MASTER THESIS

Defining the sound experience in Social VR

by Vincent Poppelaars

DESIGN FOR INTERACTION
FACULTY OF INDUSTRIAL DESIGN ENGINEERING

TU Delft

CHAIR: RENÉ VAN EGMOND
SUPERVISOR: AADJAN VAN DER HELM

GLOSSARY

Anechoic	Producing no echoes; very low or no reverberation.
Attenuation	A loss of energy; in acoustics, typically a reduction in volume.
Convolution reverb	Convolution reverb samples the impulse response from a real world environment such as a stadium, lecture hall or museum.
Direct Sound	Sound that has travelled directly to the listener without reflecting (versus reverberant sound).
DOF	Degrees of freedom. 3DOF is only looking around, 6DOF is also moving around.
Early Reflections	Reflected sounds that arrive relatively soon at a listener's location (i.e., before Late Reflections).
Head-Related Impulse Response (HRIR)	A formal characterization of the effect of sound interacting with the geometry of a particular human body. Used to create head-related transfer functions.
Head-Related Transfer Function (HRTF)	A transformation of an acoustic signal using a head-related impulse response. Used to simulate the effects of interaction of a sound originating from a specific direction with the geometry of a particular human body.
Head Shadowing	The attenuation of sound caused by the head lying between an ear and the sound source.
Initial Time Delay	The interval between the arrival of a direct sound and its first reflection.
Interaural Level Difference (ILD)	The difference in a sound's level or volume between the two ears.
Interaural Time Difference (ITD)	The length of the interval between when a sound arrives at the first ear and when it arrives at the second ear.
Impulse Response	An impulse response is the reaction of any dynamic system in response to some external change (impulse) as a function of time.
Late Reflections	Reflected sounds that arrive relatively late at a listener's location (i.e., after early reflections).
Motion Parallax	When moving objects are farther from a perceiver, their apparent speed of travel decreases; for example, a moving airplane on the horizon appears to be traveling more slowly than a nearby car. The apparent rate of travel of an object can therefore be used as a distance cue.
Obstruction	Obstruction is in an environment where a wall separates the listener from the source, that the sound reaches the listener through the wall and around it with the reflections.

Occlusion	Occlusion is when a wall separates the listener from the source and that the direct path goes through the wall and the reflections also go through the wall. It will sound differently than if the wall was not there.
Pinnae	The visible portion of the ear that lies outside the head.
Reverberant Sound	Sound that has reflected or reverberated before arriving at a listener's location (versus direct sound).
Reverberation	The reflection of sound off of a surface, or the temporary persistence of sound in a space caused by reverberation.
Room Impulse Response	Room impulse response (RIR) is the transfer function between the sound source and the microphone. It basically records what makes it different to another physical location. This can be applied to other signals.
Screen-door effect	It is the visual artifact of displays where the lines between pixels or subpixels are visible.
Sidetone	In the telecommunication field, sidetone is the effect of hearing an audible feedback of the sound picked up by the telephone's transmitter.
Sound Localization	1. The process of determining the location of a sound's origin; or 2. the suggestion of an object's location based on the manipulation of auditory cues.
Sound Spatialization	The representation of a sound within three-dimensional space.

EXECUTIVE SUMMARY

Virtual Reality (VR) grew quickly in recent years, for professionals and consumers. Through this significant progression, companies and content creators have mostly focused on the visual side of VR. However, sound has the ability to improve the immersivity and perceived quality of a visual display and virtual simulation in general. Social VR is a way of communicating with other people through VR by integrating participants in the same environment and enabling them to interact with each other. In Social VR, it is especially important to be immersed.

The purpose of this project was to evaluate current implementations of the sound experience in Social VR and to improve that experience through user testing.

The project was divided in three phases:
- Phase I: Theory & explorations. Research to understand the context
- Phase II: Basic Social VR prototype and user testing to find user needs
- Phase III: Advanced Social VR prototype to evaluate the sound experience based on previous research.

During the first phase, literature research was done, and two questionnaires were made for users and non-users of Social VR.
The main findings were that there are four crucial effects in 3D audio spatialization to tell where sounds come from:
- Head Related Transfer Function (HRTF) is used to simulate the effects of the interaction of a sound originating from a specific direction with the geometry of a particular human body.
- Interaural Time Difference (ITD): The length of the interval between when a sound arrives at the first ear and when it arrives at the second ear.
- Interaural Level Difference (ILD): The difference in a sound's level or volume between the two ears.
- Pinnae effects: Effects caused by the shape of your ear.

During explorations, it turned out people using noise-cancelling headphones in Social VR were speaking much louder than normally in order to hear themselves speak through the headphones. This can be avoided with sidetone. In the telecommunication field, sidetone is the effect of hearing an audible feedback of the sound picked up by the telephone's transmitter.
Next to that, users expect the experience to be immersing, relaxing, and give the sense of togetherness. Most of all, they want it to feel like in real life.

In the second phase, several other aspects became important: presence to judge the level of immersion, and co-presence to judge the level of interaction between users. Presence is how much the user feels "present" in the virtual environment while co-presence is how much the user feels both users are there. A prototype was made for a basic Social VR experience. The spatialization was done with Resonance Audio and it was made for WebVR.
During the testing, users indeed found realism to be better. If there was a noticeable start/stop of a voice because of the background noise turning on/off it took the user out of the immersion. Users expected more interactivity. Localization of sound sources was acceptable for left/right and down/up directions with Resonance Audio.
The biggest conclusions on the prototype itself were that visuals and real-time interactions are necessary to properly evaluate the sound experience. Sidetone should be tested. And finally, adding a digital reverberation on the sound coming from a TV in VR should be avoided.

Therefore in the third phase, a more advanced prototype was made in which two users could speak with each other, watch TV, and interact with buttons and each other with their avatar. The sound quality was also improved by using lavalier microphones.
This time users had sidetone which improved their presence, the majority preferred to have it turned on. Users were able to tell that the reverberation of the room did not match their expectations based on the room. The reverberation model of Resonance Audio was not accurate enough. They would prefer it to be more realistic which was somewhere between the low and medium reverb settings available in the test.
Hearing both voices with reverberation made it feel more like they were in the same room as the voice. Which improves co-presence. Less reverb meant better localization and better intelligibility.
A large, open and detailed environment was appreciated.
And most importantly, a virtual 5.1 setup with spatialization without reverberation combined with a large screen provided a convincingly cinematic experience.

These findings were combined in a document for developers who are interested in improving their sound design. Other findings on how to user test for Social VR were also combined.

TABLE OF CONTENT

PROJECT INTRODUCTION

INTRODUCTION

Virtual Reality grew quickly in recent years, for professionals (training, simulations, healthcare) and consumers (gaming, audio-visual experiences). Current equipment ranges from mobile solutions like the Google Cardboard or Samsung Gear VR, to a small scale room with the HTC Vive with controllers. There are even arcades with 360 running mats to emulate walking around. Through this significant progression, companies and content creators have mostly focused on the visual side of VR. The first thing the user experiences with VR is sight, as they put a large screen in front of their eyes. Most solutions focus on that aspect and an external headphone is necessary. Even the more expensive headsets such as the HTC Vive provide a basic headphone at additional costs. Most demonstrations also only focus on the visual aspect and often times the demonstrations are given in environments that do not benefit the sound experience. By limiting that experience, the general immersion takes a hit. Even at the start of virtual reality, a document from 2000 from the NASA showed that: “Another solution for improving the immersivity and perceived quality of a visual display and virtual simulation in general is to focus on other perceptual sense- in particular sound” (Begault, 2000). Dr Laurel also commented that she found that “really high-quality audio will actually make people tell you that games have better pictures, but really good pictures will not make audio sound better; in fact, they make audio sound worse.” (Begault, 2000).
Now that we have reached decent visual quality, it is time to enrich the sound quality. Not only does it positively influence other senses such as sight, it can also provide cues about the virtual environment such as size, space and localization.

The purpose of this project is to evaluate and improve the auditory experience to further increase the immersion of users in Virtual Reality, specifically for Social VR. Social VR is a way of communicating with other people through VR by integrating participants in the same environment and enabling them to interact with each other.

In this project introduction, the project and its scope will be clarified. Firstly, a short description of the company the project started at will be given. Secondly, the context of the project will be determined. And finally, a design goal will be formulated.

THE COMPANY

This project started as a proposal to TNO, a research company. The first and second phase of the project were done in collaboration with TNO during an internship of 8 months.

TNO is an independent research organization that focuses on creating common experiences of social and economic value. It was founded in 1932 as a way to provide innovations to improve the competition between companies and to improve the welfare of society. TNO focuses on nine social themes: infrastructure & maritime, circular economy and the Environment, Defence, Safety and Security, energy, healthy living, industry, information & communication technology, strategic analysis & policy, traffic and transport.

Within TNO in The Hague (figure 1), in the Networking department is a place called MediaLab where they are working on Virtual Reality experiences, with emphasis on what is called Social VR. This Social VR is a way of communicating with other people through VR by integrating them in the same environment and interacting together. The user is static in a room represented by 360 image with additional interactive elements such as a screen. They are currently focusing on two main projects: VR Together and KPN project.
VR Together is a European project involving 10 companies with the aim of radically improving the experience by innovating in how media formats are used and providing the feeling of being there together. This initiative is currently working on a VR experience in an interrogation room using 360 videos.

Figure 1. Entrance of the TNO New Babylon location in The Hague (The Netherlands) (Credit to TNO)

Mission statement:

"TNO connects people and knowledge to create innovations that boost the competitive strength of industry and the well-being of society in a sustainable way."

THE CONTEXT

Social VR presents opportunities of communication in different contexts. There are currently three distinctly different situations for Social VR TNO is currently working on.

Business stand up meeting

This is when a meeting is organized in which the attendees stand up to keep the meeting short and to the point. Having the meeting in VR would allow users from different physical locations to meet and see each other which works better than Skype meetings or using the telephone.

TV watching experience with friends/family

In this case, the participants are sitting on a couch and watching the same content while being able to interact with each other. For now only two people can sit next to each other or in front of each other in TNO's working demonstrations. The goal is eventually to be able to have more than two users at the same time.

Entertainment experience

This is a special experience that is created for multiplayer purposes and requires the communication between the participants. TNO is currently working on that experience.

Each context presents its own sound requirements, for example it could be assumed that total immersion in the virtual environment both visually and audibly is not the main goal of a stand-up meeting. That context would benefit more from clear and audible sound without distractions. However, total immersion would be of more value for the entertaining experience. The ideal user experience is therefore very different from context to context.

CHOOSING THE CONTEXT

Although the entertainment experience would benefit more from immersive and complex sound design, TNO's timeline with this project does not align with the graduation project.
However, they already have a TV experience demonstration that works where two users are sitting next to each other and watching the same content. It would be a good basis for sound experimentation while also avoiding complex visuals which would be out of the scope of this graduation project. Next to that, some of the insights from that context can still be applied to Social VR in general.

From a practical point of view, it makes the most sense to choose the context of "watching TV with a friend in VR". A design goal will be formulated for this context, and the user requirements and interaction qualities can be analysed to create an interaction vision to guide the prototyping.

DESIGN GOAL

Originally the design goal was to make the most immersive sound experience in Social VR. Throughout the research, it became apparent that this is not the only aspect that matters for the user experience. The social aspect, the content and the environment all matter. The goal has evolved into:

"Improve the user experience of Social VR through sound in the context of watching content with friends in VR"

Users and stakeholders

The users are people who currently use Social VR, but also future customers, people that already meet up in real life with friends to watch content too and might want to try meeting in VR.

Situation

The sound experience in Social VR is currently rather limited and does not contribute to a good user experience. Users want improvements and new users come with their own expectations.

Effect

The intended effect is to allow users to relax and share an experience in VR and possibly even improve certain problems that occur in real life.

Current interactions

In the chosen context of friends watching a series/film together in Social VR, there are three main categories of interactions related to sound:

- Communication: the users interact with each other, they can talk with each other, and react to the TV. But the user also hears their own voice while talking.
- TV experience: the TV speakers emit sound towards the users
- The environmental influence: the properties of the virtual room influence the way in which sound is perceived by user. This includes reverberation depending on the size, specific effects such as sound absorption of materials, and background noise.

To visualise the context and the involved interactions, please refer to Figure 2.

CONCLUSION

Since the scope and context of the project were set, the research and design could be started. Beginning with Phase I: Theory & explorations.

