen Smed & Rasmus B. Lind

Aalborg University Copenhagen
M.Sc. Medialogy 2017

AALBORG UNIVERSITY
STUDENT REPORT

Medialogy
Aalborg University Copenhagen
http://www.aau.dk

Title:
TBA

Theme:
Master (MED)

Project Period:
Spring Semester 2017

Project Group:

Participant(s):
Lind, Rasmus Bloustrød
Smed, Dina Madsen
Milesen, Victor

Supervisor(s):
Georgios Triantafyllidis

Copies: 1

Page Numbers: 65

Date of Completion:
June 2, 2017

Abstract:

In this study we are exploring how stereoscopic video and ambisonic sound contributes to the perceived quality of experience(QoE), and how these visual and aural features influences presence and motion sickness. Two tests have been conducted in this study, both an initial test and a comprehensive final quantitative test. The initial tests was conducted on 27 participants, and the final quantitative test was conducted on 120 participants. Our study showed no significant differences in the perceived quality or presence. However we found that people sensed significantly more direction in the sound while being presented to monoscopic imagery with ambisonic sound than compared to a stereoscopic imagery with stereo sound. Both of the stereoscopic productions had a higher acceptance score with regards to motion sickness, which is interesting, however this difference was insignificant. This study also reflects our efforts to propose methods to evaluate "Quality of Experience" in 360 video and VR film experiences.

AALBORG UNIVERSITET
STUDENTERRAPPORT

Medialogy
Aalborg Universitet København
http://www.aau.dk

Titel:
TBA

Tema:
Master (MED)

Projektperiode:
Forårssemestret 2017

Projektgruppe:

Deltager(e):
Lind, Rasmus Bloustrød
Smed, Dina Madsen
Milesen, Victor

Vejleder(e):
Georgios Triantafyllidis

Oplagstal: 1

Sidetal: 65

Afleveringsdato:
2. juni 2017

Abstract:

I dette studie undersøger vi hvordan stereoskopisk video og ambisonic lyd bidrager til brugerens opfattelse af oplevelsens kvalitet (QoE), og hvordan visuelle samt lyd features har inflydelse på presence og motion sickness. Vi har gennemført to test scenarier, både en indledende test samt en mere omfattende endelig kvantitativ test. Den indledende test blev gennemført med 27 testpersoner, og den endelige test blev gennemført med 120 testpersoner. Vores resultater viser ingen signifikant forskel i kvalitetsoplevelsen eller presence. Vi fandt dog at vores testpersoner opfattede signifikant mere retning i lyden når de blev præsenteret for monoskopisk video med ambisonic lyd, end stereoskopisk video med stereo lyd. Begge de stereoskopiske produktioner havde en højere acceptance score med henblik på registreret motion sickness, hvilket er interessant, dog ikke signifikant. Dette studie reflekterer desuden vores intention med at foreslå metoder til at evaluere "Quality of Experience"i 360 graders video og VR film oplevelser.

Contents

Preface

The authors of this master thesis wish to thank the studyboard of Aalborg University, Rolf Nordahl, Stefania Serafin and Lars Reng for their supervision and support during our internship in USA 2016, in collaboration with Janus Heiberg Madsen and Panorama Video. This collaboration facilitated our opportunity to capture footage from particularly the venue Antone's which is an acknowledged historic birthplace of blues and Rock'n Roll artists such as Stevie Ray Vaughan, Albert King, BB King and most lately Gary Clark Jr among many other great up and coming artists as Jackie Venson. We also wish to thank the management at Antone's in Austin Texas for trusting us and granting us permission to record and collaborate with their staff and performing artists. We are so thankful that Jackie Venson allowed us to capture the performance of her band which gave us the opportunity to evaluate some of the produced material contained in this master thesis. Furthermore we wish to thank Sven Houlberg for allowing us to capture footage at the venue RUST, also contained in this master thesis. Lastly we wish to thank Signe Ungermand and Maria Engermann for their help and support with regards to building a stereoscopic rig.

This master thesis is by us considered to be the culmination of 3 years focused studies on 360 Video and 3D Audio capture where we have been exploring several aspects of immersive, interactive and cinematic VR-Film technologies with regards to musical experiences in particular. As this is our final master thesis we wish to reflect that we have achieved professional and academic skills related to the full production, evaluation and distribution scope of high quality VR film experiences.

On a final note we wish to thank Georgios Triantafyllidis for the support and collaboration during the final master thesis semester, which allowed us to turn our ambitions, learning goals and academic vision into tangibility.

Chapter 1

Introduction

Since 2015, where we started working with 360 video and VR, it seems like the industry focus has been skewed towards high end VR-solutions for game experiences in particular. Technology providers have been introducing many interesting ideas and technologies which will probably influence the development on many of the future technological standards in VR/AR game and film. While one could say that this development is good and promising, VR and AR have still not had a groundbreaking breakthrough to the broad consumer market segments.

We believe that it is of importance to look at quality assessment of both technologies and applications in the field of VR, in order to promote better experiences and to build a better reputation for VR with regards to motion sickness, image quality and sound quality. "Quality of Experience" is a terminology and academic field introduced and defined by the International Telecommunication Union[24].

In our master thesis we will adapt this framework for quality assessment in our work, as we both wish to propose a method for quality assessment and eventually explore how 3D sound and stereoscopy has an influence on a wider spectrum of quality measures in VR film experiences.

Chapter 2

Motivation

2.1 Previous Studies

In this chapter we will briefly outline the academic work of the authors leading to this final master thesis, which is the culmination of 3 years consecutive focus on 360 video, spatial sound and interaction in VR.

In february 2015 we started exploring 360 video, as we additionally applied for funding a 360 video rig including 14 GoPro cameras. This was a groundbreaking event for us as AAU students as this event created the foundation for many new and interesting approaches to investigate the potentials of VR film. We as a student group decided to focus of concert experiences due to the scope of production and evaluation this would allow us. Our initial approach was to get acquainted with the workflow and it's complexity, while being introduced to presence theory[19] and its methodology. We also felt that concert experiences is a field most suitable for VR due to its many immersive and entertaining attributes. In our academic debut, our bachelor thesis, we conducted a study on how camera positions could possibly influence the feeling of presence in concert experiences, and thereby the title of our bachelor thesis was **"Camera Positions and Presence in VR Concert Experiences"**. This study revealed no significant difference in the feeling of presence experienced by the user when respectively being placed on stage or among the audience[21].

From september 2015 we continued our efforts with regards to VR film experiences and now with increased focus on the aural aspects of interactive VR experiences. Additionally we experiemented with implementation of holographic video which facilitated user movement within a replicated physical location in Unity game environment. By using digital photography we constructed a synthesis of a small concert venue, and in this game environment we implemented a concert footage

Figure 2.1: Illustration of the rig we used to capture 360 degree video footage for our bachelor thesis production in 2015. This photo also shows the binaural dummy we used to capture the sound with [21]

captured with 15 cameras, allowing us to create an illusion of holographic video while testing spatial audio and aspects of presence in this VR concert film/game synthesis. This study was exploratory and experimental, but based on a vision on how interactive film experiences could be in near future. At this point there were still few studies on aural spatiality and its influence on presence, which justified the conducted study. This culminated in our first semester report called **"The Effect of Spatial Sound on Presence in a Shared VR Concert Experience"**. This study revealed a degree of significance on how spatial sound contributes to the feeling of presence by the user[22].

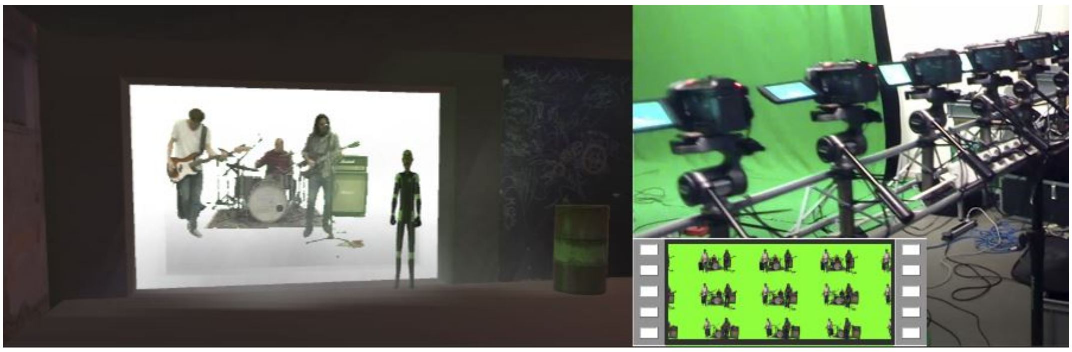

Figure 2.2: Illustration of the rig we used to capture holographic video footage for our 7th semester production in 2015. This photo also shows the holographic video format that we invented for the purpose. This was captured by 15 cameras and implemented in Unity [21]

From february 2016 we continued as a student group exploring cinematic technologies such as wave field synthesis which is a spatial sound technology utilized in cinemas, advanced sound studios and laboratories, consisting of 4 arrays of speakers surrounding a square audience/listener area. We continued our studies related to presence theory in this environment, but we also started working with the QoE framework defined by ITU and the methodologies proposed in "Quality of Experience - Advanced Concepts, Applications and Methods" from Springer in 2014[24]. In this particular study we investigated and implemented enhanced audience presence in the sound in order to investigate its impact on the feeling of social presence and the overall acceptability(QoE) of the production. We recorded a live concert with a 7-cam 360 video rig, and the sound channels was captured through the mixer, which were then mapped to the WFS Super Collider software. This VR concert experience could be experienced inside the listener area with the viewer wearing the Samsung Gear VR headset. We conducted a study with 30 participants in a "between-group" experiment. The title of the 2nd semester report was **"Sound Design in Virtual Reality Concert Experiences with Wave Field Synthesis"**, and based on our findings there were no immediate significant increasement of presence when participants were exposed to emphasis of audience in the listener area[18]. This work was also published in Los Angeles as a poster in 2017 IEEE Virtual Reality (VR) Proceedings[16], listed as **"Sound design in virtual reality concert experiences using a wave field synthesis approach"**.

Figure 2.3: Illustration of the setup we used in 8th semester to capture and replicate and sound environment in VR. This photo collage shows the mapping in Supercollider, the home build 7-cam 360 video rig, and the WFS installment at ÁAU Copenhagen [16]

Our selection of implementation platforms and distribution formats in our different projects described above, reflects that we as a study group both have been working with state-of-the-art technologies while showing willingness to think out of the box and to propose our visions for film and concert experiences in VR. Following our academic work we have through out our internship been focused on capturing concert experiences, while also working with additional fields in 360 video, which will be described in next chapter.

2.2 Internship, Panorama Video

After we finished our last academic work before our master thesis at AAU Copenhagen, we went on an internship in collaboration with Janus Heiberg and Panorama Video[17]. In this chapter we will describe our activities which lead to a large mass of captured material, including the live concert footage from Antone's in Austin Texas, that were implemented in our initial productions and tests contained in this master thesis.

From late October 2016 to early January 2017 we went to Hollywood in Los Angeles and Austin Texas to experiment with other fields of 360 video, listed as following[17]:

1) Journalism
2) 3D Sound
3) Concert Experiences
4) Sports & Events

During our 2,5 month stay we gained experience with the Samsung Gear 360 camera and the Zoom H2N 3D sound recorder which allowed us become more confident in working with spatial sound and the camera that we selected to continuously work with, in our master thesis.

Figure 2.4: Illustration of the travel rig we used in our internship semester to capture and replicate a sound environment in VR.[17]

While our primary activities during the internship was focused on concert experiences we also explored fields as journalism. We recorded interviews during the presidential election in 2016, filmed demonstrations, events and many compelling situations like Hoover Dam and Grand Canyon. The internship culminated in

Austin Texas where we were granted permission to capture several artists during two nights at Antone's. As we were filming under difficult and improvised conditions, we selected to work with the Jackie Venson sequences as they came out best. The concert with Jackie Venson was captured with the Samsung Gear 360 Camera, a Zoom H2N recorder and we were additionally permitted to capture the microphone channels from the mixer[17].

Figure 2.5: The photography shows Jackie Venson at Antone's and our rig placed on the stage. On the top right some captured situations during the presidential election.[17]

The finalized material from the Jackie Venson concert was then included as post-production, implementation and test material in this master thesis as a logical continuation of our efforts during our internship.

2.3 Research Question

As we have been working with presence theory in several of our previous projects, we are now focused on quality assessment of VR Film as a natural conclusion of our academic work. It can be said that we are both proposing methods to quality assessment while conducting research in a narrowed down scientific field. Hopefully the results of our evaluation can also reveal other interesting aspects with regards to the impact of stereoscopy and ambisonics on motion sickness, presence and immersion. These aspects will be treated in our final evaluation, discussion and conclusion chapters.

In this section we will propose our research question in order to proceed to our analysis chapter where will consider the theoretical and the technological approach to our research question which is formulated as:

"How do spatial features of sound (ambisonic) and imagery (stereoscopy/3D video) contributes to the quality of experience (QoE) in 360 vr concert experiences?"

The term "spatial features" will be divided into respectively aural spatial features (3D Sound/Ambisonics) and visual spatial features (3D Imagery/Stereoscopy), which will be treated and considered as two separate areas of our research in our analysis, implementation and evaluation.

Chapter 3

Analysis

In our analysis chapter, we have gathered the theoretical and technological foundations for our study on quality assessment of VR film experiences. We will explain the terms briefly to the reader as our final goal is to define measures of our QoE analysis chapter, which we will finally utilize in our implementation and evaluation hereof.