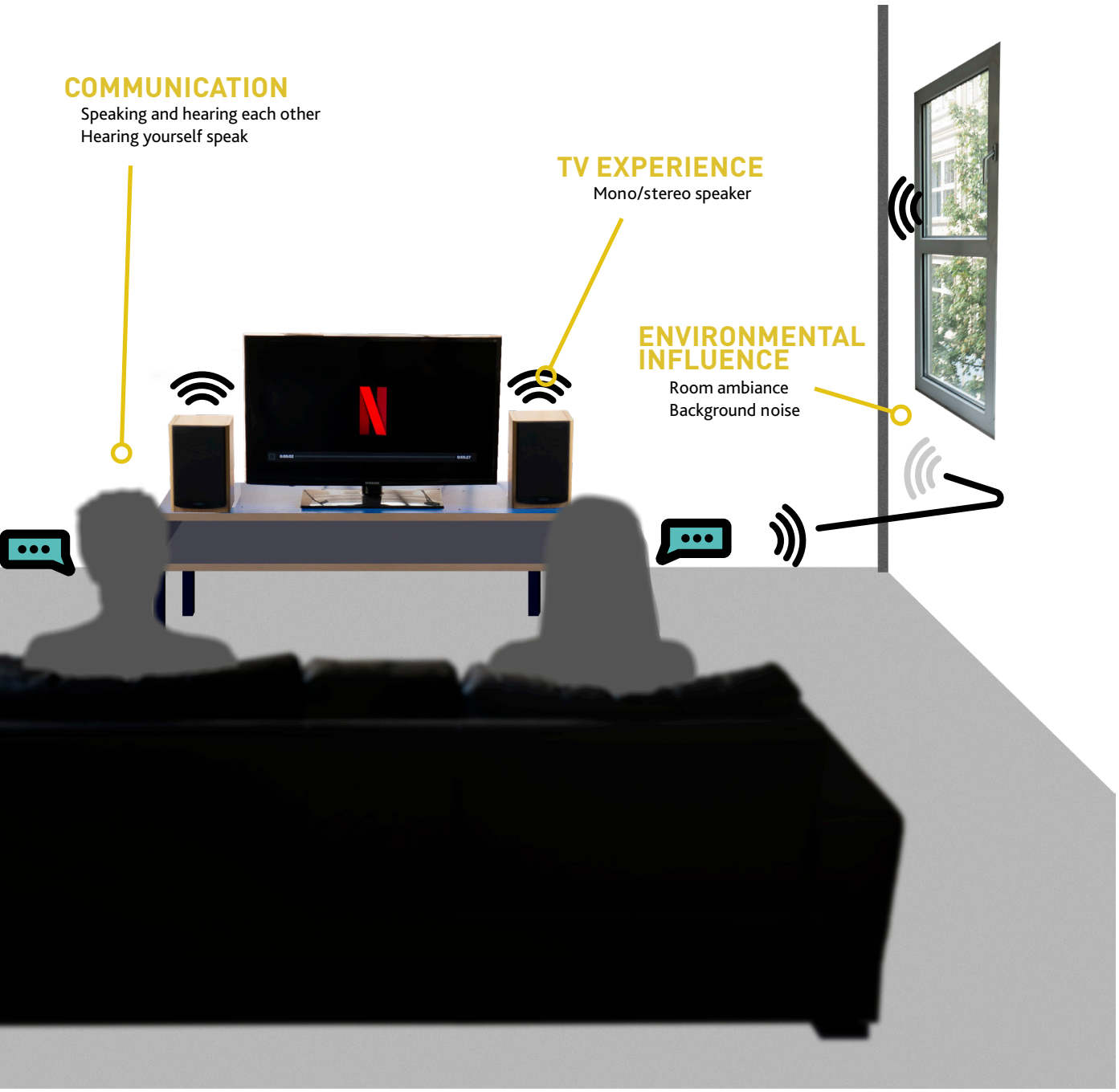


Figure 2. Illustration of the current interactions in the context of watching TV together in real life

PHASE I: THEORY & EXPLORATIONS

INTRODUCTION

PHASE I: THEORY & EXPLORATIONS

For this part of the project, literature research was done to gain knowledge on how the human hearing works, how environments influence the sounds we hear and how sound is represented virtually. An expert on sound was also approached to fill in the knowledge gaps and to be involved in the process. Different demonstrations such as the ones available by TNO and what others offer have also been tested as explorations. And finally, two questionnaires were made to be able to come up with interaction qualities that the users want and find user requirements for how the sound experience should be like. Eventually an interaction vision was formed to inspire the development of a prototype.

The main goal of this research was to understand what the possibilities are for the sound experience in Social VR, and what software and hardware can be used for a prototype.

Since the following parts involve a lot of theoretical knowledge, a glossary was made to keep up with the definitions. It can be found in pages 2 and 3.

To summarise:

1. Literature research findings:
   - Localization & spatialization
   - Environmental influence
   - 3D audio & software
   - Headphones & speakers
   - Microphones

2. Initial Social VR experience test findings

3. Expert interview findings

4. Questionnaires findings

LOCALIZATION & SPATIALIZATION

First we will look into how hearing works in real life, then how sound works in virtual reality.

Although humans only have two ears, they are able to localise a sound source in 3 dimensions, not just on the horizontal plane but also vertically. This is made possible through psychoacoustics and inference modifying the timing, phase, level and spectra. The two components of localization are direction and distance.

DIRECTIONAL LOCALIZATION

When the sound comes in laterally, it can be localised based on the delay between the time it reaches the left ear and the right one. Since we have two ears, when a sound is emitted, there will be a time difference between the time it reaches one ear then the other unless both ears are at the same distance from the source. This helps to find out the direction of the sound, especially for low frequencies. This is called Interaural Time Difference (ITD). Another cue for direction is the difference in level of intensity of the sound in each ear. The head itself obstructs sound to reach your ear. This is however only helpful for high frequencies. This is called Interaural Intensity Difference (IID) or sometimes referred to as Interaural Level Difference (ILD). The preferred localization technique depends on the frequencies coming from the sound source. Low frequencies (under 800Hz) are more difficult to experience the intensity difference (IID). For low frequencies it is therefore easier to experience the interaural time difference. For frequencies above 1500 Hz however, the inverse is true. Since the head blocks out the sound coming from the left, it makes it easier to locate with IID. In between those frequencies, both ITD and IID are used.

For sounds coming from the front or back of the head, it becomes more complicated as the sound could hit both ears at the same time and at the same intensity, making ITD and IID unusable. This is where our ears rely on spectral modifications of the sound which the brain translates into localization. Sound localization for front-to-back and elevation is especially dependent on the pinna and concha. The pinnae effect is caused by the pinna's ridges and grooves which create slight delays of the order of 0-300 microseconds (Begault, 2000) while the concha acts a resonant cavity. The asymmetry of the pinnae causes spectral modification according to where the sound source is which is translated into a spatial cue. They help determine the horizontal and vertical position of a sound source. The small delays mostly occur for high frequencies (above 5 kHz). These effects are dependent on the person as every ear is different and is learnt. Hearing a recording through someone else's ears might be worse or better for localising sounds. (Howard, 2017) The pinnae furthermore provides ambiguous ITD and IID cues. (Begault, 2000)
Another powerful way of determining front-to-back and elevation is by moving our heads: the head effects. If our head stays immobile, a sound coming to our ears creates a "cone of confusion". Using the delays, we can somewhat establish the direction it comes from. Moving the head would therefore help get a different "perspective" on the sound and localise it better. If the sound hits the ears at the same time, it creates a "circle of confusion". Modelling the effect of the head can be done with head-related transfer function (HRTF) too. What is referred to as HRTF is the combination of all of these reflections and shadowing effects creating a direction selective filter (Oculus, 2018). More on this on page 16.

DISTANCE LOCALIZATION

Loudness is one of the key components but it relies on the context. If we have no reference it becomes impossible to accurately judge the distance. However, we have learned these references through our daily life which improves our predictions.
Here too, ITD helps differentiate the time delay.

Since we have learned those references in different environments, it is important to explain what influence environments can have on the sound experience.

ENVIRONMENTAL INFLUENCE

SOUND REFRACTION

Much like with light, sound refraction happens because of the variation in the speed of sound depending on the material. When sound takes a different direction, it is called refraction. When the temperature rises, the speed of sound increases. This also explains why when a neighbour is making noise, it is difficult to determine where it comes from because it will be refracted into walls and ceilings and it does not form a direct path anymore. (Pyzdek, 2018)

SOUND ABSORPTION

All physical objects absorb sound because they vibrate when a sound wave hits them. Sound proofing foam works because it is a porous material with a very large surface area of interaction which causes frictional loss.

SOUND REFLECTION

Physical objects can also reflect sound like an echo. Because the sound will not physically move an object, it reaches a point of zero velocity so it bounces back in the reverse direction. For solid boundaries, the reflection results in a reflected pressure component that is in phase with the wave whereas a reflection from a bounded to unbounded region results in an antiphase.

SOUND DIFFRACTION

Diffraction is said when sound bends around objects. Because the variations in air pressure cannot go to zero at once after passing an object, it “bends” behind it.

In a nonanechoic environment, sound travels through space and reflects off of surfaces. When the sound starts to bounce off, it creates early reflections which help finding the direction and distance to the sound source. The other echoes create a late reverberation as it fills the room.

DIRECT SOUND

When sound is emitted from a source, it has three main behaviours in a room. Direct sound, reflections, and decay. The direct sound is the sound wave that travelled the shortest distance to the listener. It has not been influenced or modified by the environment which means it is at its clearest.

EARLY REFLECTIONS

A result from a nonanechoic environment is that there are early reflections which arrive shortly after the direct sound after reflecting off one or more surfaces. This provides cues about the size of the environment and where the sound source is located. They can however reduce the intelligibility of speech because of interference effects. The intensity of the reflections depends on distance and surface (shape and material). The material can absorb the sound with a different coefficient depending on the frequency. (Howard, 2017)

REVERBERATION

After the early reflections, the sound has reflected many times and comes from all directions from many different paths. This dense amount of reflections is what is called reverberation.

Echoes and reverberation influence localization in a significant matter, to the point of increasing the realism of the display, improving distance perception, providing cues about the room (Shinn-Cunningham, 2001). However it slightly degrades directional accuracy because the reverberant energy builds up over time (Takahashi, 2009). So although a realistic reverb helps in the localization of the source and recognizing the room, it also has a downside. If the game is about locating the source, it would hinder the user by adding reverberation.

DECAY

The time it takes for a sound to die is the reverberation time. First it build ups, then it reaches a steady state and finally it decays as the sound does not stop immediately, at least not in a room.

LATERAL REFLECTIONS

This effect is the presence of dense diffuse reflections from the side walls of a concert hall which increases the effect on binaural loudness. This effect is beneficial for concert halls as it enhances the perceived dynamic range and thus the sound envelops the listener more. (Lokki, 2011)

3D AUDIO & SOFTWARE

Now that the theory behind human hearing and environmental influence has been explained, one question remains: how can these effects and properties be applied in virtual reality? The answer is 3D audio.

3D AUDIO

Since we hear sound spatially in the real world, it makes sense to also create spatial audio in virtual reality. It adds a dimension of immersion. It provides feedback for actions and situations that occur inside and outside the user's field of view. This is where 3D audio comes in which involves technical aspects such as mimicking the human body with HRTF and shaping the environment's soundscape with reverberation.

Head related transfer function

HRTF is used to simulate the effects of interaction of a sound originating from a specific direction with the geometry of a particular human body.
The use of HRTF is crucial in 3D audio spatialization techniques, much more for VR than surround systems because of the addition of head movement in the equation. HRTF however only provides an anechoic model of three dimensional sound which therefore lacks the room ambiance (reverberation and reflections).

In today's state of VR, HTRF has to be distinguished between far field HRTF and near field HRTF. A sound in the distance hits both ears at almost the same angle and intensity. As the sound comes closer though, in the near field (<1m), that angle becomes greater and the head shadow becomes greater too. (See figure 3). The head shadow is the attenuation of sound caused by the head lying between an ear and the sound source.

Another distinction is volumetric HRTF. The sounds can come from a point, but also from a volume. This is interesting for example for trees since the sound does not come from a point, but from a large volume. The whole tree makes a noise. (Oculus, 2017)

Transmission path

To represent the transmission path from operator to listener in a 3D sound system, these models involve a source, medium and receiver (see figure 4). The source is one or more vibratory sources, usually expressed in single sources because virtual spatial hearing sets positions for each single source. The medium is the path used to reach the receiver (listener). It could be through air, metal or any material. The receiver includes the listener's hearing system from the structure of the hearing system to the way it is processed by the brain.

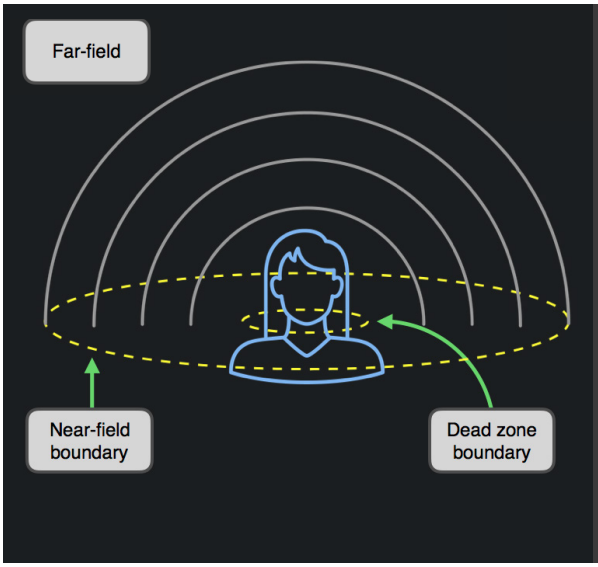


Figure 3. Illustration near and far-field boundaries (Oculus, 2018)

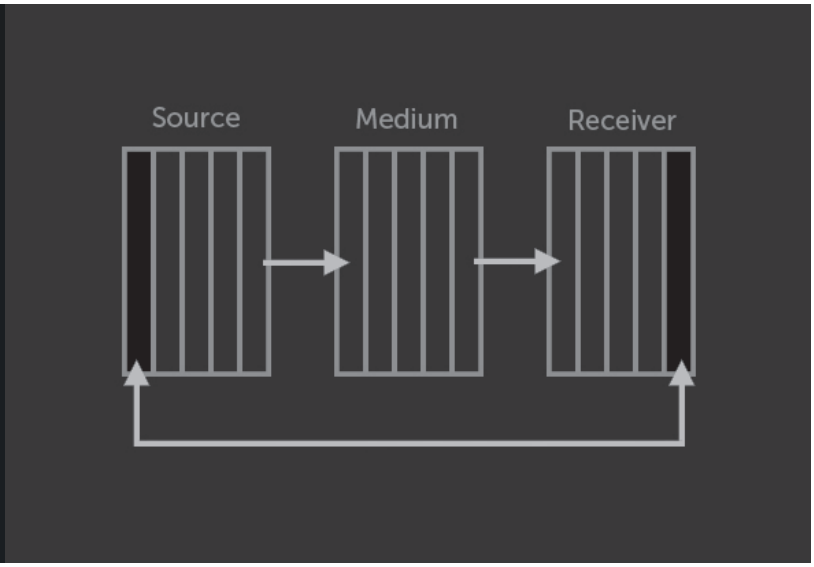


Figure 4. Illustration of the transmission path from operator to listener

Environmental context

The environmental context is the space in which the user is and how they interact with each other. Contextual cues include the effects of reverberation caused by reflections of a sound source on the surfaces or objects in a room or outdoors. These cues give information about the location of the sound source and it gives an image of the space occupied by the sound source.

If the sound source propagates in no particular direction it is omnidirectional and emits in a spherical field. In an anechoic chamber, which is a room which removes any reflections, then the sound arrives directly as a plane field perpendicular to the sound direction. In a nonanechoic environment, the sound propagates both in direct and indirect paths which provides environmental context.

For environmental context and distance, you need loudness, spectral content, reverberation and cognitive familiarity.
The hardware components to interact with the virtual world are the reality engine in which the visual, auditory and haptic illusions are created, the effectors (helmet-mounted displays (HMD), headphones, controllers) and sensors (positioning tracking for example). (Begault, 2000)

Conclusion

In 3D audio, the inclusion of distance and environmental context effects are imperative for the realism.

ENVIRONMENTAL MODELLING

Next to HRTF, the room ambiance's sound has to be modelled to create true immersion. This can be achieved by recreating reverberation and reflections of the room using different techniques.
One way is to model the room as a "shoebox room" which is a room with 6 walls to simulate early reflections and late reverberation. It is very simplistic and thus not very accurate.
Sampled impulse response reverberation is a way to record from a specific real-world location which can then be applied to a signal later on. These are however difficult to transition between different areas. Which can become a problem if the user can move around between rooms.

SOUND DESIGN IN VR

Since sound in VR comes from point sources or even volumetric sources depending on the head tracking and player position, different rules apply to the sound design.
If a sound comes from a car, it is a point source that follows a path, if it is rain though, it can be an ambisonic sound. An ambisonic sound is basically a recorded sound sphere. Much like how a 360 image is taken from one point in space, the illusion only works if you listen to the ambisonic sound without moving. To mix in music however, since it is stereo, it will sound from your head and it will tune out the other sounds. To mix it properly, what you can do is divide it into four channels (left front, left back, right front, right back) and use that as ambisonic sound.
There is also a difference between 360 video and Full VR audio.
In Full VR, the sound is tracked to the sound source which can be in movement while in 360 video it is not making Full VR more interactive. A 360 video is 3DOF (degrees of freedom) type content as in, you can only look around this spherical world. The ambisonic sound is therefore well adapted to this restriction, but it also means that you can not use ambisonic sound for Full VR. Full VR has 6DOF, because you can walk around and look around. (Earick, 2017)

SOFTWARE REQUIREMENTS

From the previous research, a list of software requirements can be made in order to choose a software for a Social VR prototype. As concluded from the theory background of psychoacoustics, several effects are needed for good spatialization and spatial awareness: ITD, IID, HRTF and pinnae effects. To be able to experiment, ambisonic files should also be supported, there should be geometry-based reverberation and far field/near-field effects if possible. "Generic HRTFs might be good enough to enable good auditory source localization in VR." (Franco, 2018) so custom HRTFs will not be considered.

CHOOSING THE SOFTWARE

A list of software was created which showed promise for this project's scope. See Appendix 1.A. for the full list of software. After consideration, Resonance Audio was chosen.

Resonance Audio

Resonance Audio is made by Google, and this powerful spatial audio technology can create realistic experiences for AR, VR, gaming, and video. It includes the most important aspects from the software requirements: it supports ITD, IID, generic HRTF, room ambiance (reverb & early reflections), and ambisonic files. It even support occlusion and sound source directivity. Sound source directivity is for example that a speaker mostly sends sound forwards, while it sounds less loud from the back. This effect is visualized in figure 5.
Next to that, it has a reverb model with is based on the shoebox model, but also takes sound absorption and reflection into account with the walls of the room. These walls can be assigned a material which has different properties and changes the reverb.

This software development kit (SDK) is available for Unity, Unreal, Audiokinetic Wwise, WebVR and DAW.

Because of the wide range of availability, Resonance Audio was chosen to handle the 3D audio part of the experience. And it included the most important features required for a realistic sound experience. It also provided occlusion and sound source directivity which could be interesting features to experiment with.

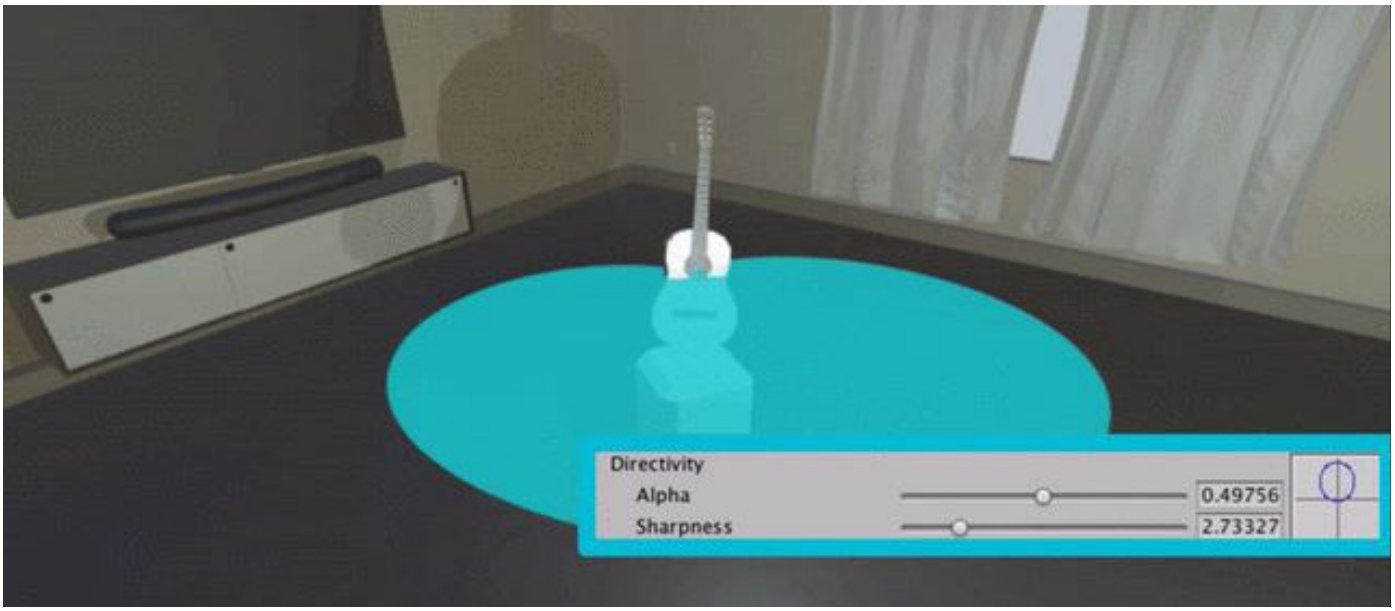


Figure 5. Sound directivity of a guitar, previewed in Unity with Resonance Audio (credit to Unity3D)

HEADPHONES & SPEAKERS

In VR, the user has a head mounted display (HMD), controllers, headphones, and a microphone. Since a basic Social VR experience will be prototyped, it is important to explore the range of hardware possible and choose which is more appropriate.

HEADPHONES

Headphones can give you the impression of being in your head because the sound hits both ears at the same time. It is the fact that it always stays like that that makes you think it must be in your head as you cannot localise it around you. They also do not replicate the IID effect.

Stereo headphones

These headphones have been the standard for most listeners and are the most popular choice. Stereo provides two channels, one for each ear much like stereo speakers. Because the drivers are perpendicular to the ear and really close to your middle ear, it creates the illusion that the sound is played in your head. This also limits the spatial information that can be provided from the hardware as it can only play sounds from the left to right and in between. (Ossic, 2017)
These traditional headphones are used for VR in combination with virtual surround sound with 3D audio.

Surround headphones

For surround headphones there are two types: true surround and virtual surround.
True surround headphones traditionally have multiple drivers per ear cup. However, they have been noticeably worse than virtual surround for years because the drivers are too close to each other to truly make a difference. (Rozario, 2017) Since there are also many different drivers, it decreases the sound quality because the drivers are less expensive and smaller in size.
As explained previously, stereo headphones negate HRTF effects and make it sound like from within your head. A more popular solution to surround sound is virtual surround. There are several technologies that have been developed to simulate a surround sound with technologies such as Dolby headphone (X), Yamaha Silent Cinema, Auro 3D Audio or Dirac VR. All of these are compatible on any headphone, however they require an external compatible device with the technology on it. (Silva, 2018)

Closed-back vs. open-back

For over-the-ear headphones, there are two types: closed-back and open-back. Closed-back has a hard enclosure at the back of the earcup which blocks the sound going outward and is better at isolating you from the world. This makes it sound like the sound is in your head. Open-back, however, leaves the back open which results in the sound also going outward. This gives the sense of listening to music in a room rather than in your head which is more realistic. (Morrison, 2013)
There is also a hybrid, the semi-open back headphone. These let less air in and out of the ear cup than a fully open-back headphone, but more than a closed-back one. Similarly, it is recommended to listen in quiet places as it leaks sound in and out. (Thomas, 2017)

Earbuds

Next to headphones, another popular alternative are earbuds which are often provided for free with phones or other electronic devices. These usually lack bass and have poor isolation, but are more practical because of their size.

In-ear monitors

Often called IEM, these offer an excellent frequency response and they offer better isolation as they are placed into the ear canal. The accurate frequency representation is needed to have a neutral sound influence from the hardware in order to properly “monitor” the sound.

Wired vs. wireless

Headphones also come with a wire, usually to connect to a 3.5mm or 6.2mm input. They can also come without wires, connected through Bluetooth, Infrared (IR) or Wi-Fi. Bluetooth headphones process sound like this, it is sent by an audio codec as digital data, then processed by an internal DAC in the headphones. In theory, the DAC in headphones

can be of better quality than the internal DAC in smartphones or computers as there is more space to play with. (Prasad, 2016)
There are currently two high definition codecs for Bluetooth headphones: Sony's LDAC and Qualcomm's aptX HD. These codecs provide 24-bit music quality, but they still have to compress sound to be sent through Bluetooth. Sound compression reduces the details which the algorithm thinks can not be heard, which also drastically reduces the amount of data that has to be transmitted from the device to the headphone through Bluetooth.

So although in terms of sound quality, the difference between Bluetooth and wired should not matter nowadays, latency can become a problem, especially combined with additional audio effects. AptX HD for example has 100ms latency while AptX LL (low latency) has 35ms. 40-50ms is usually the maximum delay before it becomes noticeable (Telme2, 2017). The HMD itself also has a delay of about 26ms without compensation (Kawamura, 2016).

SPEAKERS

Although VR demonstrations use headphones, it is important to look at the wide range of speakers to know if they are an option.
There are currently a lot of different types of speakers for different needs such as listening to music or watching a movie at the theatre. For more traditional speakers you could find at homes, there is monophonic, stereophonic and surround sound. Monophonic sound is a single channel or track of sound produced by a single speaker. Nowadays it is often used for mobile Bluetooth or Wi-Fi speakers at home such as the SONOS which can be combined to add more channels (Sonos, 2018). By adding an extra speaker, you can get stereo sound. The two speakers play separate audio channels and creates the illusion of hearing sound from the Left or Right. As for surround sound, the same principle applies. By adding speakers and independent audio channels you can get 5.1 channels which is a setup with two stereo speakers in front of the listener, two in the back, one in the centre and a subwoofer for the low frequencies. Another popular surround setup is 7.1 which adds two side-surround speakers on the sides of the listener. As explained previously, we cannot locate low frequencies well which is why most surround setups only include one subwoofer (Altunian, 2017). These are the more traditional types of setups. These mostly help for locating sound on the horizontal plane. 9.1 channels for example, adds two front height channels for an extra dimension, to also locate on the vertical plane (Dolby, 2018). On the more advanced setups, we can look at cinemas with for example the Dolby Atmos which is essentially setting up speakers on the sides and on top so each speaker can play a unique channel and match the direction of the sound that the movie is showing. Through software, sound engineers and soundtrack creators can place and move around sounds in 3D space. This technology has also been applied to the Dolby Atmos soundbar from LG which is a soundbar with 4 speakers on the sides, 1 in the centre, 2 on the top and one subwoofer. Although these speakers are not placed around the user, it still manages to recreate surround sound (Trends, 2017).
At the MoMa, a surround art installation was installed by Janet Cardiff which had hundreds of speakers in a circle and each played a different singer. If you sat in the middle you would only hear the full track, but if you walked around you could clearly distinguish the individual channels.

CHOOSING

Although there are some tricks to make it possible to use Social VR with speakers, the cost of the setup, practicality, room dedication, sweet spot and interference with the microphone make it a poor choice. In terms of headphones, a traditional stereo headphone seems to be more appropriate as there are software designed to create immersive sound for stereo headphones. An additional feature for practicality is using Bluetooth and using noise cancelling to isolate the external noises while in VR, but it would add latency.

MICROPHONES

In Social VR, there are two main sources of recorded sound: pre-recorded sounds for the environment and soundscape, and real-time recording for the communication between the players. The current offering will be analysed for both traditional microphones and more specific ones that are often associated with VR. Then a choice will be made between these options to know which microphones need to be used for the context of watching TV together in VR.

TRADITIONAL MICROPHONES

Types of microphones

There are multiple types of microphones available for different purposes: dynamic, condenser and ribbon microphones.

Dynamic microphones are more suited for loud audio recording, often used in concerts where the sound source is rather close to the microphone and where sturdiness is necessary.

Condenser microphones are more sensitive and are more balanced for all frequencies. They sound quite natural, but are not suitable for loud sounds. There are two types of condenser microphones: with small or large diaphragm.

Ribbon microphones are extremely sensitive, they pick up a lot of the surroundings (Glynn, 2016).

Directionality

Directionality and pick-up pattern determine what direction of sound is recorded such as front, rear, sides or all around and with which pattern.
Sometimes you do not want to record all surrounding sounds for which a unidirectional (a.k.a. cardioid) microphone is more appropriate as it only picks up what is in front of it.
If you want to pick up sound from the front and rear but not the sides, a bidirectional microphone does just that.
An omnidirectional microphone, however, picks up everything surrounding the microphone with no focus on a single direction.

Other microphones and setups

Next to dynamic and condenser microphones, there are other kinds of microphones with more specific purposes.