3.1 Requirements

One can say that we are both proposing methods for quality assessment of VR film experiences, while also hoping that our study can reveal to which degree a technology or method can contribute to a quality experience by its measures. Whether we seek a more general understanding or concrete approaches on quality assessment cases, we must establish some subjective "quality facilitators" as requirements for our experimental setup as follows:

Distribution Format/Medium/Platform Requirements:

1) Medium must support stereoscopic 360 degree video
2) Medium must support ambisonic sound
3) Production must be delivered on a state-of-the-art medium/platform

Video Production Requirements:

1) Must have minimal Parallax Issues, due to the limitations of camera positions in small crowded venues
2) Must support optimal conditions to avoid motion sickness
3) Format must be supported by state-of-the-art medium/platform

Audio Production Requirements:

1) Must have high quality spatial and stereo features
2) Must have minimal clipping or noise issues
3) Format must be supported by state-of-the-art medium/platform

These are the most elementary requirements, areas and conditions for us to proceed, whether we intend to test high quality 3D audio or 360 degree video footage in VR. The following chapters in our analysis will seek to address fulfillment of these requirements as widely as possible in the experimental setup and implementation.

3.2 Ambisonic Sound

In our project we have been working with monophonic sound representation, stereo sound representation and ambisonics, which we will elaborate on in this chapter.

Monophonic sound:
Monophonic sound (mono sound) is when all audio signals have been mixed and are being played though one audio channel, that being said do not confuse one audio channel for one speaker. There can be a lot of speakers producing sound, though with mono sound all the speakers would all play the same single channel audio, resulting in no indication of directional cues in the sound. With mono sound everyone will be hear the same sound representation no matter what [20].

Stereo sound:
Stereo sound has two independent audio channels instead of one as mono sound. So when the sound is played one would be able to distinguish where different instruments were placed in the sound image. Though if the listener are wearing headphones and are able to hear the stereo sound from them, then if the listener rotates 180 degrees, it would not have an effect on the sound image. So if the listener could hear a drum roll in the right ear, then (s)he would still hear a drum roll in right ear[20].

Ambisonics:
The difference between stereo sound and ambisonics is that when a person listens to the ambisonics sound, and if the listener rotates 180 degrees, then the drum roll the listener heard in right ear, will change to the left ear like a pan effect[8]. Ambisonics are in short a spherical sound field which is based on the decomposition of spherical harmonics and encoding of four channels Called: W, X, Y, and Z [29]

where the X axis represents the front-back, Y axis represents the left-right, the Z axis as up-down and the W is the pressure signal (omnidiractional) [6].

Ambisonics exists in different orders, the first order is the basic order and is what you call a B-format,[29]. The amount of channels is what determines what order of ambisonics you have [6].

The first order of ambisonics has 4 channels which is W, X, Y and Z [6].
The more channels you have, the more detail richness and data, and the higher the order becomes. So from first order with 4 channels, one must add 5 more channels to reach second order of ambisonics. To reach 3rd order, one must aditionally add 7 more channels, reaching a total of 16 channels[6].

After having presented the semantics of the spatial sound format known as "ambisonics", which will be implemented in our experimental production, we will continue in the following sections with a brief SOTA with regards to the audio capturing technology that we will consider to utilize.

3.2.1 Sennheiser AMBEO® VR MIC

Ultimo 2016 we were invited by DR (The Danish Broadcasting Corporation) to record a session with their philharmonic orchestra. Here we got hands on experience with the AMBEO® VR MIC which is highly suitable for recording of ambisonic sound with high quality. In this section we will briefly introduce the reader to its specifications.

Figure 3.1: The collage shows on left the microphone that can be seen mounted on the rig under the 360 video camera. On the right side the mic is displayed with its core parts, and its connectivity to a multi-channeled field recorder

The AMBEO® VR MIC is a condenser microphone with a frequency response between 20 Hz - 20 kHz, a maximum sound pressure level of 130 dB(A) and it delivers ambisonic A-format from its capsules. The tetrahedral microphone runs with 4 phantom powered mics that can also be used with its power supply. The microphone must be connected either to a audio interface with at least 4 inputs or to a field recorder with minimum 4 XLR input channels. [35].

In our earlier work, we found that the Sennheiser Ambeo VR delivers a very high quality audio, however this mic requires a lot of power to run in longer sessions. As we are not sure of the location's/venue's ability to provide power without interfering with audience and crowds close to our rig, we will be also considering the Zoom H2N, described in next section.

3.2.2 Zoom H2n

The Zoom H2n has five built-in microphones. Where two of them are arranged as an X/Y pair, and while the other three are configured in an MS ("Mid-Side") pattern. They can be used separately for X/Y or MS stereo recording, or combined to create 2- or 4-channel surround sound[38]. Additionally it can be used to capture B-format and it delivers 24-bit/96 kHz audio. This recorder runs on AA batteries and supports up to 32GB SD-card storage.

Image Source: http://www.sound7.be/ZOOM-H2n/en

Stereo Mode

REC Format	SD/SDHC Card Capacity				
	2GB	4GB	8GB	16GB	32GB
MP3 128kbps	34h 43m	69h 26m	138h 53m	277h 46m	555h 33m
MP3 320kbps	13h 53m	27h 46m	55h 33m	151h 6m	342h 13m
WAV 16bit/44.1kHz	3h 8m	6h 17m	12h 35m	25h 11m	50h 23m
WAV 24bit/48kHz	1h 55m	3h 51m	7h 42m	15h 25m	30h 51m
WAV 24bit/96kHz	57m	1h 55m	3h 51m	7h 42m	15h 25m

4CH Mode

REC Format	SD/SDHC Card Capacity				
	2GB	4GB	8GB	16GB	32GB
WAV 16bit/44.1kHz	1h 34m	3h 8m	6h 17m	12h 35m	25h 11m
WAV 24bit/48kHz	57m	1h 55m	3h 51m	7h 42m	15h 25m

*The maximum file size is limited to 2GB.
*Recording times are approximations. Actual times may differ according to recording conditions.

Figure 3.2: Table describes the capacity of the Zoom H2n recorder. On the right side a product image of the Zoom H2n

After having experience with both of the recording devices, the Ambeo VR Mic and the Zoom H2n during our previous study related work, we were confident that both devices could deliver high quality stereo and ambisonic formats, while also being highly acknowledged as state of the art in their field. We chose the Zoom H2n due to its ultra mobility and power supply independent ability. As we had already evaluated material captured from the Jackie Venson concert in our initial tests, we relied on it due to its relatively high acceptance score. In the implementation chapter we will elaborate further with regards to the post production and utilization of the format.

3.3 Stereoscopic Video

As our intention for this master thesis is to produce and evaluate stereoscopic video, we will in this section quickly explain what sterescopy is and describe the two available stereoscopic camera rigs that can be utilized in our production.

3.3.1 Stereoscopy

A stereoscopic display is different compared to non-stereoscopic, because it displays two different images for each eye, but you are still able to see that it is the same image. A person with normal vision has a little space between each eye, and when that person looks at an object, then because of the space between the right and the left eye, then the objects will actually be registered a little differently in each of the eyes fovea[27].

By using a stereoscopic display, you obtain more depth in the image because of something called the binocular disparity. Binocular disparity is the difference between the positions of what images the the left eye and the right eye sees, compared to each other [27].

To make the stereoscopic effect work, the objects the eyes sees have to be inside a specific area called the horopter [27]. The eyes are registering what they see at a slightly different position compared to each other, but there are an area outside of the horopfter where the stereoscsopy does not work. This occurrence is called binocular rivalry [27], and is also referred to as parallax. This occur when the difference in the distance in the binocular disparity are too great, and then the eyes will see double as a result of this. An easy example for this is if you raise your hand in front of your face while pointing to the ceiling, look at the pointing finger and then slowly pull your hand closer to your face. At some point you will end up seeing two fingers pointing towards the ceiling. Each eye are still seeing the same finger, but just at two different positions one for each eye.

3.3.2 GoPro Hero Black 14-Cam Stereoscopy

In our previous studies, as described earlier, we have been working with the 14 Camera GoPro rig which is also capable of capturing stereoscopic 3D footage and is perfect for capture of large areas with long distances to nearest objects. This rig can be considered to be a logic choice for our production, under the right circumstances. However, this type of rig is very hard to handle under difficult conditions due to many reasons. The most critical aspect of the 14-Cam GoPro rig is its parallax issue, which means that **there must be a minimum 2 meter distance from the cameras to the closest objects in the scene**. In the concert situations we have been working with earlier, we have experienced that this is really the most critical problem to be considered as concert venues tend to be very crowded and the stitching quality remains an important concern. Furthermore this type of rig develops a lot of heat, and is very sensible to high temperatures, which is causing sudden shut downs. ISO Light sensitivity goes down to ISO 400, which is equivalent with the Samsung Gear 360 Camera. The quality from the GoPros is somewhat better in low light situations, while the parallax issues stays dominantly critical for our intended use.

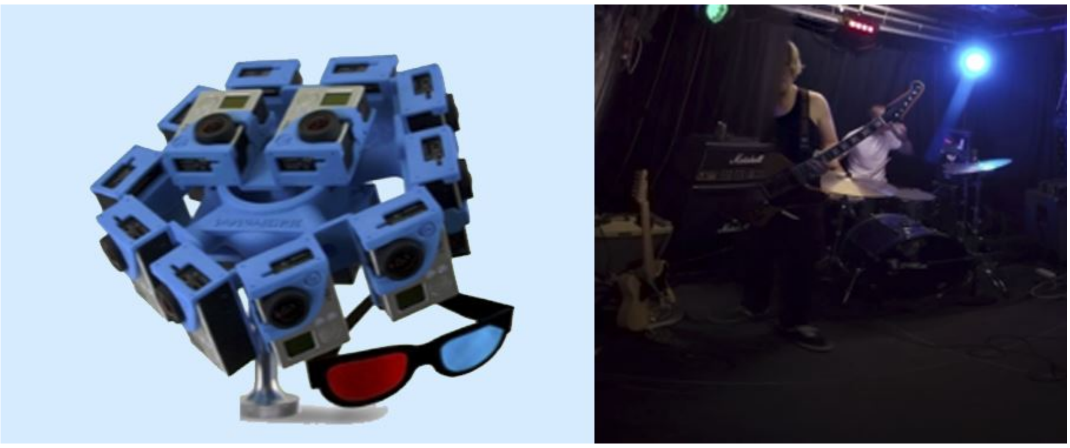

Figure 3.3: The photography shows on the left the 14 Cam GoPro Rig, and on the right an example of a typical parallax issue [21]

Due to parallax issues in particular, we decided to look for alternative ways to capture stereoscopic concert footage, which will be described in next chapter.

3.3.3 Samsung Gear 360, 2-Cam Stereoscopy

After having been working with the Samsung Gear 360 video camera during our internship, we were still not aware of its potentials with regards to stereoscopy until early february 2017 at the S.M.I.L.E Laboratory opening event at AAU. Here we met Maria Engermann and Signe Ungermand who showed us a stereoscopic footage of their short film "Separate Silences" captured with 2 Samsung Gear 360 cameras mounted and stabilized on a helmet. What was interesting about this rig was its **ability to capture stereoscopic footage with a very low degree of parallax** in particular, as we need flexibility with regards to camera positions [37].

Signe Ungermand and Maria Engermann wrote a bachelor thesis that was based on their work with a pioneering short film in VR called "Separate Silences", which was selected to enter the Cannes Film Festival in 2017 [10]. In their bachelor thesis, they conducted research on the impact of the camera distances, between the two cameras **"the stereo plate"**, on motion sickness [37]. There are considerably many causes for motion sickness but nevertheless our focus for this study is to produce material that are representative for "industrial" utilization methods, which we consider their work to do substantially with regards to minimizing motion sickness and optimizing quality experiences in stereoscopic production environment.

PLATE AND RESOLUTION TEST			
**NB = Not bothered *B = Bothered*RW = Resolution worse *RB = Resolution better*			
Sex (M/F)	**Distance between eyes (cm)**	**Reaction to GoPro footage (6,5 cm apart)**	**Reaction to Samsung footage (6,3 cm apart)**
M	5,7	NB*, RW*	NB, RB*
F	5,9	B*, RW	NB, RB
F	6,3	B, RW	NB, RB
F	6,4	NB, RW	NB, RB
M	6,5	NB, RW	NB, RB

Figure 3.4: Table showing results of comparison of perceived resolution and to what degree participants were bothered or stressed by the "stereo plate" [37]

According to Signe Ungermand two out of five test subjects were bothered by the GoPro stereoplate and none were bothered or stressed by the Samsung cameras. Most importantly all 5 participants thought that the Samsung Gear 360 cam provided a better resolution [37].

After having presented the reader for our initial options with regards to stereoscopic camera rigs, their limitations and our minimal requirements with regards to parallax, and with respect to motion sickness, we can now proceed with a decision on the most suitable technology for our experiment. Based on Signe Ungermand's

Figure 3.5: The photography shows Signe Ungermand demonstrating the helmet mount used in "Separate Silences""[37]

related tests on the stereo plate, with regards to perceived quality of resolution and motion sickness and the minimum parallax issues, we conclude that this this capture method is the most suitable for our experimental implementation. In the implementation chapter we will elaborate further on the rig that we build for our production.

3.3.4 Head Mounted Display

There are a lot of different HMD's on the marked right now, however we chose to look more into three of the state-of-the-art HMDs. These are:

HTC's Vive consumer version 1
HTC's Oculus Rift consumer version (C.V.) 1
Samsung Gear VR C.V. 1

In table 3.1 we compare the three HMDs specifications and requirements. We chose to work with Samsung Gear VR out of the three HMD's, since first of all, it had video players which supported stereoscopic imagery and ambisonics sound, it was one of the lighter HMD's and Samsung Gear VR is the only one of them which does not require the HMD to be wired to an external stationary computer in order to work.