Headsets are microphones worn around your head especially used for customer service calls. These are usually omnidirectional microphones.

Lavalier microphones are microphones attached to clothing not far from one's mouth. They are usually omnidirectional condensers (Yamaha, 2018) but there are also cardioid lavaliers which are more suited for really loud environments or when trying to avoid feedback from speakers.

Shotgun microphones are used for interviews for example and are very directional which makes it easier to aim at a source and exclude surrounding sounds (Boyd, 2014).

With delay stereo: using two omnidirectional microphones spaced at a reasonable distance, it accounts for delays (ITD) but the movements of the sound sources are usually not drastic enough that the intensity changes (IID).

With intensity stereo: two directional microphones pointing at the left and right. Because they are so close to each other, it does not change the delay (ITD), but it changes the intensity (IID).

MICROPHONES FOR VR

These microphones have been separated from the others because they need to be better looked at since these are more specifically associated with VR.

Binaural microphones are using two omnidirectional microphones, placed in an artificial head at a 18 cm distance to create 3D sound and represent the pinnae effect and ITD. Some include the artificial head to include the ITD effect, some only have ears and rely solely on the pinnae effect. It hears both the intensity and the delays which creates a believable illusion. It essentially combines both the delay stereo and intensity setup in one (Howard, 2017).

Ambisonic microphones are using four different channels with sub-cardioid microphones

in order to record from every different direction, but also record on the horizontal and vertical plane. It records a sphere of sound which is well suited for VR or surround systems such as the Dolby Atmos. There are different formats too, B-Format is 4-channel type, A-Format must be further processed to separate the 4 channels and there are higher-order Ambisonic (Virostek, 2017) with more channels. The sound from ambisonic can be made binaural with 3Dception (Oculus, 2015). The first order uses 4 channels, the higher order works by adding 5 new channels for the second order, 7 new for the third etc. (Bates, 2017).

There are however problems with ambisonic microphones as it has been recorded with a microphone in one specific spot, the illusion of a sphere of sound only for a specific location.

In a study designed to let blind people explore architectural places through auditory virtual reality only, they found that first order ambisonic recordings performed more poorly than binaural only if it only used the head tracking (3DOF). Including displacement (6DOF) improved the experience though by using ambisonic. This research shows that auditory virtual reality is advanced enough to match real sounds in terms spatial configurations of environments and the location of sound sources. Reflections are very important for blind people together with interactive sounds such as footsteps or finger snapping. Using low order reflections has to be avoided, at least for blind people as they are heavily dependent on an accurate representation. However, increasing the quality of the recordings (higher order ambisonic) together with high quality reflections made it unfeasible to play in real time, it had to be pre-recorded for this study.
Ambisonic microphones have therefore proven themselves to be accurate and realistic, at least in its representation of the environment the sounds came from (Picinali, 2014).

CHOOSING

Through this research and by looking into competitors in Social VR (Appendix 1.B.), a combination of microphones can be chosen. Social VR has two different needs for recording: one for the environments, to create and record the ambiance properly of the environment that you try to mimic, and one for the users interacting with each other which will have different requirements. According to Michael Wohl (award winning filmmaker), for a 360 video you need an ambisonic microphone for the ambiance right next to the camera. Or, in case of a 3D environment, the sounds can be mono and spatialized in the 3D environment since ambisonic sounds would not work well.

For specific sounds though such as human voices, you need a more traditional microphone such as a lavalier microphone as close to the source as possible to get the least reverb and widest frequency range (Casman, 2018). This way it only picks up the person speaking. It would need a pop-filter to avoid unwanted spikes of saturation.

TESTING SOCIAL VR

Next to the theory, some available demos of Social VR were explored to understand the state of Social VR.

TNO DEMONSTRATION

TNO currently has a demonstration in which 4 users can interact with each other in the same virtual environment, a meeting room with a table and a display going through a presentation. They are integrated together by using Kinect sensors to separate the person from their background through depth and colour. They are then integrated in the same environment. (See figure 6). To understand the users' experience better and immerse myself in the context, I tested it myself.

The experience is that we are watching a presentation that slides through on its own while in a 360 picture of an office. If you move around, the scaling is wrong in the virtual environment. There were a lot of sound issues though, one microphone would saturate, two microphones were only played on the left ear and none of it was dependent on the head orientation. During the experience, I had no idea how loud I was talking. The noise cancelling headphones did a great job at eliminating the noise from the busy room we were in. There were no effects or sounds coming from the environment or from the content.

Conclusion

The sound experience to me was mixed, it would help feeling more immersed if it had 3D audio so I could hear the sound coming from the person I am looking at. The quality of the microphones could also be improved and not hearing myself was confusing.

Full details of how it was tested it in Appendix 1.C.

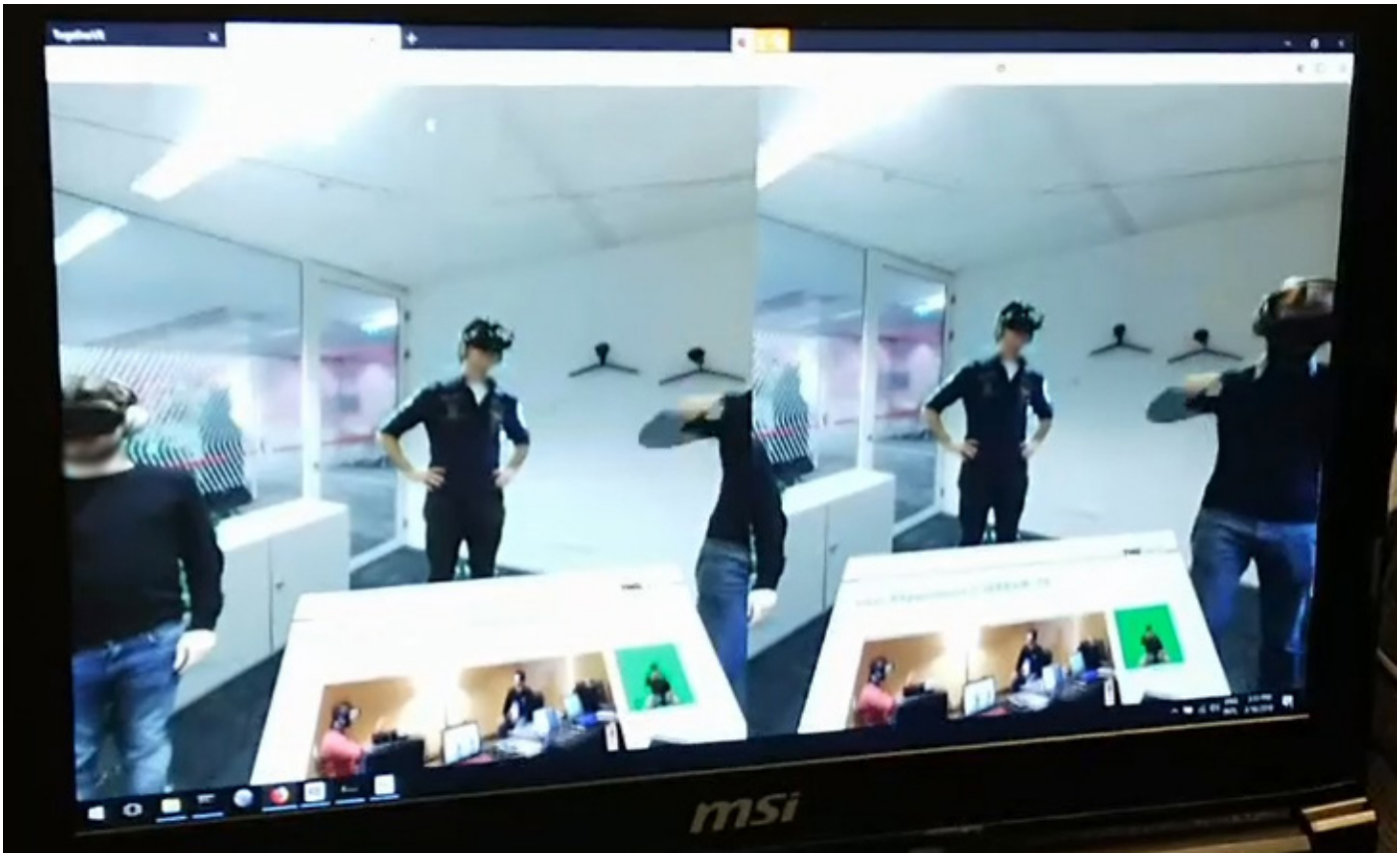


Figure 6. Image of what the user sees in their HMD.

BIGSCREEN

There are several experiences available publicly that are considered Social VR, but the closest and most popular competitor for a TV watching experience with friends is BigScreen. The way it works is that there are multiple environments available in which you share your desktop with others, or go to a cinema theatre to watch content together. The users can interact with each other by talking, they can also show more expression by moving their hands. There are a select few gestures possible. They have an avatar representing their head and hand movements, it also lip syncs and shows mouth movements (as shown in figure 7). The recording is through the integrated microphone of the HMD. Although it does not provide content itself, it focuses on the social aspect of it (Reality Check R, 2017).

I tried the application and connected with strangers to watch a movie. The sound from the movie comes without modifications and is just stereo sound. It does not change with head movements. Other than that, the sound quality of the movie is rather good without noticeable latency. Sound coming from users either pan fully to the right or fully to the left depending on your head position, no in between. This kind of panning was distracting to me and I thought the sound coming from the other users was pretty loud compared to the content. I have also used it regularly alone for a few weeks and although the content is more pixelated than it would be on a normal screen, the experience of being in a dark cinema allows to focus more on the content and be less distracted. The sound not tracking your head movements does not matter that much when being alone because I do not move my head around as much.

Recommendations

That being said, it would be a much better experience if you could get the Dolby Atmos experience or at least surround sound. The voices should also be spatialized with 3D audio and HRTF. With or without post-production effects is debatable, but being able to localize the sound would help and could be more intelligible because when the sound is either left or right and the content is in stereo, both sounds come from the same space.

Figure 7. BigScreen screenshot of 3 people sharing desktops and talking in VR (credit to BigScreen)

VR SCREEN

Next to looking at how Social VR is representing sound, looking at experiences that purely focus on watching content can help finding out how sound can be implemented. Many applications were found. But even the ones focusing solely on the movie experience such as Cmoar and CINEVEO, were really advanced on the visual side with 3D, realistic lighting and light reflections yet lacked any sound features other than stereo sound. However, there is the application called VR Screen which includes 3D audio with HRTF and headtracking paired with surround sound for the speakers. Though when I tried to use it, I could only get the HRTF sound to work. It was less convincing than Resonance Audio's implementation as the sound panned too quickly and the difference in elevation was unnoticeable. Next to that the content was only in stereo although the source was 5.1. This could be a hardware problem though. Aside from sound, the application currently only includes a screen in a "void" which is a completely black room aside from the screen (as seen in Figure 8).

CONCLUSIONS

Even experiences focusing on the cinema experience keep the sound experience really basic and focus on improving the visuals. Looking into those experiences was meant to find out what is currently possible. It seems like sound is a weakness all around.

Using a questionnaire to find out if this is the type of experience people expect for Social VR would help determine what is worth improving and how.

Figure 8. VR Screen screenshot

EXPERT

Next to experiencing current Social VR applications on my own, to get an idea of the type of interactions involved in communication, John Beerends, a TNO employee, was consulted as an expert on the field of speech intelligibility, speech processing and auditory perception. There are distinct characteristics in speech that can be evaluated:
Quality of a voice recording (voice quality and audio quality)
Talking quality (hearing one's own voice)

Adding effects in post-production to clean the sound/improve the quality?

Through one of his papers, "On the Assessment of High-Quality Voice Recordings including Voice Postprocessing" (Beerends, 2015), the added effects in post-processing of the voice such as timbre optimization, loudness optimization, de-essing, room reverberation optimization and noise suppression did not improve the perceived quality of the voice recording. Since it has no influence on the perceived quality of a voice recording, it means adding those effects on a live recording voice in SocialVR is not necessary.

What about hearing your own voice?

In one of his other papers, "On the quality of hearing one's own voice" (Beerends, 2002), sidetone is addressed. Sidetone is hearing the sound of your own voice in your headphones/headset/earpiece when talking to someone else through a medium (telephone or virtual reality). In virtual reality, the PSVR does this by using the integrated microphone in the headset (Mitsuownes, 2017), but is mostly used to be able to hear what is going on in the room, not just the voice. However, the feedback from users (Reddit, 2017) is that it actually improves their immersion as their voice is no longer muffled and they can hear themselves breathing, exhaling and talking like in real life.
An interesting effect that was observed while testing the demo from TNO, is that people talk loudly if they cannot hear themselves in VR through the noise cancelling headphones. The real life environment they were in was noisy, however, since they used noise cancelling headphones, them talking louder is not a result of the Lombard Effect. That effect is the reflex to speak with an increased effort when the perceived decibel level is louder than normal (Lau, 2008).

Do you want to be put into that virtual space or include those elements in your own space?

It is feasible to be in a virtual environment in two ways: either you are included into the virtual space or the virtual space comes to you. The first way relies on room ambiance that is simulated on the environment and voices through software by using the cleanest sounds and adding effects over them. The advantage is that you have full control on the position of the sound sources and the room ambiance. The second way relies on using the cleanest sounds too, but they will be placed through speakers in the room and therefore utilizes the room ambiance of your own room. Although the setup is simple and does not rely on software, it comes with many disadvantages. The users cannot move around as the speakers are stationary, you have no control over the room ambiance and you are very likely to get interference between the microphone and speakers.

CONCLUSIONS

From talking with an expert, I found out more about how to evaluate perceived audio and speech quality and the importance of sidetone. Evaluating whether sidetone is necessary for VR will be necessary.
John Beerends was consulted again during the second phase, for the user test of a basic Social VR experience.
He also mentioned that using an ambisonic sound for the environment is probably not necessary. The same effect could be achieved by using spatialized mono sounds such as a fan noise.