Table 3.1: Specifications for Oculus rift, HTC Vive and Samsung Gear VR [3], [4], [25], [31], [32]

Categories	HTC Vive C.V. 1	Oculus Rift C.V 1	Samsung Gear VR C.V.1
Hardware Requirements	Graphic card: NVIDIA GeForce™ GTX 1060 or AMD Radeon™ RX 480, CPU: Intel™ Core™ i5-4590 or AMD FX™ 8350, equivalent or greater, Memory: 4 GB RAM or more, Video Output: 1x HDMI 1.4 port, or DisplayPort 1.2 or newer, USB ports: 1x USB 2.0 port or newer, OS: Windows™ 7 SP1, Windows™ 8.1 or later or Windows™ 10	Graphic card: NVIDIA GTX 1050Ti / AMD Radeon RX 470 or greater, CPU: Intel i3-6100 / AMD Ryzen 3 1200, FX4350 or greater, Memory: 8GB+ RAM, Video Output: HDMI 1.3 video output, USB ports: 1x USB 3.0 port, plus 2x USB 2.0 ports, OS: Windws 8.1 or newer	Galaxy S8+, Galaxy S8, Galaxy S7, Galaxy S7 edge, Galaxy S6, Galaxy S6 edge, Galaxy S6 edge+
Weight	563g (excluding cable)	470g (excluding cable)	318g for HMD, 152g for phone, in total 470g
Resolution	2160 x 1200	2160 x 1200	2560 x 1440
Field Of View	110 degrees	110 degrees	96 degrees
Tacking Area	15 x 15 feet	5 x 11 feet	Only head tracking
Refresh Rate	90 Hz	90 Hz	60 Hz
Platform	PC	PC	Galaxy S8+, Galaxy S8, Galaxy S7, Galaxy S7 edge, Galaxy S6, Galaxy S6 edge, Galaxy S6 edge+

3.4 Theory

The following section will describe the different theories of psychology that are considered and used during this project. These are mostly focuses around presence, immersion and quality theory.

3.4.1 Presence

In this chapter we will introduce the reader to the theory of "presence", mainly as formulated by the often cited Matthew Lombard and published in volume 3, issue 2, Journal of Computer-Mediated Communication, first published in 1997 [19]. Matthew Lombard is in our opinion the most flexible but still precise in his proposed definitions of presence. He has also kept his work updated, and is by many considered as go-to reference with regards to understanding and utilizing elements of presence in games, film and VR experiences. Presence is by us considered to be a central quality measure, and we will be embedding the aspects of presence into our final questionnaire in order to quantify (by Mean Opinion Score) elements of presence in the subjective QoE assessment. We will not conduct a comprehensive presence evaluation as such but only focus on what we find to be the most relevant initial quality measures with regards to the feeling of presence in VR film experiences.

Matthew Lombard has proposed 6 general definitions of "presence", briefly summed up as:

1) Presence as Social Richness describes how richly the conveyed environment facilitates communication with other participants[19].

2) Presence as Realism describes how the viewer perceives the imagery and sound as plausible or even authentic. "True to life" (Social Realism) replication of the captured environment through imagery and sound plays a central role in conveying experiences with elements of realism[19].

3) Presence as Transportation is defined as the users feeling of "being there", "it is here" and finally "we are together"[19]. The key element here is the feeling of being transported into a scene or a location which effectively contributes to the suspension of disbelief.

4) Presence as Immersion is defined as both perceptual and psychological. The perceptual immersion is related to the compelling visual and sound environment replication which involves and immerses all senses. The psychological immersion being described as engagement, engrossment and the feeling of involvement[19].

5) Presence as Social Actor Within Medium is defined as having the opportunity of interacting with virtual peers or characters through artificial intelligence and such[19].

6) Presence as medium as social actor is defined as providing cues for the media user leading the user to treat the medium as a social entity[19].

Many variables have influence on different aspects of presence which is described as influential factors, including spatial sound quality attributes and loudness. Stereoscopic images can according to Matthew Lombard contribute to presence which has been shown in previous studies[19].

After having briefly summed up the 6 generalized definitions and their influential factors and based on our design choices/limitations we have chosen to work with the 3 specific definitions that we will evaluate in our production: Presence as Realism, Presence as Transportation and Presence as Immersion. As designers of an experience our goal is to evaluate the perceived and psychological aspects, for us to better understand and optimize any concrete VR concert experience or production. Therefore we will elaborate further on how questions were formulated in the evaluation chapter.

3.4.2 Engagement & Continuation Desire

The following section will look at user engagement and continuation Desire. These concepts are mostly used in game and interaction research, but can be applicable to other experiences.

The concept of engagement involve the factors that makes a player or user want to continue an game or experience [34]. An as such continuation desire are the expression of how much a user are in engaged in an experience [34].

Although there concepts are mostly used in interactive experiences. We will use this as an expression of how much a user feel that the experience has been immersive. In this, a higher continuation desire is an expression of a quality measure of the experience.

3.4.3 Motion Sickness

In this section we will look into field of motion sickness, as it is a highly regarded quality measure, that people in general and that our test subjects have seamless experiences so we can better conduct our comparative study under comfortable settings.

Motion sickness also known as "VR Sickness" is still a relatively new and unexplored academic field. However, the US Army has due to their need for many simulation types and training purposes conducted a lot of research on the topic[12].

3.4.3.1 Susceptibility

There are many influential factors that causes "Motion Sickness" also known as "VR Sickness". Studies have shown that **age** has significant influence for people between 2-12 with regards to susceptibility and it has been suggested that VR follows the same pattern[14]. **Postural stability** is known to increase susceptibility to visually-induced motion sickness, nausea and disorientation[14]. **Flicker in the display** is also known to cause motion illness[14] with regards to system influential factors. **Experience with VR** is known to be a human influential factor that decreases the chance of the user having discomforts as they develop familiarity with VR experiences[36]. **Gender** is in fact also known to be a human influential factor as women has a wider field of view and due to their hormonal differences from men[13]. Furthermore women are also more susceptible to motion sickness during ovulation[2]. This corresponds with other human influential factors related to the **health condition** of the user when being fatique, under sleep deprivation, during viral or respiratory illnesses[26]. **Mental Rotation Ability** is also known to reduce motion sickness, especially for users who are specifically trained in this ability[26]. Lastly **Motion Sickness Sensibility** is lastly, if not the most common term for people who have increased sensitivity for motion sickness are in high risk of "VR Illness"[7].

3.4.3.2 Technical Factors

There are many influential factors in play as causality for motion sickness. The most common IF' are known to be; mismatched motion[7], mismatched "field of view"[15], mismatched motion parallax[11] and mismatched viewing angle[30]. Based on our own experiences with 360 video, bad stitching and distortion of the visual space has become our greatest concerns with regards to preventing motion sickness.

3.4.3.3 Symptoms

Motion sickness symptoms can vary, but is often described as nausea, dizziness, headache which is caused by eyestrain and other oculomotor distress[30]. It is also known that long time spent in VR experiences increases the chances of getting symptoms of motion sickness[30].

Since we have now established knowledge with regards to motion sickness, we can now proceed with considerations that minimizes the risk for our test participants to develop motion sickness symptoms during our experiment. This will be reflected through our questionnaire and in our evaluation chapter.

3.4.4 Quality of Experience

In this chapter we present the theoretical framework of QoE that we have also based our previous studies on, with regards to defining subjective quality measures and proposed methods for evaluation hereof.

3.4.4.1 definition of QoE

ITU, International Telecommunication Union, who are working with standardization of technologies has proposed a definition of "Quality" as quoted:

> **"The outcome of an individual's comparison and judgment process. It includes perception, reflection about the perception, and the description of the outcome. In contrast to definitions which see quality as "qualitas", i.e. a set of inherent characteristics, we consider quality in terms of the evaluated excellence or goodness, of the degree of need fulfillment, and in terms of a "quality event"." [1]**

Furthermore the definition can be elaborated as quoted:

> **"Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state." [1]**

ITU suggests both evaluation of services and applications, and for this matter the definition of an application is as quoted:

> **"A software and/or hardware that enables usage and interaction by a user for a given purpose. Such purpose may include entertainment or information retrieval, or other." [1]**

As our variables are both ambisonic sound and stereoscopic video, these can be described as "features":

> **"A perceivable, recognized and namable characteristic of the individual's experience of a service which contributes to its quality." [1]**

3.4.4.2 Absolute Category Rating (ACR)

In **subjective quality evaluation** the ACR rating method is a method where the test participant is presented to stimuli and asked to separately rate specific aspects of an experience on a Likert scale[24].

The ACR method typically uses a five-level rating scale as following[23][5][24]:

5 Excellent
4 Good
3 Fair
2 Poor
1 Bad

After the test participant has been exposed to stimuli, he or she will be presented to a questionnaire with rating of subjective quality measures, known as "Mean Opinion Score" which is described in next section.

3.4.4.3 Mean Opinion Score (MOS)

In this section we will briefly introduce the reader to the term MOS, which is a central part of our final evaluation.

The MOS is generally used to quantify aspects and features related to the subjective evaluation method[5][24]. The MOS is a vital instrument to judge the overall aspects of a system, and is recommended by ITU, International Telecommunication Union, for evaluation of applications and services related to audio and video.

3.4.4.4 Quality Measures

In this section we will finally propose our subjective quality measures for a given VR Film experience in our experimental setup. These measures will particularly be used to quantify the QoE, based on what we consider most important quality aspects in a VR concert film experience.

In our experimental setup we are embedding values in our questionnaire designed to reveal how quality is perceived by the test participant. The quality measures reflects what we believe is defining a quality VR experience as:

"An immersive and compelling experience that does not make the user sick, with high quality imagery and sound".

Image Quality Measure
In our image quality measurements we will ask general questions related to the perceived quality. We will additionally simplify the questions in order to ask the participant about perceived "depth" as we are concretely working with evaluation

of 3D imagery.

Sound Quality Measure
In our sound quality measurements we will ask general questions related to the perceived quality. We will additionally simplify the questions in order to ask the participant about perceived "direction" as we are concretely working with evaluation of 3D sound known as ambisonics.

Presence as Quality Measure
As we are trying to convey both immersive and compelling experiences, we will try to ask simplified questions related to presence (realism, transportation and immersion)[19].

Motion Sickness as Quality Measure
Lastly we are prioritizing the risk of inducing motion sickness to our users with great concern. We will ask questions directly related to discomforts so we can identify eventual mismatched features in our production.

The quality measures contained in this section will be reflected/embedded in our QoE analysis and the questionnaire which we will elaborated on in the evaluation chapter.

Chapter 4

Implementation

In this chapter will showcase how the different solutions for the two different tests was implemented. The first part of this chapter will showcase the production made for the initial test and the last part of this chapter will present the productions and recording made for the final test.

4.1 Scenario 1: Jackie Venson, Antone's

In the following text we will talk about the pre-production and post-production for Scenario 1, it is this scenario that are used to gather the data for the initial test.

4.1.1 Recording Scenario 1 : Jackie Venson, Antone's

In this section we will talk more about the specific recording mentioned in section 2.2 at Antone's. This to explain what data we have worked on the the following sections for Scenario 1.

Audio Recording

For the audio recorded at Antone's we used two methods, one where we used the the Zoom H2n that can record ambisonic sound (however only the horizontal changes) and sound from the mixer (this is higher quality audio).

When using the Zoom H2n to record spatial audio, one has to change the recording setting to spatial audio. Another aspect to be aware of is to place the Zoom H2n recorder correctly so the front of the recording follows the font of the video recorded.

When recording from the mixer, we used Reaper to capture the recording. For this to work, we needed to get and install the plug-in belonging to the mixer. With this, each instrument and the singer was recorded on separate tracks. It is these tracks that are used in section 4.1.2.1 to make the ambisonics.

Video Recording
For the video recording itself, a single Samsung Gear 360 was used. Before the recording, we made sure that it would record in the highest resolution and that the white balance was set correctly (in this case auto seam to do a good job). For the placement of the camera we decided to put it on stage. This placement was chosen, because from there we would get more light and more movement in the image as the band would be almost in front the camera. Furthermore the Zoom H2n recorder was placed with the camera on the rig.

4.1.2 Audio Scenario 1: Jackie Venson, Antone's

In this section we will elaborate in details on how we went from post production to finalization with ambisonics.

The recording session at Antone's provided us with material captured from the mixer in 96 kHz 24 bit quality per channel. Jackie Venson and her band was recorded through a traditional microphone setup, where each instrument was captured through microphones and stored through the mixers USB-output onto a laptop with multiple simultaneous recording channels in Reaper[17].

All files were separately normalized to 0db, and then imported to a multi-track mix down project. Here, all the recorded tracks was balanced in order to be suitable for two purposes:

1) A balanced mono version (Lossless .wav format)
2) A balanced mix down with separate tracks (Lossless .wav format)

Both "productions" were based on the same settings and adjustments with regards to normalization and eventual useful enhancements like minimal remastering techniques. There was done "as little as possible" besides optimization of loudness, bass- and mid range adjustments only to translate well on headphones and speakers.

The files from the Jackie Venson were now ready for the finalization process which will be described in next section.

4.1.2.1 Scenario 1: Finalizing Ambisonic, Jackie Venson, Antone's

The following will describe how the audio stem was taken into Reaper and how it was spatialized. This spatialization was made using Ambisonic Tool Kit (ATK) for reaper. As the Zoom H2n files are B-format ready for ambisonic, they only needed to be synchronized and cut to fit the song chosen, to fit the video.