QUESTIONNAIRES

Using a questionnaire as quantitative research brought the insights from the target group which resulted in an interaction vision and user requirements. In this case, there were actually two contexts that needed to be looked at. The first wass people talking together while watching a movie/ show in their own living room in real life. This wass relatable to anyone and necessary to judge what people were looking for in that kind of context. The second target group, were the users of Social VR. In this case, users of BigScreen VR because it was already an established VR application with many returning visitors. This way, the interaction qualities missing from BigScreen VR could be extracted. These users could also compare it to real life experiences which makes their insights valuable as to what the user experience should be.

RESEARCH QUESTION

For both questionnaires, the main research question was similar. Their goal was different though.

What interaction qualities would people want to have for sound in Social VR?

GOAL

The general goal was defining the context better with the input of people meeting with friends to watch content. The goal from having two questionnaires was, firstly, to see if there was a difference in interactions in real life versus VR and also to see if the ideal experience was similar for people with little to no experience with Social VR with people who were experienced with it.

METHODS

Since a lot of questions was open, the responses were shortened into words. The most popular responses were included into a bar chart to give an overview of the most popular answers. The colours go from very light green to dark green, from less to more popular.

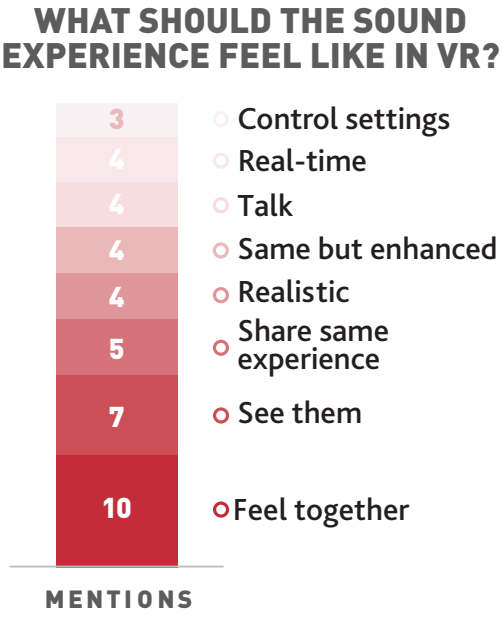
SAMPLE SIZE

The first questionnaire had 38 responses, 19 from ages 25-44, 18 from ages 15-24 and 1 for ages 45-64.
The second questionnaire had a significantly smaller target group which was more difficult to come in contact with. It resulted in 12 responses,
8 from ages 25-44, 2 from ages 15-24 and 2 from ages 45-64.

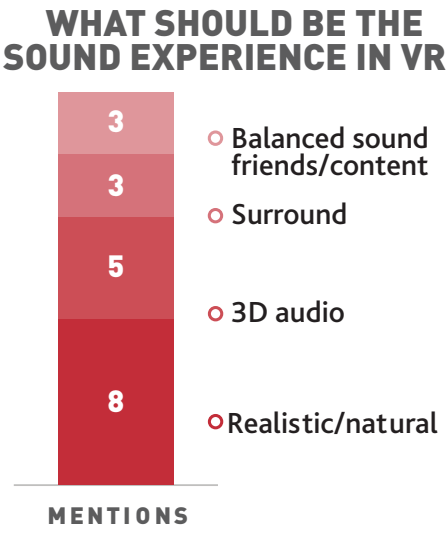
QUESTIONS AND RAW DATA (SEE APPENDIX 1.D. AND 1.E.)

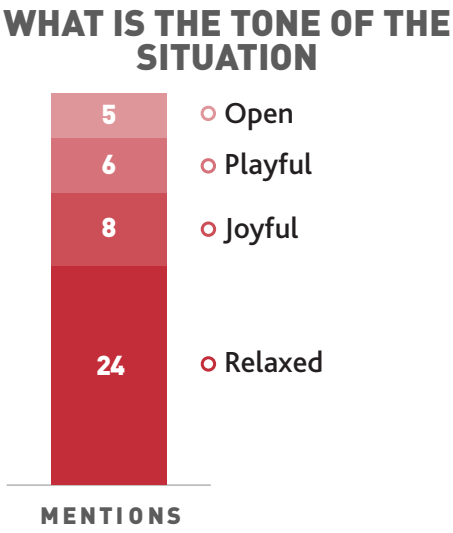
RESULT QUESTIONNAIRE 1

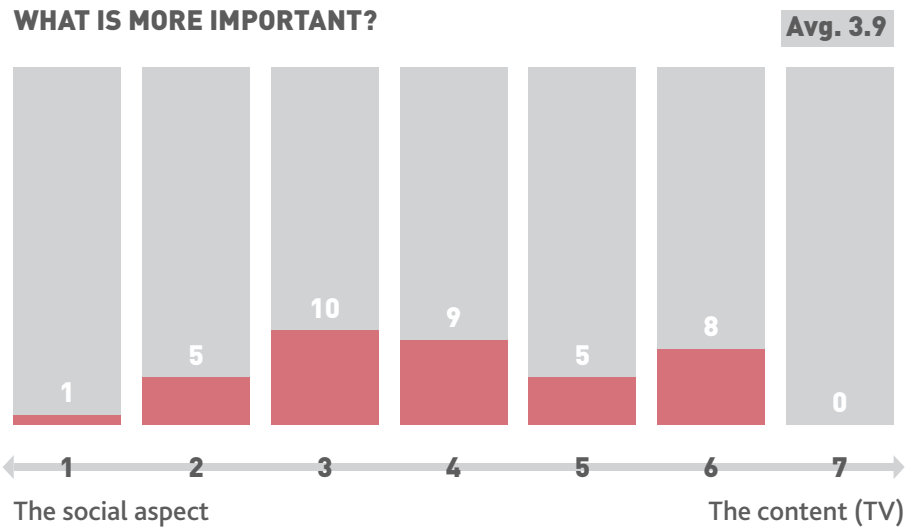
As mentioned previously, the raw data was processed into keywords. Due to the amount of responses, these mentions could be ranked to have an overview of what was most mentioned and how often, in the shape of bar graphs. If it was only mentioned once or if there are clearly outliers, they were not included in the graph.



From the first questionnaire, the most important reason to watch content with friends is to share an experience together so they can talk about it while watching or afterwards.
The tone is relaxed and the



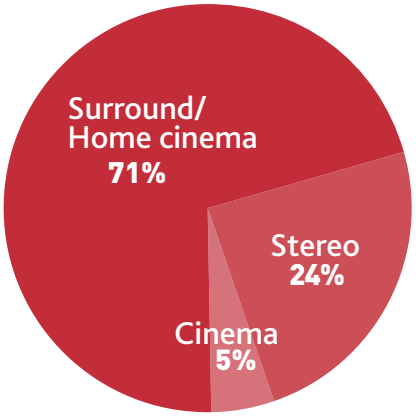


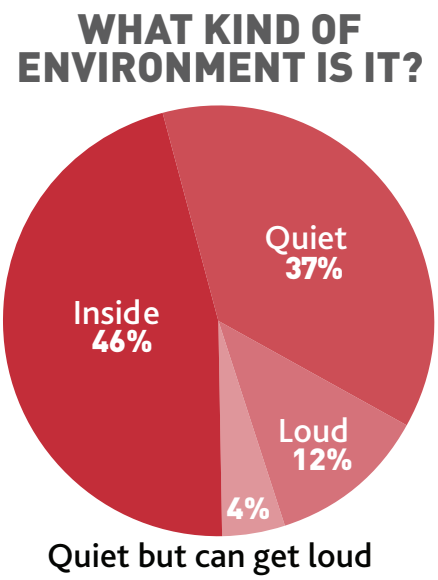


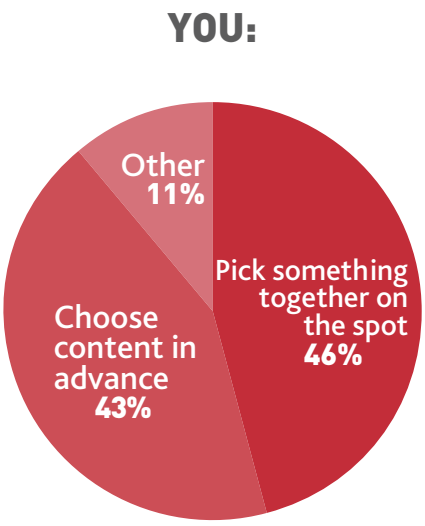
experience is fun. People primarily watch it in a quiet room inside. It can get loud though, but they want to keep the sound down to avoid annoying other people. Sound quality of the sound system/video file and outside noise were also mentioned as experienced problems. In terms of intelligibility, 29% of the respondents often ask their friends to repeat themselves and 60% does it sometimes.

HOW OFTEN DO YOU ASK FRIENDS TO REPEAT THEMSELVES ?

WHAT SOUND QUALITY WOULD YOU PREFER COMING FROM THE TV?







CONCLUSION

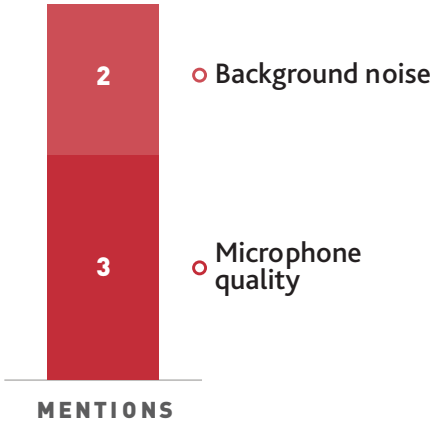
The biggest interactions are talking, laughing and eating. The most important interactions qualities that could be extracted from the answers was that it should feel like they are sharing an experience together, in a relaxed yet fun ambiance and that the experience in VR should be close to how it feels in real life. Intelligibility is definitely a problem since 92% ask their friends to repeat themselves from sometimes to always. Something that could be improved in VR.

QUESTIONS AND RAW DATA IN APPENDIX 1.F. AND 1.G.

RESULTS QUESTIONNAIRE 2

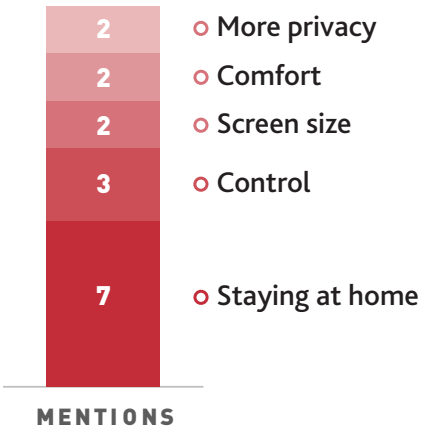
From the second questionnaire, let us start with the similarities. The reason to watch content with friends was the same: share an experience. Some respondents felt like it was also a way to have a social experience with strangers. The tone was also relaxed and fun with talking and laughing as the main interactions. Bad intelligibility though was not only caused by background noise, but also microphone quality. The biggest reason to use VR for it would be to stay at home. The downside of VR vs. IRL is that the graphics and comfort are worse. The HMD being the culprit of discomfort for long periods. The improvement users wanted to see was to have 3D audio for the users and for the content. The preference in map choice is for the Big Cinema, which is essentially looks like a cinema in which you can change seats to be closer or further from the screen. If the server is public you can join strangers too.

BAD INTELLIGIBILITY CAUSED BY:



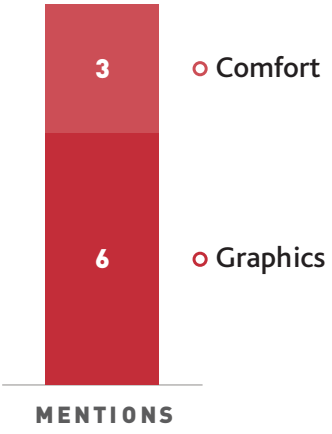
WHY WATCH CONTENT WITH FRIENDS IN VR?

WHICH ELEMENTS ARE BETTER THAN IRL?

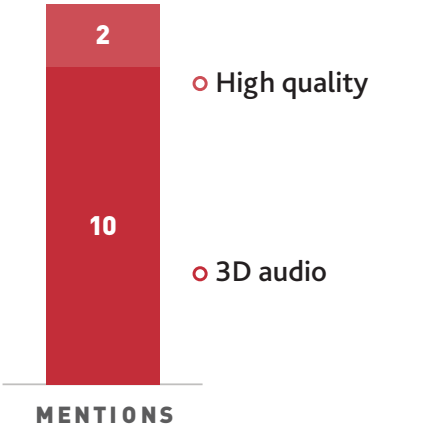


WHAT IS THE TONE?

WHICH ELEMENTS ARE WORSE THAN IRL?

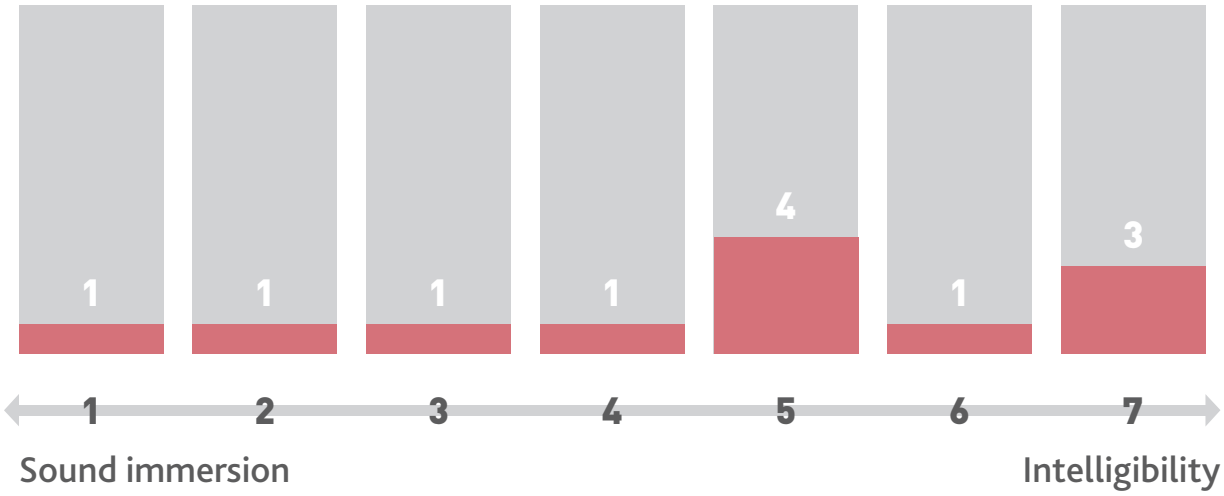


WHAT ARE YOU LOOKING FOR IN TERMS OF SOUND QUALITIES?