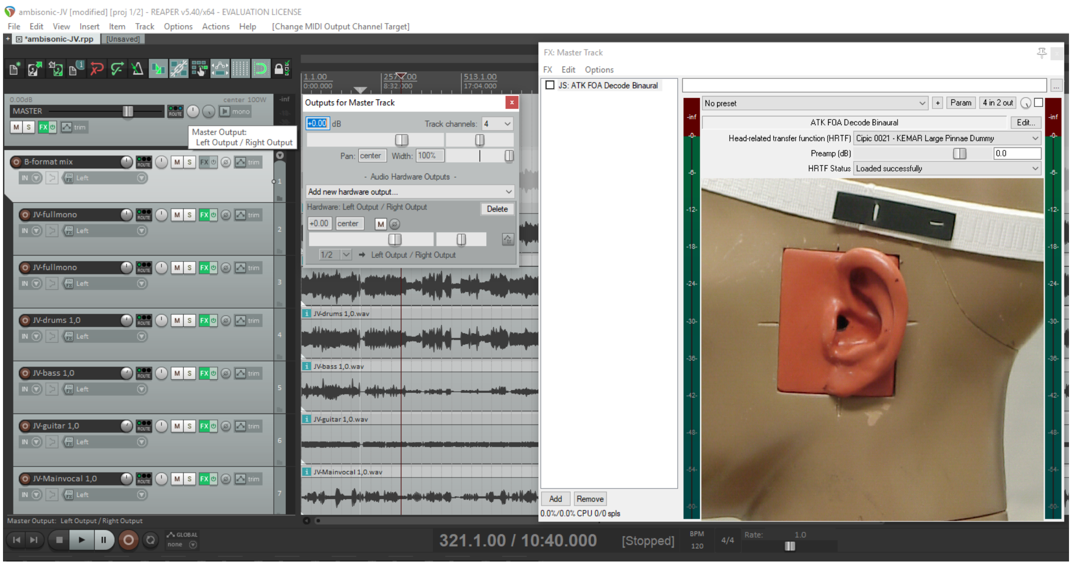

Figure 4.1: Here are a screen shot of the program reaper with the setting used when making ambisonics in this program. The window that is open show the binaural decoder being used while making the spatialized audio

First setting up for making ambisonic in reaper we needed to change a few settings. First we have to change the master channel to 4 instead of 2, this is because we are making 1st order ambisonic that have these 4 channels, as can be seen in figure 4.1. This has to be done for all the channels that we use further on. Next we have to setup a binaural decoder on the master channel, here we used the KEMAR doll setting in the JS: ATK FOA Decode Binaural. This is because when we are placing the different sound sources we need to actually hear where they are placed and a binaural decoder help us do this. We can see the plugin in figure 4.1.

Next we make an empty track that will gather the 1 st order ambisonic. We call it B-format mix. Next up we have to add each separate sound that we want to place in the sound space, here we used the JS: ATK FOA Encode Planewave to place sounds in the ambisonic space, see figure 4.2. This is then done for all the audio sources bass, guitar, vocals ect. When exporting these files one has to make sure that all the channels are exported.

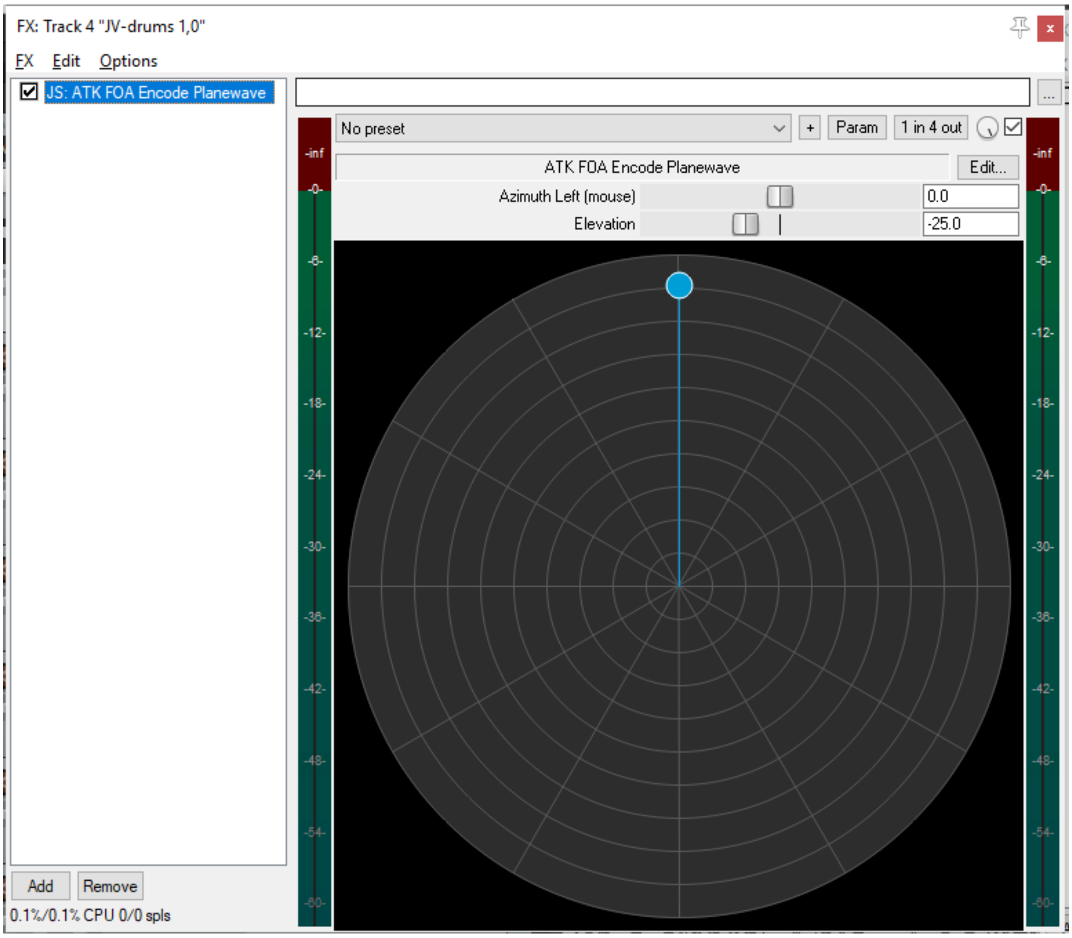

Figure 4.2: This is the specific plugin we used to place sound in space, the dot is where the sound are, then we can change the azimute to change where the sound are place horizontaly and we can change the elevation to change if the sound comes from above or below. In this case it is the drum

The next step is to synchronize the audio to the video file. This is best done by taking the audio into Reaper and listen for where the match is and then looking at the waveform to get them to match up. After the synchronization we need to get 2 versions from this mix, one for the mono version and one for the synthesized ambisonic. This is done by rendering the sound two times, one where we render 4 channels out and one where the sound is compressed to one channel.
When this is done we need to mux the sound onto the videos this is done using the facebook spatial encoder. This will be explained later in this implementation chapter in section 4.2.3.

4.1.3 Video Scenario 1: Jackie Venson, Antone's

For the first scenario we are going to use the footage gathered from the internship mentioned in section 2.2. With this footage as used only one camera, we can not make it stereoscopic. Because of this there are a lot less post-processing, but as one can see in fig 4.3a, there are a lot of noise in the initial output we had from the recording. The following will briefly explain what was done in order to improve the footage.

The process used to improve the video and remove overexposure was done in Adobe Premiere Pro. Here we corrected the light by the use of gamma correction by masking one side of the video as the noise are only present on one side of the video. This was done during multiple passes, first we masked the larger effected area and turned down the light. This was followed by making the noisy dark area black. The last step was to light of the spotlights in the image as they had gone to dark for light. The finished version can be seen in figure 4.3b.

(a) This is the first Video of Jackie Venson before any optimization was done on the video

(b) This is the final video used when testing, -with the optimized light and noise removed

Figure 4.3: This figure are meant to be a comparison of the optimization of the video done for scenario 1

4.1.4 Scenario 1 Test Conditions

The test condition we where going to make of Scenario 1 are all monoscopic with different audio on, these are the audio condition:

- Zoom H2n ambisonic Conditio 1
- Synthetic Ambisonic Condition 2
- Mono sound Condition 3

Condition 3 was rendered out from premiere pro with a mixdown of the audio data we got from the mixer. For the two ambisonic version we neede to synchronize the audio to fit the video. Then we need to mux the audio onto the video file, this is done in the Facebook spatial encoder this process will be explain in more details in the end of this chapter.

4.2 Scenario 2: Rest in Beats, Rust

4.2.1 Recording Scenario 2 : Rest in Beats, Rust

Building our own rig

We have earlier looked into two different types of stereoscopic rigs we could use for our recordings and we decided the setup used in Separate Silence were the way to go. Though their setup was build as a head mount which is good if you wish to have a body representation, but we have not focused on that for our project therefor chose we to build a our rig as a mount for a tripod We build the a very simple rig out of a piece of wood, two Manfrotto 015 Adapter 1/4" to 3/8", hot glue and a tripod and two Samsung Gear 360 cameras and unitape. We build the rig by cutting a wood out in around 15 cm and glued the adapters to the wood with 6,3 cm distance between the center of the adapters. In between the adapters we drilled a hole so it could be attached to a tripod. We finished the rig by adding unitape to the adapters since we were not sure of what was front and back of the adapters when they were glued on so by adding the tape we could better fasten the cameras with the front in the desired direction.

Setup and Video Recording

Our stereoscopic rig was placed to the left in front of the stage on a plateau surrounded by a railing, with the front of the rig pointing towards the stage.
At the event at Rust were we not able to properly test the lighting for the imagery since they did not make a proper light check before the event started. Though we were told that the light would be turned off and they had a light show during the event. Because of the information given to us changed we the cameras settings and put ISO to 400 turned the exposure compensation up til max which was +3, and set the White Balance on auto. The HDR was turned of since it did not make that much of a difference on the imagery while we were testing the cameras before the event started.

Audio recording

The recorder used for the Rust event was the Zoom H2n It recorded with a 48 kHz and 24 bits on all 4 channels. It was attached to the same tripod as our stereoscopic rig. Both the rig and the audio recorder front was pointing towards the stage.

4.2.2 Audio Scenario 2: Rest in Beats, Rust

In this section, we will describe the basic steps that were needed to process the captured audio in order to prepare for final implementation in our production or distribution format.

Initially all the files from the Zoom H2N were imported to local drives and edited in Adobe Audition as described in the following steps:

Step 1, Normalization: The entire audio footage was normalized independently on all channels to ensure better presence of all directions (channels) in the mix. This was especially considered important for the preparation for Ambisonics implementation. We particularly considered to effectively optimize aspects of spatial loudness, clearness and eventually restore the bass heavy sound environment without applying too much filtering and effects as we to the highest possible degree intended to keep the authenticity and originality of the actual sound capture.

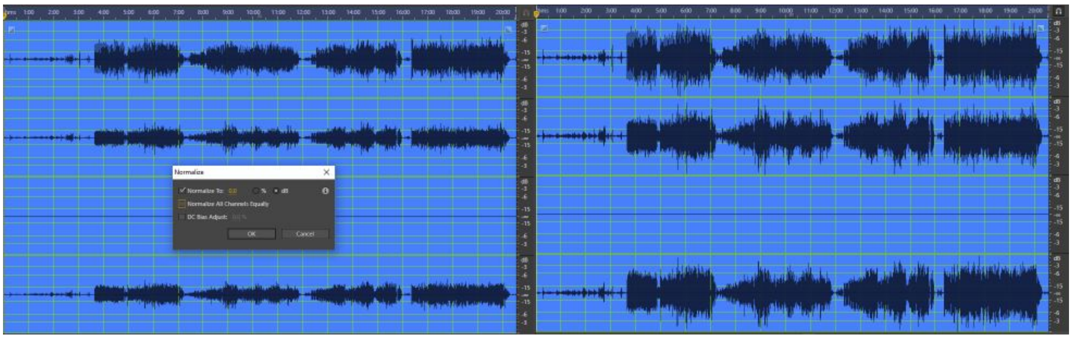

Figure 4.4: Illustration of how we in step 1 applies independent normalization on the signal, raising the amplitude to its maximum on all channels.

Step 2, Remastering: As the Zoom H2N was set to have recorded on a relatively careful level in order to avoid clipping sequences and unpredictable amplitude levels we were forced to leave headroom while recording the captured audio signal. This results in a "loss" of bass heaviness and some unbalance in the represented frequency spectrum to be restored into its original, by us judged subjective characteristics. Secondarily a gentle remastering also lowered the potentially clipping low frequencies, in order to have a "flatter" sounding and better representation of all frequency details in the music, without emphasis on any specific attribute in the mix. This was done while monitored on two Yamaha HS8 studio monitors, to check for well representation of mid frequencies, and on Beyer Dynamic DT770 PRO 250 ohms to check for well representation of lower frequencies in the mix. The main goal for the mix is to translate well on all possible devices, including lower grade headphones and speakers.

Figure 4.5: Illustration of how we in step 2 applied the flat remastering which increased the general amplitude of all frequencies represented and thereby restoring detail richness.

Step 3, Compression: As we were now working with a well translating flat mix with proper loudness, some of the emphasis on bass now needed to be heavier represented as actual live performances are are often perceived as very loud and bass heavy, depending on the music genre, and in this case an actual musical art form with explicit emphasis on bass in their live performances. For this reason we applied a gentle manipulation of the dynamics in the bass representation, to restore a subjective realistic live concert sound experience as close as experienced and remembered by us. Furthermore, this was also done to ensure better translation to consumer grade speakers and headphone devices while optimizing the conditions for quality sound experiences on consumer grade equipment.

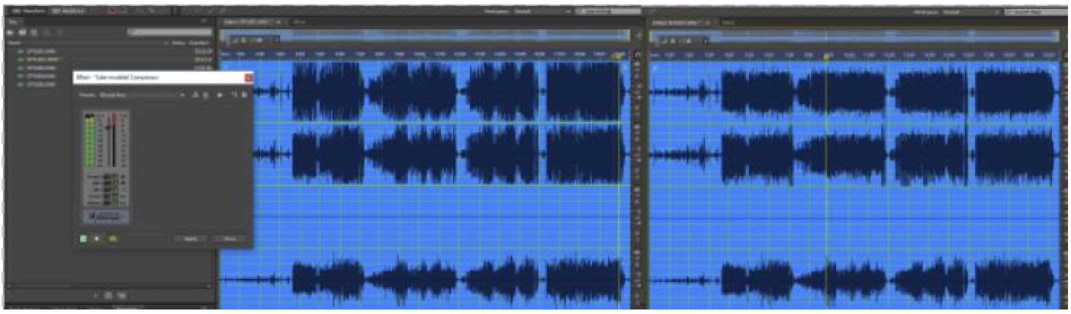

Figure 4.6: Illustration of how we in step 3 applies the compressor with gentle emphasis on bass representation generally lowered the signal as a wanted side effect that keeps the characteristics of the sound but enhances bass dynamics.

Step 4, Final Amplitude Adjustment: After having both optimized the lower and higher frequency representation in the mix we now have an overall well sounding and well translating distribution format prototype. The headroom left over from the gentle compression can now be carefully decreased by increasing the amplitude of the clear sounding mix. Now we have taken final control over the amplitude of the overall sound production and it can be utilized identically in both Stereo and Ambisonic productions on consumer grade audio hardware.

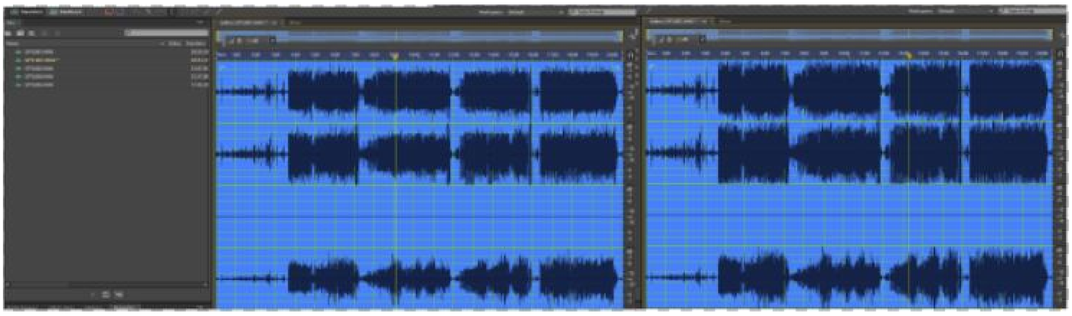

Figure 4.7: Illustration of how we in step 4 are gently increasing the amplitude to maximize the loudness effect with all sound attributes of a live concert experience kept intact, enhanced or restored

One can say that the main goal for the sound production was to restore a high degree of authenticity, while also manipulating features that optimizes the experience with respect to sound clarity, loudness, width among many concerns, but also to ensure that the production lives up to technological standards, variations and limitations. The final sound production was now ready for export providing a high degree of flexibility with regards to further customization's and most importantly ready for implementation in our Stereo and Ambisonic 360 video solutions. Further implementation hereof will be elaborated on in the video implementation chapter.

4.2.3 Video Scenario 2: Rest in Beats, Rust

In this section we will go through the steps that was used to make the video content used in the scenario 2 that was used for the final test. As these video was recorded stereoscopic there are few extra step to stitching and finalizing the images.

Step 1 : Autopano Video/stitching
The first step to make the 360 video after import of data is to stitch the videos together. As the footage was recorded with Samsung gear 360 we have 2 solution to make the stitching. One we can use the software that comes with the camera, Action Director, the up side to this is that it is fast and make a good stitch, the downside is that we will loose some quality in the image as we can not control the bit rate. second solution is to use the stitching program Autopano Video video, here we will get a slightly worse stitch but we can control the bit rate and we have more control of the image inside Autopano Video.

As the recording was made at a dark location and the camera that we use do not handle dark environments well, we could not afford to loose more quality in the image so Autopano Video was used.

The first step is to loaded the video into Autopano Video. After this one have to stitch the videos together for doing this in Autopano Video one have to get a file that the stitching can follow, our Template source are from michael pearce [28] (we will make our own for stitching the video for the other eye). Autopano Video needs 2 video in order to stitch, and as Samsung Gear 360 file have both video feeds on the same file, one have to make a identical copy and import it to Autopano in order to make the stitch. After the stitching the video have some weird balance as one can see in figure 4.8a. Because of this one have to move the video around until it looks like it has a good balance to the image, the balancing for our video can be seen in figure 4.8b. When doing this it is a good idea to render small snippet out to see if it looks good.

(a) This is how the video looks just after it has been stitched

(b) Here are the video after it has been moved and corrected to have the right center point

Figure 4.8: Here are an comparison of the video before and after finding balance in the video

When we have a good picture we open Autopano Giga from here we save the panorama. This panorama are going to be used to stitch the other files as it has all the change we have made. Because of this we should get the other video to look like the one we just made.

However when we did this we noticed that there was a small offset in the height between the two videos. As this will ruin the stereoscopic effect we had do adjust the height manually. After doing this we can render both video for further edition in premiere pro.

Step 2 : Final edit for Stereoscopic

When we now have Videos for both eye we need to make the final edit, so the videos can be used to be uploaded to youtube and for the different 360 video player. For this both video have to be rendered onto the same file in the right setup. The most common setup used when displaying 360 stereoscopic videos are over / under which means that the two video are placed above and below each other, the video for the left eye has to be on top.

This is simply done in Adobe Premiere Pro, by doubling the height of the sequence and moving the videos so that the left eyes video are on top as can be seen in figure 4.9. When rendering this video we have to used H.265 as if we use H.264 we can not export in full resolution.

Figure 4.9: Here we see how the final stereoscopic video looks like

The last problem that we have to take hand of before we export the video is that some of the other eye lens are visible on the video. For removing this, if we look at the right eye we will take the left eye video crop everything but the area where the lens are and remove it, and use it as a patch at the spot where the lens are. This will give some artifact in the video but it is better than the lens. This is then done for the other video also, ending out with the video we can see in figure 4.9.

Step 3 : producing the 4 files we are going to use in the test
In the end we need more than the stereoscopic video for the final test. we need monoscopic video to compare with and on both videos we need a stereo track and an ambisonic track. In total we need 2 monoscopic video and 2 stereoscopic video, with two different sound settings on them.

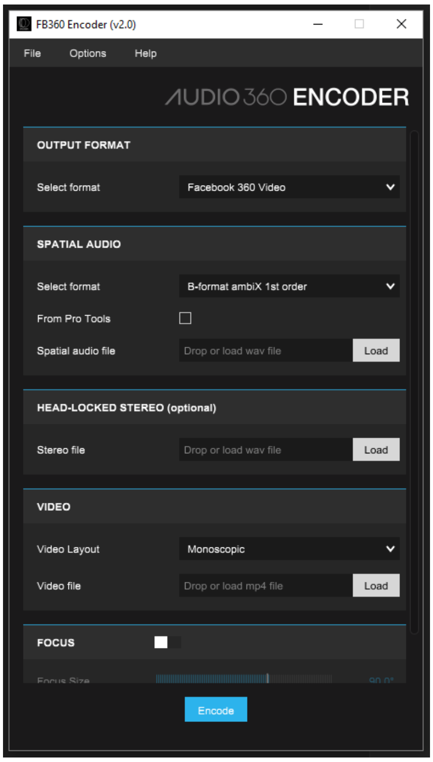

Figure 4.10: This is the interface of Facebook 360 spatial encoder

4.2.4 Scenario 2 Test Conditions

The following will describe how the 4 finale test conditions was made. These 4 conditions are:

- Video with monoscopy and stereo sound, condition 1
- Video with monoscopy and ambisonic, condition 2

- Video with stereoscopy and stereo sound, condition 3
- Video with stereoscopy and ambisonic sound, condition 4

Stereo Conditions For these condition we took the audio talked about in section 4.2.2, and imported it into premiere pro. This audio was then synchronized to fit the audio for the video. When this was done we could render condition 1 by taken half of the stereoscopic video to render it out monoscopic, this will apply the optimized audio in a stereo sound format. The same was done for condition 3 but with the full stereoscopic video.

Ambisonic Conditions For these versions we took the files rendered with stereo sound. We then in audition took the optimized audio from section 4.2.2 and cut this sound to fit the audio on the video. This was then rendered out in 4 channel format. The last step is then to "mux" this audio onto the video, this can be done using Facebook Spatial Encoder (see the interface on figure 4.10). In the Facebook Spatial Encoder we need to choose format to be Facebook 360 Videos, the spatial audio is B-format ambiX 1st order. Then we load the ambisonic 4 channel file into the program and the video. As we can see on figure 4.10 the video layout are set to monoscopic this can also be set to stereoscopic over/under, this is how we make our stereoscopic and monoscopic ambisonics versions.

4.3 Apparatus for Initial Test (Jackie Venson)

For the initial test we used the first consumer version of Samsung Gear VR together with a Samsung s7 smart phone with the resolution of 2560 x 1440 (Quad HD) [33]. The player used was Oculus' standard player Oculus Video version 2.5.23.

The videos we tested on were rendered out in the format of 264, the videos had 3840 X 1920 in resolution and were at 39.97 frames per second. The headset we used was a pair of Phillips Bluetooth headsets with the serial nr. SHB8850NC, though we used a wire from the headset to the HMD since we wished to avoid having any possibility of losing the sound because of Bluetooth connectivity problems .

4.4 Apparatus for Final Test (Rest In Beats)

For our final test we chose to use Samsung gear VR head mounted display (HMD) first consumer version with a Samsung galaxy S7 smart phone with a resolution of 2560 x 1440 (Quad HD) [33] in it. We chose to this display because of multiple reasons, first of all because of the resolution and because of the handiness of not having any wires between the the headset and an external computer, constraining the movements of the test participants.

For our final test used the 360 video player Samsung VR player version 1.71.1 since the standard video player Oculus Video player, which was the player we used in the first test, had some difficulties with playing our videos with ambisonics sound. The reason why the player worked for our initial test but not our final test is because in our initial test our test focus was on the audio, and therefore were our videos rendered out in the format 264. Whereas for our final test we tested on both the image and sound for and therefore were those videos rendered out in the format of 265 since this format allowed us to render our stereoscopic out with 3840 X 3840 frame rate, which was not possible in the 264 render format. Our videos were rendered out at 29.97 frames per second. The audio headset used for the final test were the same as we used for our initial test which were a pair of Phillips Bluetooth headsets the serial nr. SHB8850NC, and again as in the initial test the audio headset were wire from the HMD to the headset to avoid any possibility of Bluetooth problems.

To sum op the differences in what we used in our initial test and in our final test take a look at the table 4.1.

Table 4.1: Specifications for Initial test and Final test

Categories	Initial test	Final test
Frames Per Second	29.97	29.97
360 Player & Version	Oculuc Video , Version 2.5.23	Samsung VR, Version 1.71.7
Head Mounted Display	Samsung Gear VR, consumer version 1	Samsung Gear VR, consumer version 1
Resolution of Samsung s7	2560 x 1440 (Quad HD)	2560 x 1440 (Quad HD)
Resolution of Rendered Video	3840 X 1920	3840 X 3840
Video Render Format	264	265
Audio Headset	Philips SHB8850NC Bluetooth	Philips SHB8850NC Bluetooth

Chapter 5

Evaluation

The following chapter will seek to explain how the different implement solution was tested and the evaluate the results that was found. First part of this chapter will be on the initial test, first an explanation on the test scenario, followed by a presentation of the findings during the test. In the results we will present the results along with a evaluation of them. The second part of this chapter will contain the final test which is a bigger study with more condition. First there will be a presentation of test setup and method followed by a evaluation of the data gathered in the test.

5.1 Initial

The following will describe the test setting of the initial test. This is to shed light on how the the data for the initial test was gathered.

5.1.1 Test procedure

The test took place on Aalborg University Copenhagen Campus. The S.M.I.L.E lab in building A was used to conduct the test in, as it is a location without unnecessary noise.

The process during the test was that when a test participant started the test they would get a condition that was picked at random. For the randomization of the test we had written the different condition on pieces of paper. when we picked on we put it aside until the others also had been pick then put them all back so they could be picked. They then heard 3 minutes of the song and was sent to answer a questionnaire.

5.1.1.1 questionnaire

The questionnaire was divided into different section with questions, listed as follows:

- **Demographic**
  1. Age
  2. Gender
  3. How would you rate your desire to continue?
  4. Have you tried 360 Video in VR before?
  5. Do you consider yourself to easily get symptoms of motion sickness?
  6. Are you wearing glasses or contact lenses?
  7. Do you suffer from hearing disabilities?
- **Motion sickness**
  1. Did you experience motion sickness during the experiment?
  2. With regards to motion sickness, how would you rate the quality of the experience?
  3. Please elaborate or comment (optional)
- **Imagery**
  1. How would you rate the quality of of imagery?
  2. Did you sense spatiality in the imagery?
  3. To what degree did spatiality in the imagery contribute to the quality of your experience?
  4. Please elaborate or comment (optional)
- **Sound**
  1. How would you rate the quality of of sound?
  2. Did you sense spatiality in the sound?
  3. To what degree did spatiality in the sound contribute to the quality of your experience?
  4. Please elaborate or comment (optional)
- **Presence**
  1. To what degree would you rate the realism of the imagery?
  2. To what degree would you rate the realism of the sound?

3. To what degree did you feel immersed into the experience?
4. To what degree did your feel transported into the scenery/location in the experience?

- **Quality of Experience**
  1. Please rate the overall quality of the sound?
  2. Please rate the overall quality of the imagery?
  3. Please rate the overall quality of the experience?

The questions that asked the test participants to make a rank was based on a 5 point lickert scale. After the test participant had answered the questionnaire they would be asked to try all three condition. This was followed up by a small interview where we primarily wanted them to rank the three conditions.

5.1.2 Initial Test Results

The following are an overview and evaluation of the data gathered in the initial test. First we will look at the demographic that was involved in our test, this is sex distribution, age, if they have tried VR before and if they have any hearing and / or seeing disabilities. This will be followed by a presentation of the rest of the data in the questionnaire and some evaluation of that data, we will go through each each section of the questionnaire, ending with a evaluation between the sections. Lastly we will present the qualitative data that was gather in this initial test and do a comparison to the quantitative data gathered.