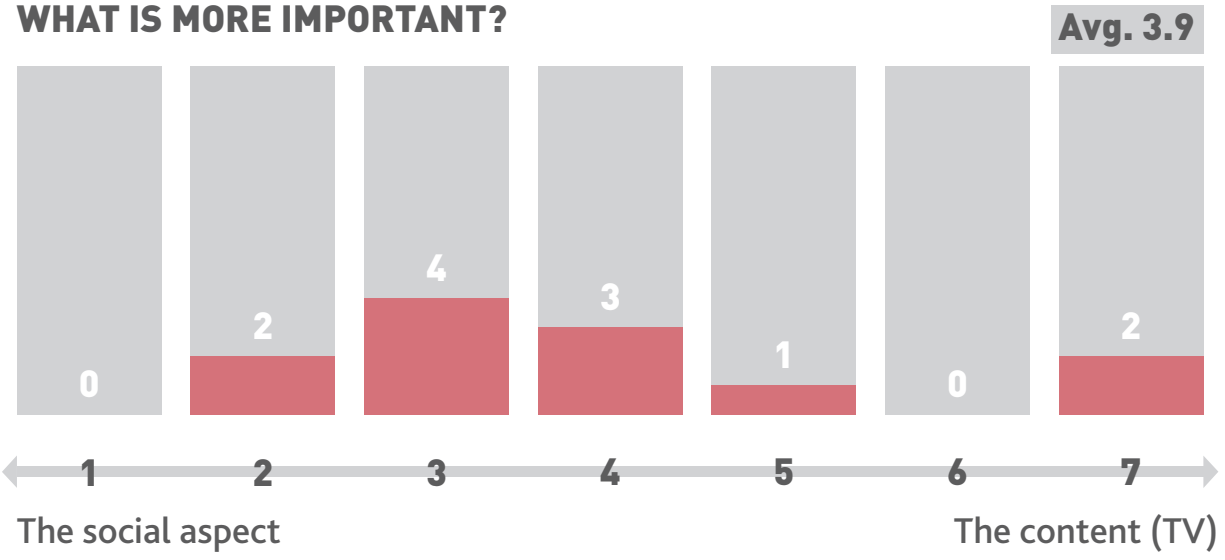
WHAT IS MORE IMPORTANT?

Avg. 4.8

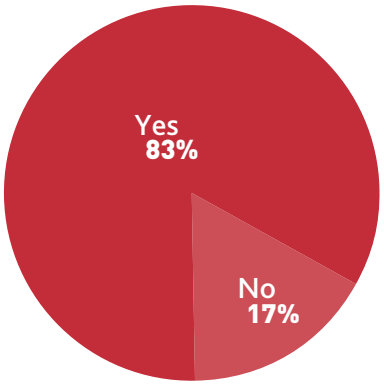


WHAT IS MORE IMPORTANT?

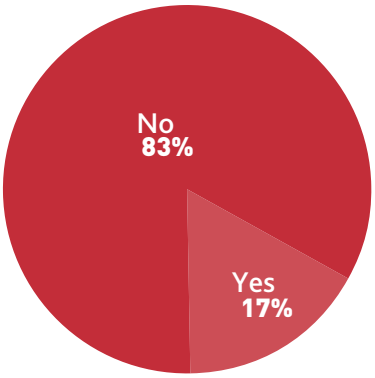
Avg. 3.9



WOULD YOU LIKE TO BE MORE IMMERSED THROUGH BETTER ROOM AMBIANCE EFFECTS (REVERBERATION, EARLY REFLECTIONS, BACKGROUND NOISE)?



WOULD YOU LIKE TO HEAR YOURSELF THROUGH THE HEADPHONES/EARBUDS WHILE TALKING? (SIDETONE)



CONCLUSION FOR QUESTIONNAIRE 2

An interesting difference in interactions is the lack of eating and drinking. Only one person mentioned drinking in VR. The assumption is that the HMD gets in the way of glasses or food. You also can not tell where it is since you cannot see it in VR.
One interesting aspect after reading the responses was to see if there was a bond between preferring the content over the social aspect and the map choice. The assumption was that people who preferred the content would watch it in the Big Cinema environment since its focus is on the content. People who preferred to talk would join smaller maps. There was no clear bond from the answers though. One of the respondents who preferred the social aspect over the content mentioned that "it's more comfortable to sit in a big cinema because there's no expectation to chatter". They also mentioned for more socially oriented sessions or for watching movie with friends, they would choose a smaller environment to enhance the closeness. Nothing sound related, but it is interesting how the environment choice creates different expectations and how one can be more inviting when you do not know the people you are joining.

In general, both experienced Social VR users and novice/ medium VR users have the same interests in the context of watching content together. A relaxed and fun shared experience with a sense of togetherness. If people in VR did not want the social aspect, they would use an app that solely focuses on the content.

The user needs that appeared from both questionnaires are: good speaker quality, 3D audio, no outside noises, people need to be intelligible and surround sound would be preferred for the content.
Some users also mentioned that the experience should be the same as in real life yet enhanced.

The surprising response was that 10/12 people did not want sidetone in VR which went against what John Beerends said. The new assumption was that sidetone was a feature that would need to be experienced before being able to judge whether it is useful or not. As previously found, users of sidetone in Sony's PSVR liked it, reinforcing this assumption. This feature should therefore be tested to see whether or not it is necessary.

INTERACTION VISION

From the research, observations and questionnaire, an interaction vision was created to indicate how the experience should feel like in VR.

Figure 9. Representation of the interaction vision

RELAXING

A majority of the respondents mentioned the relaxed tone while watching content with friends. Keeping this relaxed feeling is important to fit the tone and could be achieved through enhancements that were mentioned in the questionnaire results such as being able to mute or lower the sound of a person. Understanding what makes an environment relaxing would also be important.

TOGETHERNESS

What users wanted was to have the sense of feeling together in the same room. To feel like they were sharing an experience and being able to freely discuss it.

IMMERSING

Immersion and realism was often mentioned and presents a real need from both the people used to VR and those who are not. It also ties in with the sense of togetherness. This also touches on the sound quality and speech quality. One important aspect though about immersion is that it should not stand in the way of intelligibility. There should be a balance between the two.

MAIN FINDINGS

All findings of this phase about sound and visuals in Social VR are listed here. This will be done for each phase in order to combine them all into recommendations.

SOUND AND VISUALS

Current demos focus on visuals rather than sound. None provide 3D audio.

The use of HRTF is crucial in 3D audio spatialization.

Several effects are needed for good spatialization and spatial awareness: ITD, IID and pinnae effects.

Echoes and reverberation influence localization in a significant matter, to the point of increasing the realism of the display, improving distance perception, providing cues about the room.

More reverb means less accurate directional localization.

If the quality of the included microphone of the HMD is not satisfying, a lavalier microphone would be preferred.

For the preferred way of listening, a traditional stereo headphone seems to be more appropriate as there are software designed to create immersive sound for stereo headphones.

The biggest interactions in the context of watching TV together are talking, laughing and eating in real life. In VR the eating and drinking interactions are mostly absent.

In this context, they look for sharing an experience together, in a relaxed yet fun ambiance.

Users want the Social VR experience to be close to how it feels in real life.

Intelligibility in real life is already a significant problem. This needs to be taken into account for VR to avoid making it worse.

Without having experienced sidetone, the majority of users were not interested in it.

Those who experienced sidetone with the PSVR liked it.

CONCLUSION

In this phase, the theoretical knowledge required to get started with prototyping a Social VR experience was acquired. A direction and focus for this project was also established with the design goal and interaction vision. A way to test and prototype a Social VR experience will need to be figured out in Phase II, but it should aim to provide an experience in line with the interaction vision.

PHASE II: BASIC SOCIAL VR PROTOTYPE

INTRODUCTION

PHASE II: BASIC SOCIAL VR PROTOTYPE

Following the research in Phase I, initial ideas were thought out to create the experience of watching TV together in VR.

Considering how long prototyping takes to develop in VR, these ideas could not be roughly prototyped and tested quickly. Setting up experiments was a task in itself. Therefore a different approach was necessary to allow enough time to prototype and test several directions at the same time. The goal was to create a basic Social VR experience of two people watching content together and learn how to prototype it.

Since the qualities that emerged out of the questionnaire were relaxing, immersiveness and togetherness, those were the key terms to orient the test in the first iteration of the experience. Starting from the ground up, the idea was to create a simple sound and visual experience in VR with different environments to understand how users react to the qualities each environment provides. It was also necessary to learn how much visuals had an influence on the perceived sound.

This was not a classic conceptualisation phase, rather an exploration through prototyping to understand how users want the experience to be. There were too many unknowns, and not enough evidence of user's preferences since Social VR is in its infancy.

This section of the report will detail the process of prototyping for Social VR, how the user tests were conducted and how the project was continued.

DETERMINING THE EXPERIENCE

The idea was to create a simplified Social VR experience with sound and visuals with different soundscapes to understand how users react to the qualities each environment provides. This would be manageable enough to prototype yet valuable as a basic experience to iterate on based on the feedback. The experience itself was influenced by previous questionnaires' results.

ASSUMPTIONS & CHOICES

Environments evoke emotions, so before selecting the environments, assumptions were made on what makes an environment relaxing, emulate togetherness and immersion. For example, wind blowing through tree leaves or river sounds, which are consistent and almost white noise, but offer a little variety. These are often used in products to help people fall asleep. To increase the sense of immersion, the assumption was that realism was proportional to it. Creating the soundscape of a highway while it looks like a living room would take the user out of the experience, it would be too unexpected. It was therefore chosen to select environments that were bound to reality. For the togetherness aspect, the environments needed to have a difference in terms of closeness and feeling of cosiness. The assumption was that physical closeness and the feeling of being in a small room was influencing togetherness.
There was also the limitation of having only 360 degrees pictures of real environments which partially influenced the choice of going for real environment rather than imaginary places.

SELECTING THE ENVIRONMENTS

The goal of the selection was to test the interaction qualities, but also to compare environments based on their soundscape.
A closed space would result in reverberation if non-anechoic, while an open space without reflecting surfaces would not have any reflections.

Three real environments were chosen in order to create a wide range of soundscapes. They were different soundscapes yet familiar environments to the participants. A small living room, a large office and a forest were selected. These illustrate the difference between open and closed environments, and each has a specific soundscape that hit the interaction qualities participants were looking for in the questionnaire.

See figure 10 for the environments.

The small room

The small room created a smaller place which could be considered as cosy, creating a better sense of togetherness and potentially more relaxing because living rooms are usually away from stress. The TV was visualised in front, which looks like a wallpaper and then the trailer of the movie Avengers played.

The office

The office was the middle ground, a big room with a fan noise, potentially less relaxing and decreasing the sense of togetherness by increasing the distances. The picture used was from TNO's office, The MediaLab, which has their own wallpaper-like TV on the wall. The picture was also taken from the POV of a person sitting on the couch. The trailer for the movie Ant-Man was used.

The forest

The forest was based on the assumption that people would go there to relax as many sounds found in forests, such as birds, wind, or rivers, are used in relaxing compilations. Next to that, it was the lesser realistic scenario considering this is usually not a place where you would watch a movie. Considering the setting, the TV would look unnatural either way so it was set to float in the air and a couch was photoshopped in to give the sensation of sitting. The trailer for Venom was used.

Figure 10. The three virtual environments

DETERMINING THE CONTENT

After selecting the environments, the content needed to be chosen. This content was changed based on a meeting with the supervisors from TU Delft (see Appendix 1.A.) and a meeting with the supervisor from TNO. The environment needed to be experienced for around 5 minutes, so the participants could get used to it. All the sounds had to be pre-recorded, that was one of the limitations of the prototype method. This required making a scenario and using a video long enough to fit in those 5 minutes. For choosing the content, a suggestion from Hans Stokking was to use trailers in user tests rather than full length series or movies. This way the participant has a similar experience as watching a movie, except it is much shorter.
Next to that, a scenario was created. The experiment started with a narrator telling the participant when the experiment started. It was in mono, so the difference was clear between this voice and the sounds in the experience. Then a voice came from as if someone was sitting next to them. The voice was thanking them for coming and introduced a high pitched sound depending on the environment. For the forest the high pitched sound was a bird flying from right to left above the tree. For the office it was a coffee mug and for the small room it was a tea mug, both including the sound of the spoon hitting the mug. Then the voice lead towards watching a trailer, and the trailer started. After the trailer ended, there was an interruption, either a doorbell ringing or a knock on the door. Then the voice asked what they thought of the trailer and whether they should see the movie. Finally the narrator announced the end of the experiment.

By going through this scenario, it felt closer to the actual experience aimed for: watching content with someone else. Yet it still included different sounds, such as the high pitched triggers that are positioned in more awkward places such as the back or from above.

Since this test was more to get a general feedback about the experience of watching content together, the video was in stereo and spatialized in 3D, while the other sounds were mono and spatialized in 3D.

SOUNDSCAPES

Each environment had its soundscape. The constant between these soundscapes was hearing a trailer from the front, hearing distractions with high pitch and a voice speaking from their left.

The small room had a tea mug that was placed on the table in front of the camera. At some point a doorbell rang, and the other person in the room walked from the couch to the door and back. The room in Resonance Audio was set to the right dimensions to recreate the right reverberation. For the office, the room was set to the size of the office, but the reverberation was accentuated more, to make the difference between both environments more apparent. It also had a tea spoon moving in the mug, someone knocking on the door to get inside the room. It had a fan noise continuously playing in the background.
The forest had a stream flowing on the right, wind blowing in the tree above them, both continuously. At some point a bird flies from above/right to above/left. Since a forest was in the open, the room settings in Resonance Audio were changed to remove reflections.

STRUCTURE

After determining the content and environments, the final choice was the structure of the experience for the user test. Since the focus of the project is on the sound experience, the first part of the user test only included sound. Then the influence of visuals on the sound perception needed to be observed either through comments or by letting the users compare the same environment with or without visuals. Finally, in VR there are degrees of freedom differences. 3DOF+, as TNO would call it, is a 3DOF environment where you can rotate your head and move it a little bit without feeling strange. Normally, 3DOF only offers head

rotation. There is also 6DOF where it tracks head rotation and position in such a way that the user can move around. The assumption was that being able to move around would augment the immersion so both 3DOF and 6DOF had to be compared to see if it adds value in the context of watching TV in VR. For that purpose, 9 different experiences were made in total. That is, three phases per environment.

The phases, as shown in figure 11, were as follows:
- Sound only. The HMD was still to be placed on the user's head to benefit from the head tracking and to "blind" the users by displaying a black screen. Considering the focus on sound, in this phase, three environments were shown to experience all three soundscapes.
- With visuals. It was the same sound experience except the visuals were added. This means the background/room around the user and the video of the trailer were shown.
- With movement, with basic visuals. The sound experience was the same, except the user had the freedom to walk around. The sound sources were visible so the users could move around and experience the difference in sound directivity. The background and video were removed though since being able to move in 6DOF with a static 360 degrees picture is nauseating.

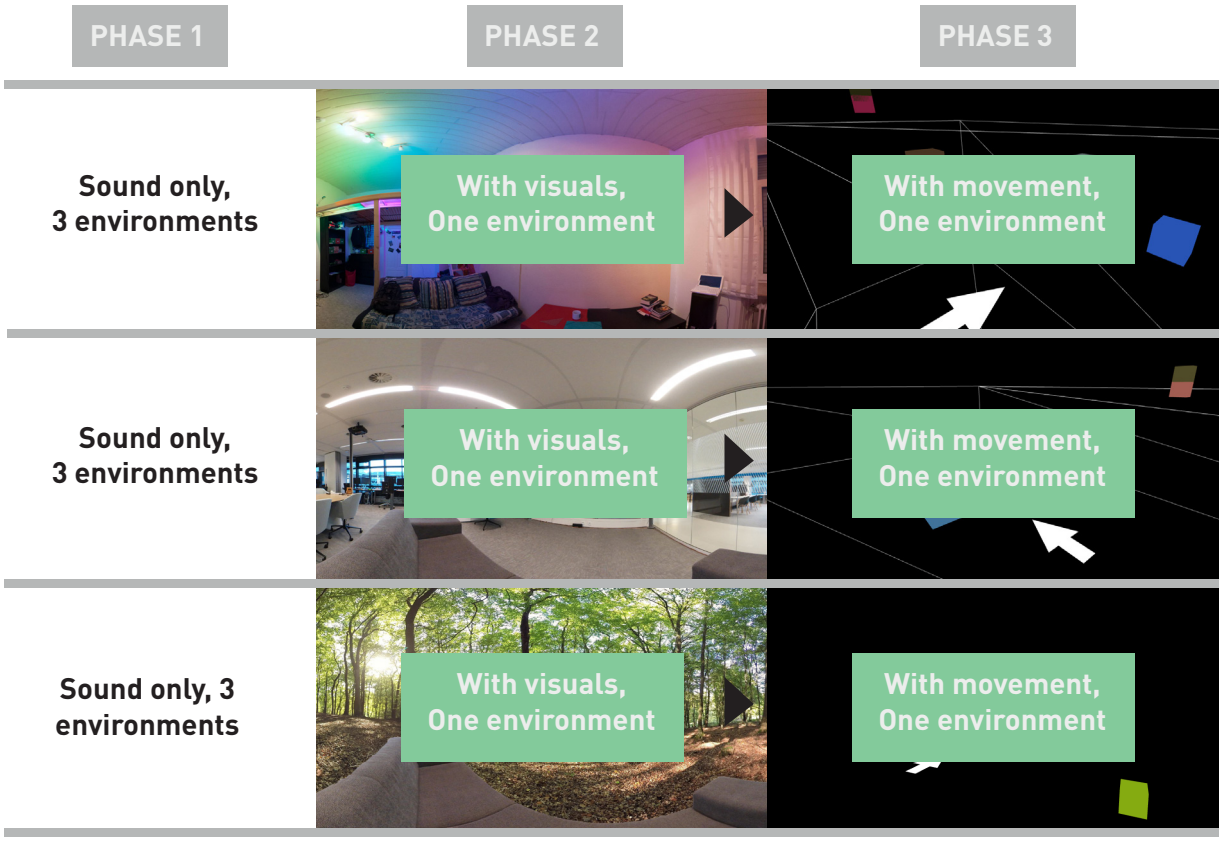


Figure 11. All the different experiences that required to be prototyped

PROTOTYPE V1

These nine experiences needed to be prototyped with three unique soundscapes.

COMPLEX PROTOTYPING

The options for prototyping were laid out in Phase I. As a reminder, TNO have their own implementation of Resonance Audio in WebVR which includes real-time integration of the other person in the virtual environment by using a kinect. They also included real-time communication, however, it did not go through the Resonance Audio settings. The other option was using an existing demo of Resonance Audio available on Github as a starting point to learn how to work with it. This was more adequate for the experience and would be more manageable. The choice of going for a VR experience as a stepping stone turned the experience from low fidelity to high fidelity in one go. The assumption was that starting out with a "flat" experience, i.e. without head tracking/movement, would give results that would not translate well into VR.

The choice was made to continue using Resonance Audio in WebVR with A-Frame, on the premise that it could be used on following prototypes either using TNO's implementation or moving on to another platform compatible with Resonance Audio.

HOW A-FRAME WORKS

A-Frame is a web framework for building virtual reality experiences using HTML for the HTC Vive, Oculus Rift and other HMDs, desktop or mobile. Adding elements can be done through <a-(element)>. A-Frame offers different shapes (box, sphere, cylinder, planes...), but it also includes sounds, images, videos. By adding scripts made for A-Frame, these simple elements can be expanded on. Using Resonance Audio adds realistic reverberation to the sounds depending on the room settings, player position and gives it a sound directivity.

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <title>Hello, WebVR! • A-Frame</title>
    <meta name="description" content="Hello, WebVR! • A-Frame">
    <script src="https://aframe.io/releases/0.8.2/aframe.min.js"></script>
  </head>
  <body>
    <a-scene background="color: #ECECEC">
      <a-box position="-1 0.5 -3" rotation="0 45 0" color="#4CC3D9" shadow></a-box>
      <a-sphere position="0 1.25 -5" radius="1.25" color="#EF2D5E" shadow></a-sphere>
      <a-cylinder position="1 0.75 -3" radius="0.5" height="1.5" color="#FFC65D" shadow></a-cylinder>
      <a-plane position="0 0 -4" rotation="-90 0 0" width="4" height="4" color="#7BC8A4" shadow></a-plane>
    </a-scene>
  </body>
</html>
```

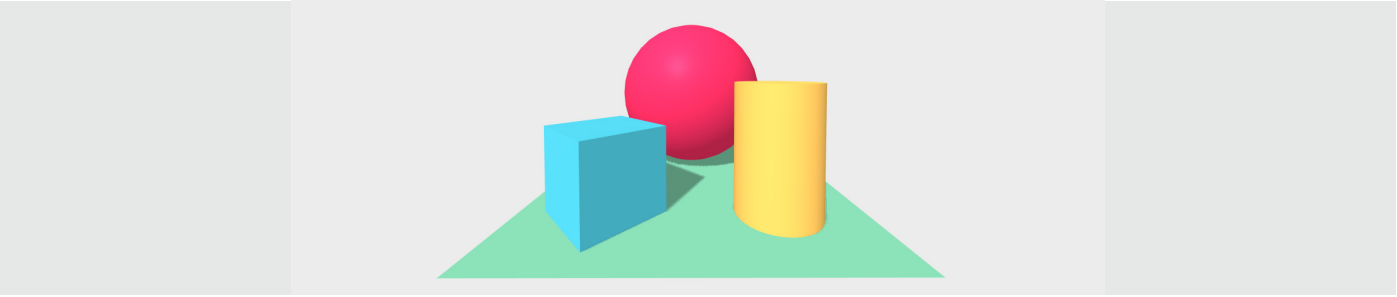


Figure 12. Example of HTML code with A-Frame, which, once opened in a browser, looks like the bottom image.

FILE STRUCTURE

For a basic experience and by only using the elements included in A-Frame, there is just an HTML file, like in figure 12.
As soon as local files and Resonance Audio are added, the project goes from a single HTML page to a more elaborate folder. Figure 13 is a simplified version of what the folder of the prototype ended up looking like. To edit what the experience shows, there is the HTML file. To make resonance audio work in A-Frame, there is a javascript file with the code from Google. With the help of a TNO employee, that was also where the gain, sound width, roll-off and directivity pattern could be changed for each sound source. That javascript file has more than 8000 lines, but by looking through them, several other options can be modified such as the intensity of the reverberation. The extra script.js was to provide more control over when the sounds started. The rest are the files that the HTML calls for. Images, sounds, videos and models for the speakers.

Models	30-1-2019 12:04	Bestandsmap	
Forrest (with visuals).html	7-8-2018 13:32	Chrome HTML Do...	5 kB
aframe-resonance-audio-component.js	7-8-2018 00:14	JavaScript-bestand	769 kB
script.js	7-8-2018 00:14	JavaScript-bestand	1 kB
black.jpg	10-7-2018 19:13	JPG-bestand	26 kB
forest.jpg	7-8-2018 00:09	JPG-bestand	2.617 kB
Music video.jpg	10-7-2018 18:37	JPG-bestand	387 kB
Amazing Natural Bird Sounds (2).mp3	6-8-2018 13:01	MP3-bestand	2.830 kB
bird.mp3	6-8-2018 13:15	MP3-bestand	1.493 kB
musicleft.mp3	18-7-2018 14:21	MP3-bestand	1.184 kB
musicright.mp3	18-7-2018 14:21	MP3-bestand	1.184 kB
narrator.mp3	27-7-2018 06:25	MP3-bestand	5.860 kB
stream.mp3	10-7-2018 20:44	MP3-bestand	5.365 kB
venomleft.mp3	27-7-2018 06:26	MP3-bestand	5.860 kB
venomright.mp3	27-7-2018 06:26	MP3-bestand	5.860 kB
wind.mp3	11-7-2018 00:54	MP3-bestand	11.721 kB
venomtrailer.mp4	6-8-2018 13:29	MP4 Video File (VL...	6.647 kB

Figure 13. Folder structure of the first prototype

THE JAVASCRIPT FILES

Bellow are the relevant parts of the javascript file aframe-resonance-audio-component.js. The first part allows to set the gain and directivity pattern. If a sound source, like a speaker, has a directivity pattern in real life, it can be edited here to mimic that. The gain was helpful to mix the experience realistically.
The second part is an example of the other settings available in the file. In this case the reverb gain could be amplified which would increase the loudness of the reverberation of the environment. The default is set to be realistic while it could also be totally removed. Most other lines in the code are to make Resonance Audio's features work, and some set default settings if not specified in the HTML file.

Underneath it, is the javascript file "script.js" which was a custom code required to get more control than what A-Frame offered. Since the audio files for the trailer were separated from the visuals of the trailer, the video and sound had to be synchronised somehow. The trailer does not start directly

```
setTimeout(function(){
    document.getElementById("trailer").pause();
}, 0);
setTimeout(function(){
    document.getElementById("trailer").play();
}, 54700);

window.setTimeout(function () {
    console.log('play audio');
    document.querySelectorAll('a-resonance-audio-room').forEach(function(room) {
        room.object3D.el.audioElement.play();
    });
}, 7000);

var el = document.querySelector("#voice1"); el.setAttribute("visible",false);
var el = document.querySelector("#teaontableuser1"); el.setAttribute("visible",false);
var el = document.querySelector("#teaontableuser2"); el.setAttribute("visible",false);
var el = document.querySelector("#doorknock"); el.setAttribute("visible",false);
var el = document.querySelector("#voiceatdoor"); el.setAttribute("visible",false);
var el = document.querySelector("#TVleft"); el.setAttribute("visible",false);
var el = document.querySelector("#TVright"); el.setAttribute("visible",false);
var el = document.querySelector("#TVon"); el.setAttribute("visible",false);
Utils.DEFAULT_REVERB_GAIN = 0.035; //default is 0.0035
```