5.1.2.1 Evaluation Method

The following will present the reasoning for what has been done with the data gathered from our questionnaire. First considering the quantitative data we will start by looking at what participants that we have in our test. This will be followed by a table with the Mean opinion score to give an overview of the data gathered.

Followed by a statistical evaluation of the data. First to choose what statistical we have to use we will consider the data gathered. This will be considered as non-parametric as it is gathered using a questionnaire, this is because it is interval data as we can not make consider each rank on the scale to have the same rational difference [9]. A statistical test suitable when having more than one condition and having non-parametric data is a Kruskal-Wallis (this is a non-parametric equivalent of a ANOVA test) [9]. If a significant different i found one can use a box plot to see what condition is higher or one can use a repeated use of Mann-whitney test, however doing this will increase the risk of making a type 1 error, so making a bonferroni correction to lessen this risk will be used [9]. Lastly we have quality answers to consider here we have ask them to make a ranking of the three versions, and furthermore look for common answers and quantify them.

5.1.2.2 Demographic

The first part of the questionnaire use in this test was mostly focused on what participant we had in our test, this part can be used to as a indicator of how reliable the test is. As an example if we have very few that have tried VR before, an good experience might just be an wow effect on the user. In the test we have gathered answer from 27 different people (27 n). When finding these 27 test participant we were aware of the division between the sexes, in the sense that we tried to have an equal distribution of men and women, in the end we found 14 men and 13 women. The participant had a mean age of 24.0741 with a standard deviation of 2.6736, by this we have a young group of people. One can see in figure 5.1 that 1 / 3 of the test participant had not tried VR beforehand, this might have influenced the test because it is new to people. Only 7.4 % of the test participant had any hearing

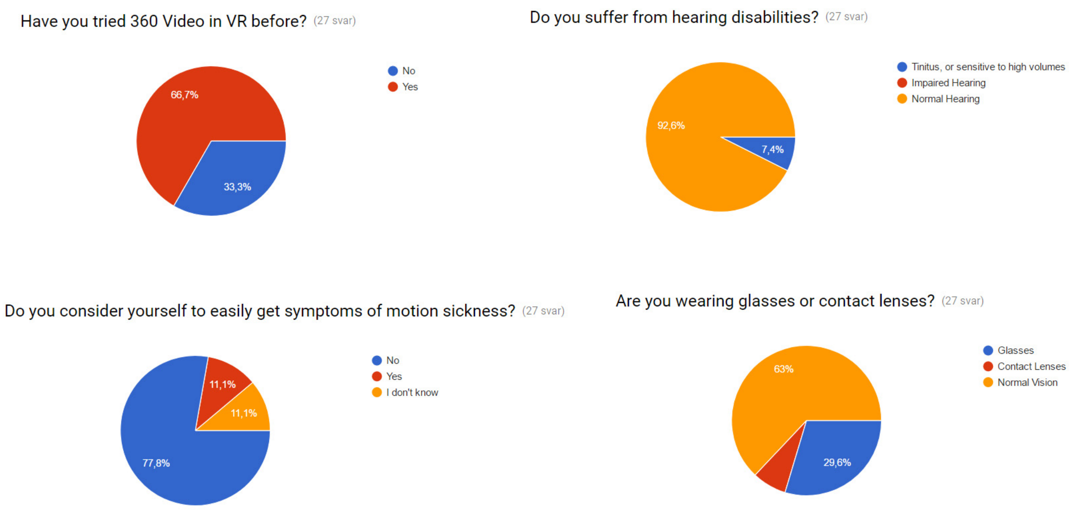

Figure 5.1: Pie chart of some of the demographic

disability (see fig 5.1), as it was tinnitus and sensitivity to high sound and only 7.4 % we do not consider this to have had a great effect on the answers. Furthermore we asked if the participant normally would have a tendency to get motion sick. Here most answer no, with only a few (11.1 %) saying yes and the same on don't know (see figure 5.1). Lastly in figure 5.1 we can see how many of the participants that had glasses/contact lenses and how many that had normal vision. Most had a regular vision but around 1 / 3 used glasses, this might have a negative effect on the results at it can be more cumbersome to have the headset on with glasses.

5.1.2.3 Mean Opinion Score (MOS)

The following will look at the mean opinion score that can be seen in table 5.1, this part is mostly to get an initial overview of the data collected. look at the table one can see that it is synthesized ambisonic and mono that have the highest means on average, however it is only in the first question that we find a big difference and overall it is only small differences.

Table 5.1: Here are presented the Mean Opinion Score (MOS) over the three condition

Question	Zoom	synthesized ambi	mono
How would you rate your desire to continue?	3.111	3.7778	4.111
Motion sickness			
With regards to motion sickness, how would you rate the quality of the experience?	4,333	4	4.222
Imagery			
How would you rate the quality of of imagery?	2.222	2.6667	2.8889
To what degree did spatiality in the imagery contribute to the quality of your experience?	3.2222	3.3333	3.5559
Sound			
How would you rate the quality of of sound?	4	3.7778	4.111
To what degree did spatiality in the sound contribute to the quality of your experience?	3	3.7778	3.8889
Presence			
To what degree would you rate the realism of the imagery?	3.4444	4.1111	3.6667
To what degree would you rate the realism of the sound?	3.777	4.1111	4.2222
To what degree did you feel immersed into the experience?	3.222	4	3.444
To what degree did your feel transported into the scenery/location in the experience?	2.7778	3.7778	3.5556
Quality of Experience			
Please rate the overall quality of the sound?	3.778	3.6667	3.8889
Please rate the overall quality of the imagery?	2.6667	2.8889	3.3333
Please rate the overall quality of the experience?	3.4444	3.8889	3.7778

5.1.2.4 Continuation Desire

In the first section we asked if the participant would like to continue the experience, as we intentionally stopped the experience after 3 minutes. Here we find a slight positive mean at 3.6667 over all three conditions with B (synthesized ambisonic at 3.7778) and C (mono at 4.1111), Zoom H2N recording had a mean at 3.1111 (for an overview of the score see boxplot in figure 5.2). Testing for a significant difference with our non-parametric test we get a p-value at 0.0529, so no significant difference even though we are close at significant value at p 0.05. As we are close to an significant value it can make sense to make more evaluation with the different

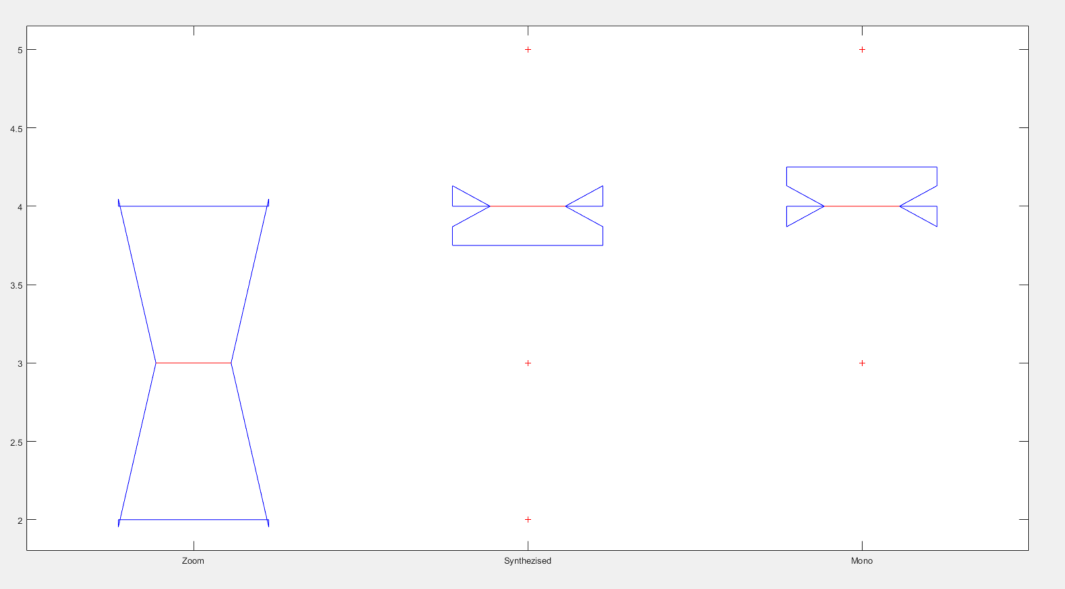

Figure 5.2: A box plot of the question "How would you rate your desire to continue?"

condition paired, with a Mann-Whitney test to see where the potential difference is. Here we are going to make a bonferroni correction because each time we test there are a 5 % chance of making a Type 1 error. As a bonferroni correction can quickly make a critical value very low we will test the Zoom H2N up against the two other version, as Zoom H2N are the only one that are not mixed afterwards. Doing this the critical value end up at 0.025, with the pairs:

- 1 Zoom H2N / synthezised ambi
- 2 Zoom H2N / mono

Here we find that there are the biggest difference between the Zoom and the mono at a p value of 0.0346 (so significant without the bonferonni correction), with the other comparison at a p value of 0.1794. In the end we found no significant difference between the 3 conditions.

5.1.2.5 Motion Sickness

The next section are on motion sickness. In this section we only ask into of any felt motion sickness had an effect on the perceived quality of the experience. Looking back a table 5.1 we can see that for this question "With regards to motion sickness, how would you rate the quality of the experience?" we have mean above 4 on all three conditions. This tells us that the potential motion sickness that the participants have felt did not have a negative effect on the experience. Also worth notion is that on 14.8 % of the participant felt any motion sickness (see figure 5.3).

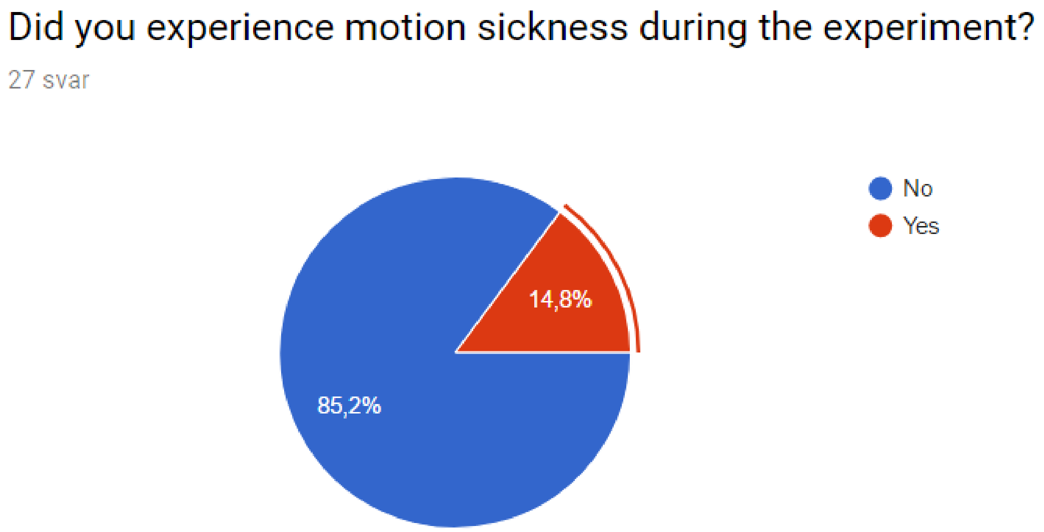

Figure 5.3: Here are a pie plot for if people reported the fealing of motion sickness

Looking for a statistical significant difference using the kruskal-wallis test we find no significant difference at a p value of 0.7384. This is as expected from looking at the mean score in table 5.1 where they are almost the same.

Lastly from looking at the elaboration question in the end of this section we find 8 that said that they had no motion sickness and 3 that had. This does not completely match was we can see in figure 5.3 but this is properly because only 11 out of 27 answered on this question.

5.1.2.6 Imagery

Here we will evaluate the data from the Imagery part of the questionnaire. This part of the questionnaire contains two questions that ask about the quality of the imagery and how the participant felt that the imagery had spatiality. For imagery we had two questions "How would you rate the quality of imagery?" (with mean score just below average) and "To what degree did spatiality in the imagery contribute to the quality of your experience?" (with a mean score just above average).

Looking for statistical difference within each question we find a p-value at 0.2527 for "How would you rate the quality of imagery?" and p-value at 0.8058 for "To what degree did spatiality in the imagery contribute to the quality of your experience?". Based on this we can not report any significant between the different experience, which makes sense because the variable are on the sound and the imagery on both are the same.

5.1.2.7 Sound

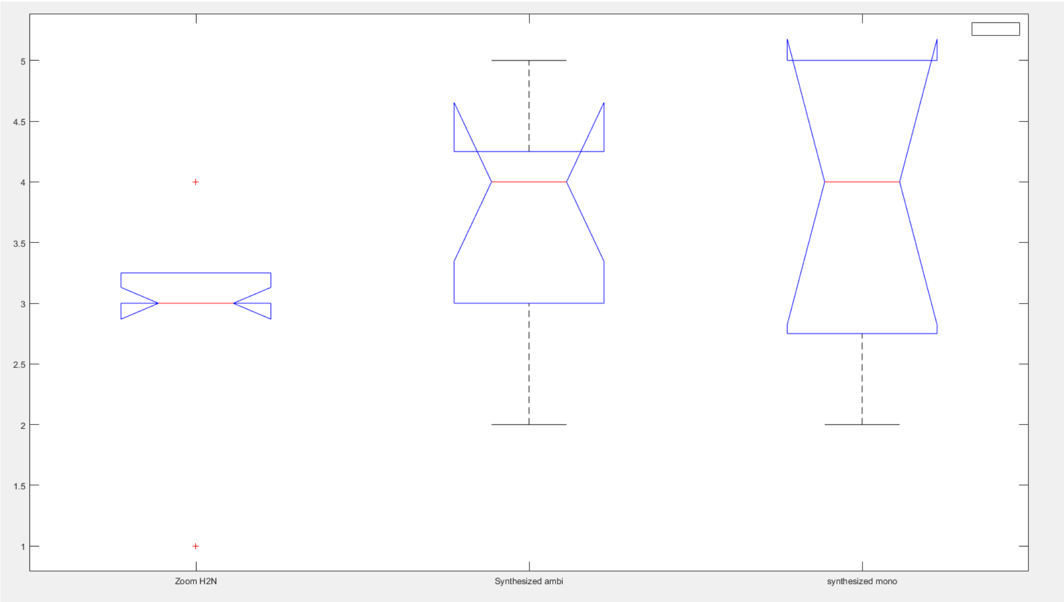

Figure 5.4: Here are a boxplot of the question "To what degree did spatiality in the sound contribute to the quality of your experience?"

This section will evaluate the section in the questionnaire about that ask into sound specifically. Here we like in the imagery section ask into the quality and how they felt the spatiality of the sound added to the quality. Here the mean score of the quality question is 3.9630 over all the conditions and the spatiality question have a mean score of 3.5556 over all the condition (for the individual mean score see table 5.1.

Testing for significant different for the quality question we get a p value at 0.5922 and with that we can not report a significant difference for the first question about quality.

Secondly test for significant difference for the spatiality in the audio we find a p value at 0.1523 and as such the 3 different condition are more different that the quality question but still not at a significant level at 0.5. Looking at figure 5.4 we can see that the difference in the distribution does not have a big difference, however that the two synthesized are group similar. However we will not report any significant difference on the two question in this chapter.

5.1.2.8 Presence

The following is a evaluation of the section in the initial questionnaire about different Presence variables. In this section we ask about realism (in sound and video as in 2 qustions), immersion and transportation, in total 4 questions.
looking at just the mean values from the section (looking a table 5.1 we find tat the sound at seam a bit more real for the zoom and mono condition but the same for the synthesized ambisonic and at a higher rank. further more if we look at immersion mean scores we see above mean at them all with the synthesized getting the highest rank. Lastly looking at transportation we again find that the synthesized audio getting the highest rank, with the audio from the Zoom H2n being the lowest.

Table 5.2: A table of the question in the Presence section of the questionnaire and it's p-value, to right in the table there are 0 because non of the questions had a significant different between the conditions

Question	P-value	Significance
To what degree would you rate the realism of the imagery?	0.3383	0
To what degree would you rate the realism of the sound?	0.4669	0
To what degree did you feel immersed into the experience?	0.2813	0
To what degree did your feel transported into the scenery/location in the experience?	0.0965	0

The statistical test for significance show that non of them had a significant difference, for an overview of the specific p-values see table 5.2.

5.1.2.9 Overall Quality of Experience

This section will evaluate the last section of the initial questionnaire about the overall felling of quality of the experience. Here we have three questions one on the quality of sound, on for the imagery and the last for the overall experience.

Looking at the mean score for this section we can see that the quality of the imagery is performing worse that the sound but had only had a small effect on the overall averages. Further more it is interesting to note that the mono sound condition has an higher mean than the spatial audio condition (see table 5.1).

Making a non-parametric statistical test we find that non of the condition is significantly different (see table 5.3).

5.1.2.10 Qualitative Evaluation

This part will take a look on the ranks that was given during the interview at the end of the test.

Table 5.3: Here are a table for the P-values found when testing the Quality of Experience part of the questionnaire

Question	P-value	Significance
Please rate the overall quality of the sound?	0.8961	0
Please rate the overall quality of the imagery?	0.2601	0
Please rate the overall quality of the experience?	0.4376	0

The ranking that was given in the recorded interview was 5 for A, 5 for B and 3 for C. In general people had a tendency to like that there was a spatiality to the sound with only a few preferring that the sound was mono.

5.1.2.11 Sum up of initial test

as we have rather small sample size the effect of this test is rather small. However we find in the qualitative test that most people like the two different ambisonic sound. For the statistical test we find no significant difference between the 3 conditions and as such we can say that people do not dislike added spatiality.

5.2 Final Test

The following chapter will describe the final test that was performed. this test aim to both evaluate the imagery in two different setting and the sound of a 360 movie seen with a Head Mounted Display (HMD). For the imagery there are prepared two different versions, one with a mono image which are the most common in 360 movie and a 3d / stereoscopic version which mimic the real world more with two slightly different images for each eye. For the sound we have a stereo sound and ambisonic sound, one of each on both kind of imagery. This ended us up with 4 different conditions

- Video with monoscopy and stereo sound, condition 1
- Video with monoscopy and ambisonic, condition 2
- Video with stereoscopy and stereo sound, condition 3
- Video with stereoscopy and ambisonic sound, condition 4

The following section will first describe the test setup, along with the questionnaire and procedure. This section will be followed up by a description of the evaluation method ending with the evaluation.

5.2.1 Test Setup

The following will describe the test procedure followed by an explanation of what the test participants was asked during the test. Ending with an explanation of how the following evaluation of the test will be performed.

5.2.1.1 Test procedure

The test was performed in two different rooms. One room for the test participant to try one of the 4 condition, this room had to be relative silent as sound from outside source could bias the results. When they came to the room they got a short introduction to the Head Mounted Display, afterward they got an random condition and where asked to choose and see that condition on the HMD.
The randomization was done in a structure random way where we had 4 paper with the different condition on them, one was chosen and removed until all the different condition had been removed, they where then added back.
After they had tried the Virtual Reality experience they where moved to another room where they was aske to answer a questionnaire.

5.2.1.2 Questionnaire for final test

The questionnaire was divided into 4 different section with the different question under each section, each ranked question was ranked on a 7 point scale:
The first section is to get an idea of what people we have tested on. And if the test participant had disabilities that could bias the test.
The second section of the questionnaire concern the imagery in the experience, here we ask into if the participant felt depth in the image, along with questions about Presence variables, ending with different quality variables.
This section was followed by a section that asked into the same variables but for the sound. In the motion sickness section we asked but one question if they felt motion sick.

5.2.2 Test Method

The following section will describe how the test result are to be evaluated. The following evaluation will start out by looking at the demographic to see what the sample that we have gathered contains. This is followed by a Mean Opinion Score table to get an overview of what condition performed best based on the mean score. Following this we will use a statistical test to see if any of the differences in the scores have a significant difference to each other.

Table 5.4: Here are a table that show the questions in the questionnaire, also one can see by the Qn how each question will be referenced henceforth

Demographic
Age
Gender
Please rate your level of experience with VR
Visual Impairment
Hearing Impairment
Do you you consider yourself to easily get symptoms of motion sickness?
Imagery
Q1 : To what degree did you sense depth in the imagery?
Q2 : To what degree would you rate the realism of the imagery?
Q3 : To what degree did you have a feeling of being transported into the visual location of the experience?
Q4 : To what degree would you rate the sharpness of the imagery?
Q5 : To what degree would you rate the level of noise in the imagery?
Q6 : To what degree was the imagery stressful for your eyes?
Q7 : Please rate the quality of the imagery?
Sound
Q8 : To what degree did you sense direction in the sound?
Q9 : To what degree would you rate the realism of the sound?
Q10 : To what degree did you have a feeling of being transported into the auditorial(sound) space of the experience?
Q11 : To what degree would you rate the clearness of the sound?
Q12 : To what degree would you rate the loudness of the sound?
Q13 : To what degree would you rate the level of noise in the sound?
Q14 : To what degree was the sound stressful for your ears?
Q15 : Please rate the quality of the sound?
Motion sickness
Q16 : With regards to motion sickness, how would you rate the quality of the experience?

5.2.2.1 Statistical Test

To determine what statistical test we would like to use we have to consider what data we have collected. In our case because we have self report of the experience we have ordinal data, as we do not know of the ratio that people connect between the different score [9]. because of that we have to look for a non-parametric statistical test the difference between each conditions. as we have more than 4 different conditions we have to make a non-parametric equivalent of a ANOVA test, in this case where we have independent measures we will use the Kruskal-wallis test [9]. Furthermore when we find a significant differences we only know that there are a significant difference and not where it is, so for a post hoc test we will make repeated man-whitney test to find where the difference is [9]. Here we need to make a bonferroni correction as the type 1 error can build up each time we make the test.

5.2.3 Test Results

The following section is the presentation of the data and the statistical evaluation on the data. This is so we can make a conclusion on our research question in the conclusion chapter. First of we are going to look at the sample we have gathered and if there are any big bias that the test participant could have, like if halve of the sample are wearing glasses this could have an effect on the results as it can be cumbersome to have both glasses and the HMD on at the same time. Following this we are going to present the MOS table where we can see all the means of all condition and all questions. This will give us an overview and we can roughly see which condition performed better than the other. After this we will continue with a statistical evaluation of each question, see section 5.2.2.1 for what statistical test that will be use. Few comment will be made in each section and the results of the test will be further discussed in the next chapter.

5.2.3.1 Demographic

In total we gathered answers from 120 participant (120 n) divided onto all 4 condition equally. in all 30 participant in each conditions. The mean age of the participants was 25.6917 with the mean within each condition being around this number at 25.2667 on the lowest mean within the condition and 26.7 at the highest mean. By this we can see that we have a somewhat young sample of people. The division in the gender was 1/3 female (40) and 2/3 male (80).

For the visual impairment among the participant we find that 43 of the participant had some impairment (16 did wear contact lenses and 27 did wear glasses), so this might have effect the test with some bias. If we count in each condition how many had visual impairment in each condition we find that they have been each distributed with:

- Condition 1 : 12
- Condition 2 : 13
- Condition 3 : 13
- Condition 4 : 12

So the effect of the bias that this would have will effect each condition equally.
For hearing impairment we find that only 6 persons have reported any and that only 2 was sever enough to wear a hear aid, and 4 said other. So the bias from hearing impairment is minimal.

Lastly in the Demographic section we asked the participant what their experience with Virtual Reality was. Here we find different means at:

- Condition 1 : 3.5
- Condition 2 : 4

- Condition 3 : 3.4667
- Condition 4 : 4.3667

over all the condition we find that the test participant on average had an average experience with Virtual Reality. Looking for is there are a significant difference between each condition using Kruskal-Wallis test, we find a p-value at 0.222 and those each condition do not differ significantly in their experience with Virtual Reality. And because of that the different condition have not been significantly differently bias by more or less experience with virtual reality.

5.2.3.2 Mean Opinion Score (MOS)

The following will talk about the mean scores that we find in table 5.5. These mean score are a expression of the quality and the amount of presence that the test participant felt during the 4 different conditions.

Table 5.5: Mean Opinion Score of all the question in the final questionnaire, Qn represent the 16 different questions in the questionnaire, see question in section 5.2.1.2, The Means summed Quality score are calculated from the question in each section except for Q2,Q3,Q9 and Q10 that asks about Presence variables and are used for the Means summed presence. The acceptability score in the bottom of the table are the weighted means of the means summed and motion sickness, and as such this is a expression of the overall quality of the experience.

Questions	Monscopi stereo sound	Monoscopi ambisonic	Stereoscopi stereo sound	Stereoscopi ambisonic
Imagery				
Q1	5,4333	4,7667	4,6667	4,9333
Q2	4,7	4,5333	4,7333	4,8
Q3	5,0333	4,8333	5,2667	4,9
Q4	2,8333	2,6667	2.8333	2,5333
Q5	4,4333	3,7667	4	4.0333
Q6	5	4,7333	4,4333	4,9667
Q7	3,6	3,1	3,8	3,7
Means summed Quality	4,26	3,8067	3,9467	4.0333
Means summed Presence	4,8666	4,6833	5	4.8500
Sound				
Q8	4,9333	5,5	4,2667	4,9
Q9	5,8667	5,5	5,4667	5,5
Q10	5,5667	5,1333	4,8333	5,3
Q11	5,6667	5,5333	5,0667	5,4333
Q12	5,6667	5,6333	5,4	5,6667
Q13	5,5	5,3	4,9667	5,3
Q14	5,6	5,8667	5,2	6,2333
Q15	5,8	5,5	5,2667	5,6667
Means summed Quality	5,5278	5,5556	5,0278	5,5333
Means summed Presence	5,7167	5,3167	5,15	5.4
Motion Sickness				
Q16	5,7667	5,5333	6,0333	6,3333
Total Presence Mean	5,2917	5	5,075	5,125
Acceptability Score	5,2276	4,9791	5,0316	5,23

A quick glance at the Imagery section we can see that the monoscopic image with stereo sound in general have slightly better scores closely followed by the stereoscopic versions that has almost the same scores, with the monoscopic image with ambisonic sounds on average performed worse. We can also see that the image has given more presence than quality, and that it does not seam like Presence and Quality follows each other.

If we further look at the table on the sound section we find much the same tendency with monoscopy with stereo sound (condition 1) highest and and stereoscopy with ambisonic following just after it. Here on the other hand it is monoscopy with ambisonic sound that follows closely and stereoscopy with stereo sound that are lacking far behind. If one looke that the means summed in this section we see that the values follows each other more here than the Imagery section.

Next in the Motion Sickness section we find that the stereoscopic condition have the best grade. This shows us that people have felt less motion sickness in these condition.

Lastly if we look at the Acceptability score (which are the weighted means) we find that Condition 1 and Condition 4 have the best score following each other closely, and condition 2 & 3 also are similar.

5.2.3.3 Statistical Test

The following is the statistical test that see if we have found any significant differences between the 4 conditions, for the method use see section 5.2.2.1. Here are the list with the resuls of the kruskal-wallis test that was performed:

- Q1 p value : 0,1012, h : 0
- Q2 p value : 0,8313, h : 0
- Q3 p value : 0,5098, h : 0
- Q4 p value : 0,7830, h : 0
- Q5 p value : 0,5266, h : 0
- Q6 p value : 0,6445, h : 0
- Q7 p value : 0,1479, h : 0
- Q8 p value : 0,0243, h : 1
- Q9 p value : 0,6904, h : 0
- Q10 p value : 0,1567, h : 0

- Q11 p value : 0,2220, h : 0
- Q12 p value : 0,7924, h : 0
- Q13 p value : 0,3438, h : 0
- Q14 p value : 0,1342, h : 0
- Q15 p value : 0,5477, h : 0
- Q16 p value : 0,1385, h : 0

Looking at the list of p-values we only found a significant difference in one question, Q8 "To what degree did you sense direction in the sound?". First we can look at the box plot of the questions to get a deeper look into the data, both where the median are and the general spread of the data. In figure 5.5 we see this plot, here it seams as the biggest difference are between monoscopy and ambisonic versus the other as it distribution are more gathered in the high end.

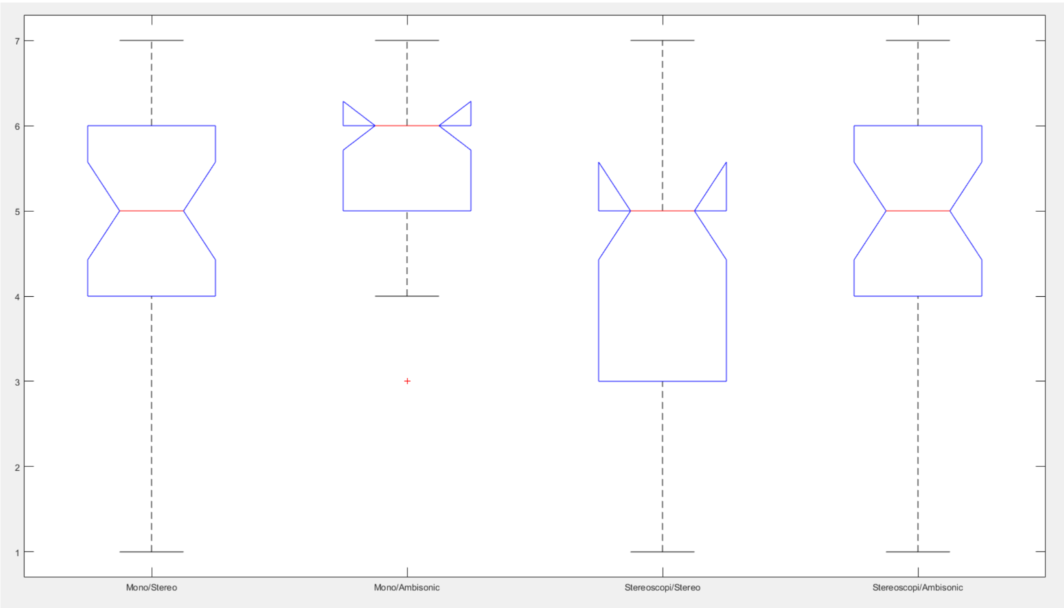

Figure 5.5: a boxplot of the only question with a significant difference Q8

However we can not say anything for sure by just looking at these boxplot. So to find where the significant difference are, we will perform the post hoc test described in section 5.2.2.1. For doing this test we have to make a bonferroni correction. The coreection take the critical value that we have 0.05 and divided it by the amount of test that we are going to perform. For the amount of test needed we see table 5.6, which givs us 6 test to perform. With the amount of test determined we can make the calculation.

$$CV = 0.05/6 = 0.0083$$

Table 5.6: A table explaining how we are going to make the Mann-Whitney test

Mann-Whitney test matrix	monoscopy Stereo sound	monoscopy Ambisonic	stereoscopyc Stereo sound	stereoscopyc Ambisonic
monoscopy Stereo sound	-	Q8MW1	Q8MW2	Q8MW3
monoscopy Ambisonic	-	-	Q8MW4	Q8MW5
stereoscopyc Stereo sound	-	-	-	Q8MW6
stereoscopyc Ambisonic	-	-	-	-

With the critical value found at 0.0083 we can continue to make our post hoc test, and the results are listed here:

- Q8MW1 p-value : 0,1601
- Q8MW2 p-value : 0,1078
- Q8MW3 p-value : 0,9756
- Q8MW4 p-value : 0,0026
- Q8MW5 p-value : 0,1277
- Q8MW6 p-value : 0,1093

Using our new critical value and looking over the table we find that the 4 comparison have a significant difference as the only one, this test is between condition 2 and 3. Further more we can calculate the effect size of this we have a z-value for that comparison at 3.0107 and that will give us.

$$Effectsize = 3.0107/\sqrt{(60)}$$

This will equal to 0.3887 which represent a medium effect.

5.2.3.4 Sum up of final test

In the end we can report that we find a significant different in the perceived directional in sound. This significant different is between condition 2 and condition 3, with condition 2 having the highest score. This make sense as condition 2 have ambisonic and condition 3 have stereo sound.

Looking at the MOS table we find that the highest scores on the sound and images in average are for condition 1 however not by much. look at the motion sickness part we find that people actually felt less motion sickness using the stereoscopic version.

Chapter 6

Discussion

This chapter will be used to make the final comments on what have been done and what has been accomplished during this project.

In the initial test we had a somewhat messy approach to the questionnaire. We were messy in our questionnaire for our initial test since we have imagery and sound questions in different sections, and since in the the last questions we are asking for a acceptability score, which would have been better to calculate from the other section. In our statistical test for significance, did we not find any significant difference between the condition in any of the questions. One of the reason why we do not find any significant different could be because of the small sample size we have from the test. With only 27 persons and 3 different groups can we not conclude much for the general population based on those numbers. However to make up for this we had a listening test in the end of the test where people had the chance to rank the 3 version. From this we did find that many people liked the spatial aspect to the sound.

In the final test we look at difference for sound and imagery spatiality. Because we both have changes in the imagery and the sound this test tells us more about our research question. From this the final test we only found one significant difference in Q8 "To what degree did you sense direction in the sound?". When we investigated this difference we found that the specific difference was between condition 2 and 3, where condition 2 had the higher scores. From this we can say that people did sense the spatiality in the sound and they found the sound less directional when we had stereo sound with depth in the image. One could speculate that the depth in the image lessen the effect of the limited direction in stereo sound. These two are also the condition with the lowest acceptability scores. This could be because a flat image with non interactive sound seem more natural than a flat imagery with interactive sound, looking at the presence score between these two could also indicate this as condition 2 have lower scores than condition 1, see figure 5.5. This also goes together with that we find that a fully spatial version (condition 4 stereoscopy and ambisonic) rates better in acceptance score than the stereoscopic video with stereo sound. This acceptability score show a small preference towards that a flat 360 video with stereo sound and the 3D video with ambisonic. This might be because when people have spatiality in sound but not imagery they might feel a disconnection, and the same when they have a 3d video they want the sound to be more interactive as well.

During this study we have proposed method for evaluating Quality and presence in 360 film. Here we will highlight the method used in the final test, as this is the final iteration on the method. Here we have presented a questionnaire that base on Imagery, sound, presence and motion sickness produce a acceptability score, that are a measure of quality of the experience.

Chapter 7

Conclusion

During this study we have explored the concepts of Quality of Experience and presence, this have lead to a proposed method to evaluate these concepts in 360 films.

This study show us that there are no significant difference on presence or quality if a user tries a stereoscopic 360 video or a monoscopic 360 video. However we found that people sense significant more direction in the sound when people get a monoscopic image with ambisonic compared to a stereoscopic image with stereo sound, however they do not feel significant more direction in the sound when the imagery are in monoscopy with stereo sound or when the imagery are stereoscopic with ambisonic. Regarding motion sickness we find that the two stereoscopic versions have a higher acceptance, however not a significant higher score.

Spatial feature do not contribute significantly to a quality of experience. However we see a small improvement on the acceptance score when we have monoscopic video with stereo sound and with stereoscopic video with ambisonic sound.

Bibliography

[1] Kjell Brunnström et al. "Qualinet white paper on definitions of quality of experience". In: (2013).

[2] Stacy A Clemes and Peter A Howarth. "The menstrual cycle and susceptibility to virtual simulation sickness". In: *Journal of biological rhythms* 20.1 (2005), pp. 71–82.

[3] HTC Corporation. *RECOMMENDED COMPUTER SPECS*. June 17, 1. URL: https://www.vive.com/ca/ready/.

[4] Alex Davies. *Oculus Rift Vs. HTC Vive Vs. PlayStation VR*. June 17, 1. URL: http://www.tomshardware.co.uk/vive-rift-playstation-vr-comparison,review-33556-3.html.

[5] Chenwei Deng et al. *Visual Signal Quality Assessment*. Springer, 2015.

[6] Malham D. G. "Higher order Ambisonics systems for spatialization of sound". In: *ICMC Proceeding 1999* 1999 (1999).

[7] Eric L Groen and Jelte E Bos. "Simulator sickness depends on frequency of the simulator motion mismatch: An observation". In: *Presence* 17.6 (2008), pp. 584–593.

[8] Varun Nair Hans Fugal. *Spatial audio — bringing realistic sound to 360 video*. May 17, 28. URL: https://code.facebook.com/posts/412047759146896/spatial-audio-bringing-realistic-sound-to-360-video/.

[9] Andy Field & Graham Hole. *How to Design and Report Experiments*. An optional note. SAGE, 2003. ISBN: 978 0 7619 7382 9.

[10] Danish Film Institute. *A Heightened Sense of Presence*. URL: http://www.dfi.dk/Service/English/News-and-publications/News/May-2017/Separate-Silences-to-Cannes.aspx.

[11] Chompoonuch Jinjakam and Kazuhiko Hamamoto. "Study on parallax affect on simulator sickness in one-screen and three-screen immersive virtual environment". In: *Proceedings of The School of Information and Telecommunication Engineering Tokai University* 4 (2011), pp. 34–39.

[12] David M Johnson. *Introduction to and review of simulator sickness research*. Tech. rep. DTIC Document, 2005.

[13] Robert Samuel Kennedy and Lawernce H Frank. *A review of motion sickness with special reference to simulator sickness*. Tech. rep. DTIC Document, 1985.

[14] Eugenia M Kolasinski. *Simulator Sickness in Virtual Environments.* Tech. rep. DTIC Document, 1995.

[15] JJ-W Lin et al. "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment". In: *Virtual Reality, 2002. Proceedings. IEEE*. IEEE. 2002, pp. 164–171.

[16] Rasmus B Lind et al. "Sound design in virtual reality concert experiences using a wave field synthesis approach". In: *Virtual Reality (VR), 2017 IEEE*. IEEE. 2017, pp. 363–364.

[17] Rasmus Bloustrød Lind, Dina Madsen Smed, and Victor Milesen. "Internship at Panorama Video". 3rd Semester Internship Report, Medialogy Master Study, 2016. Jan. 2017.

[18] Rasmus Bloustrød Lind et al. "Sound Design in Virtual Reality Concert Experiences with Wave Field Synthesis". 2nd Semester Report, Medialogy Master Study, 2016. Dec. 2015.

[19] Matthew Lombard and Theresa Ditton. "At the heart of it all: The concept of presence". In: *Journal of Computer-Mediated Communication* 3.2 (1997), pp. 0–0.

[20] Inc Mc Squared System Design Group. *Sound Systems: Mono versus Stereo*. May 17, 25. URL: `http://www.mcsquared.com/mono-stereo.htm`.

[21] Victor Milesen et al. "Camera Positions and Presence in VR Concert Experiences". Bachelor Thesis, 2015, Medialogy, Aalborg University Copenhagen. May 2015.

[22] Victor Milesen et al. "The Effect of Spatial Sound on Presence in a Shared VR Concert Experience". 1st Semester Report, Medialogy Master Study, 2015. Dec. 2015.

[23] Sebastian Möller and Alexander Raake. *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*. Springer, 2014.

[24] Sebastian Möller and Alexander Raake. *Quality of Experience - Advanced Concepts, Applications and Methods*. Springer, 2014.

[25] LLC. Oculus VR. *Recommended PC Specification*. June 17, 1. URL: `https://www.oculus.com/rift/`.

[26] Donald E Parker and Deborah L Harm. "Mental rotation: a key to mitigation of motion sickness in the virtual environment?" In: *Presence: Teleoperators & Virtual Environments* 1.3 (1992), pp. 329–333.

[27] Robert Earl Patterson. *Human Factors of Stereoscopic 3D Displays*. Springer, 2015.

[28] Michael Pearce. *Gear360Tools*. URL: http://www.dropbox.com/sh/0iopn0v63bcuece/AADV06864yYNN1KvMEvjbwwWa?dl=0 (visited on 06/01/2017).

[29] Rui Penha. "Distance encoding in ambisonics using three angular coordinates". In: (May 17, 8), pp. 1–3.

[30] Roy A Ruddle. "The effect of environment characteristics and user interaction on levels of virtual environment sickness". In: *Virtual Reality, 2004. Proceedings. IEEE*. IEEE. 2004, pp. 141–285.

[31] Samsung. *Compatible Devices*. June 17, 1. URL: http://www.samsung.com/us/mobile/virtual-reality/gear-vr/gear-vr-with-controller-sm-r324nzaaxar/#specs.

[32] SAMSUNG. *Gear VR (2015)*. May 17, 31. URL: http://www.samsung.com/uk/wearables/gear-vr-r322/.

[33] SAMSUNG. *Specifikationer*. May 17, 31. URL: http://www.samsung.com/dk/smartphones/galaxy-s7/more/#!/spec.

[34] Henrik Schoenau-Fog. "The Player Engagement Process – An Exploration of Continuation Desire in Digital Games". In: *DiGRA 2011 Conference: Think Design Play* (2011).

[35] Sennheiser. *AMBEO® VR MIC*. June 17, 2. URL: https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic.

[36] KC Uliano et al. *The effects of asynchronous visual delays on simulator flight performance and the development of simulator sickness symptomatology*. Tech. rep. DTIC Document, 1986.

[37] Signe Ungermand. "Separate Silences". Bachelor Thesis. Jan. 2016.

[38] Zoom. *Zoom H2n*. June 17, 2. URL: https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h2n-handy-recorder.