```
// Use with: window['resonance_audiosource']['room1'] or window['resonance_audiosource']['room2'] ...
window['resonance_audiosource'] = window['resonance_audiosource'] || {};
window['resonance_audiosource'][this.el.id] = source;


if (this.el.id == 'voice1') {
    window['resonance_audiosource']['voice1'].setGain(0.75);
    window['resonance_audiosource']['voice1'].setDirectivityPattern(0.5, 0.5)
    window['resonance_audiosource']['voice1'].setRolloff('logarithmic')
    window['resonance_audiosource']['voice1'].setSourceWidth(0);
}

if (this.el.id == 'teaontableuser1') {
    window['resonance_audiosource']['teaontableuser1'].setGain(0.5);
    window['resonance_audiosource']['teaontableuser1'].setDirectivityPattern(0.5, 2)
    window['resonance_audiosource']['teaontableuser1'].setRolloff('logarithmic')
    window['resonance_audiosource']['teaontableuser1'].setSourceWidth(0);
}
```

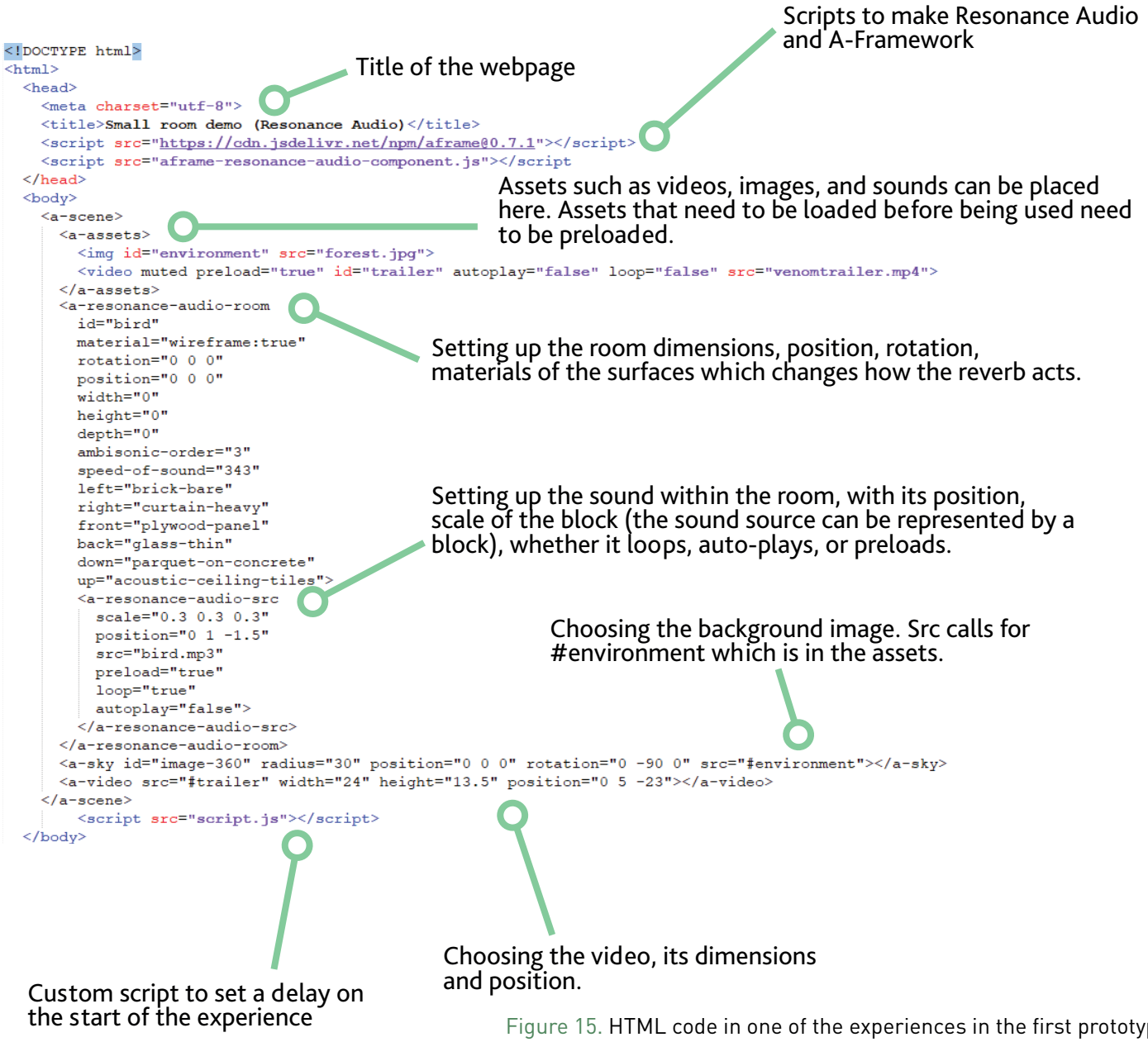
Figure 14. Relevant parts of the javascript code

either, so a delay was added with the first 6 lines so it started at the right moment in the predefined scenario. By tweaking the number, it could be synchronised better with the sound. Another issue was that the left and right audio files for the trailer would start at separate times even though they were preloaded. By adding a delay (setTimeout for Play Audio), it would make sure both files were preloaded long before they were played.

The last lines are to show or hide the sound sources' blocks used to visualize where they were placed in 3D.

THE HTML FILE

The code bellow is what the HTML file looked like with A-Frame using Resonance Audio. This is a simplified version of the first prototype. In the version for the tests there were more sounds.




```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8">
    <title>Small room demo (Resonance Audio)</title>
    <script src="https://cdn.jsdelivr.net/npm/aframe@0.7.1"></script>
    <script src="aframe-resonance-audio-component.js"></script
  </head>
  <body>
    <a-scene>
      <a-assets>
        <img id="environment" src="forest.jpg">
        <video muted preload="true" id="trailer" autoplay="false" loop="false" src="venomtrailer.mp4">
      </a-assets>
      <a-resonance-audio-room
        id="bird"
        material="wireframe:true"
        rotation="0 0 0"
        position="0 0 0"
        width="0"
        height="0"
        depth="0"
        ambisonic-order="3"
        speed-of-sound="343"
        left="brick-bare"
        right="curtain-heavy"
        front="plywood-panel"
        back="glass-thin"
        down="parquet-on-concrete"
        up="acoustic-ceiling-tiles">
        <a-resonance-audio-src
          scale="0.3 0.3 0.3"
          position="0 1 -1.5"
          src="bird.mp3"
          preload="true"
          loop="true"
          autoplay="false">
        </a-resonance-audio-src>
      </a-resonance-audio-room>
      <a-sky id="image-360" radius="30" position="0 0 0" rotation="0 -90 0" src="#environment"></a-sky>
      <a-video src="#trailer" width="24" height="13.5" position="0 5 -23"></a-video>
    </a-scene>
        <script src="script.js"></script>
  </body>
```

Figure 15. HTML code in one of the experiences in the first prototype

THE SOUNDS

For the sounds, the voices and Foley type sounds such as walking, knocking on a door, putting a mug down were recorded with a MXL 770 microphone. The colouration to the sound added by the microphone was somewhat corrected by ear using an equalizer (EQ). Most of the noise was removed using a digital denoiser. An aggressive denoising would result in losing a lot of frequencies which would make the voices especially unnatural, therefore it was used with care. The background noise, such as a stream and wind, were created using files found online and selected based on whether they felt real or not. Each environment had its own scenario, so three

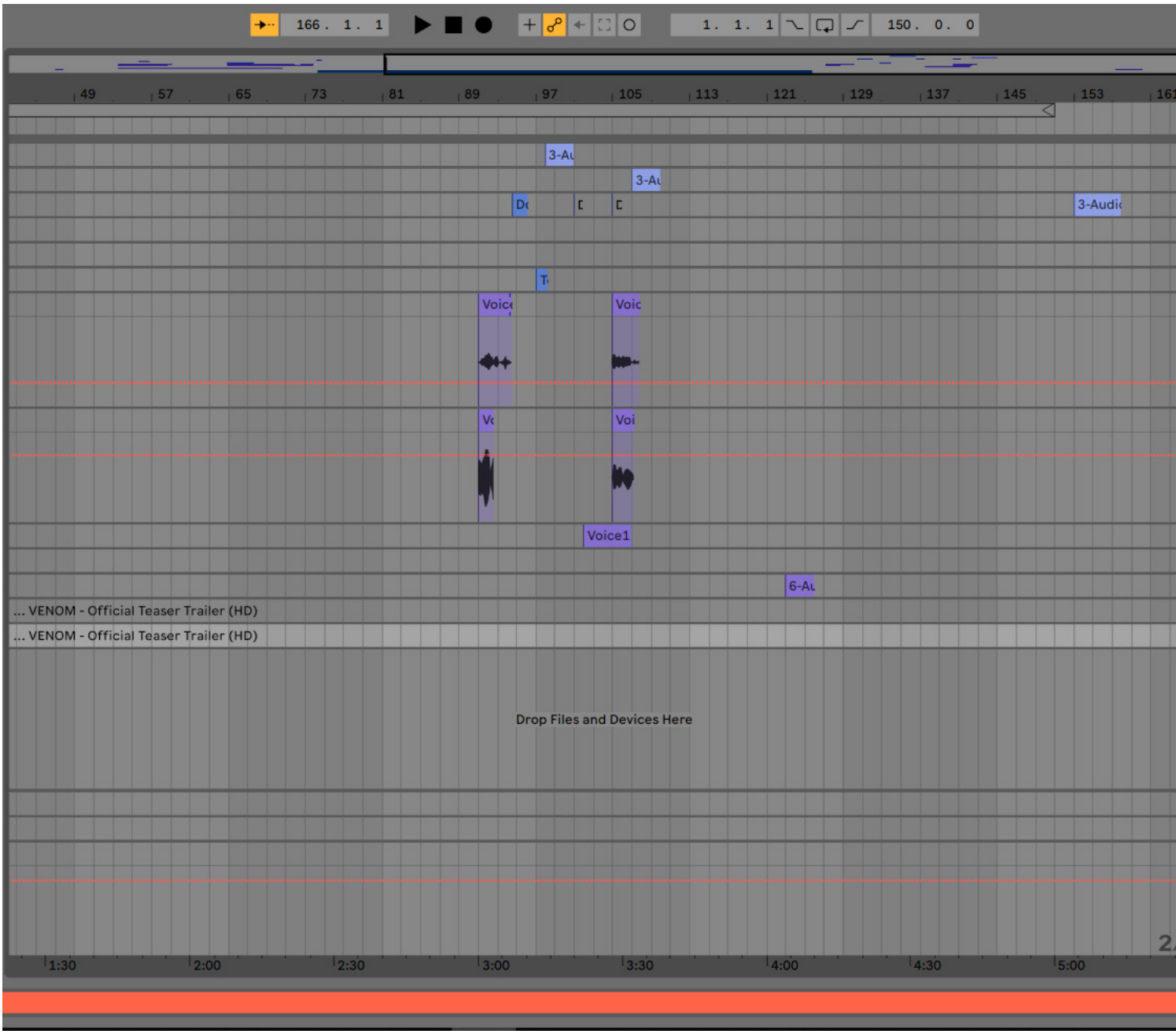
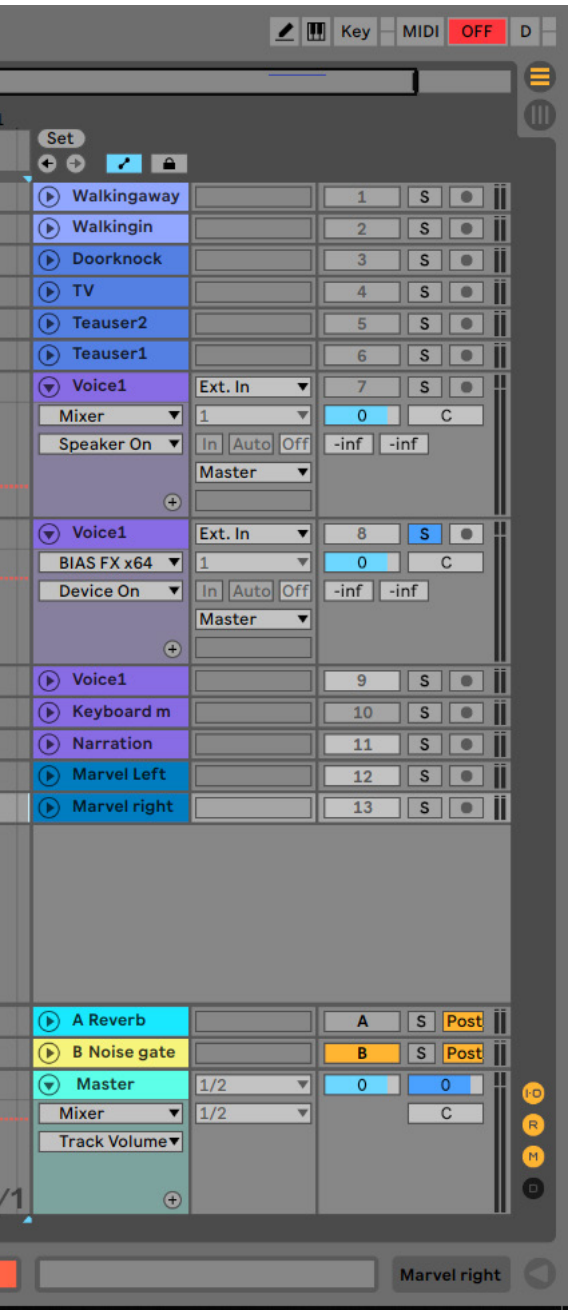


Figure 16. Ableton arrangement for one of the experiences

timelines were created in Ableton Live 10 so the sounds would trigger at the right moment. One timeline is shown in figure 16)



SETUP

COMPLEX SETUP

Next to complicated prototyping, the setup needed to test the experience was complex. There needs to be a powerful computer, an HMD, its trackers, space to move, controllers and a silent room. Installing the setup also took up an hour and utmost care. There were also many different variables that could go wrong.

Launching the experience kept going unexpectedly. For example with the experience not being shown in the headset or not having any sound.

The way the experience was designed, the participant was sitting and in front of them would be a TV from which the sound came from two speakers and the participant would be at sitting height in the virtual environment too. What happened quite often however, was that the room would position itself in the wrong orientation and/or wrong height. Tracking would also be lost and reposition the tester in the wrong place. This was incredibly frustrating during the prototyping, because settings would be tweaked, then checked with the HMD and the experience would be different than what was expected. There were also some bugs within Resonance Audio.

This resulted in a very inconsistent experience which was unwanted for user tests. Because of it, a protocol needed to be figured out with the exact steps to follow to avoid most, if not all, errors. Especially necessary to keep the participant at the right height, at the centre of the virtual environment and facing forwards. This would ensure a consistent and expected experience each time the headset was put on. This is only valid for the HTC Vive.

PROTOCOL TO RESPECT

- Clear out an area of at least 1,5 by 2 meters where the user will be able to move.
- Set the trackers at opposing corners if possible so they cover front/back and each side of the participant. Set them high up to avoid having objects in the way or the person blocking the HMD.
- Make sure the HMD is always visible before even opening SteamVR, that there is not anything in the way of the trackers.
- Connect the HMD and trackers to the computer.
- Launch SteamVR and go through the room setup if the trackers have moved since last time. Select room-scale, trace the area's outline by using the controller and set the HMD on the floor. When it asks to point to your "screen", position yourself on where you want the middle of your room to be, point the direction you want your room to be facing forwards and validate.
- Set down a marker on the physical floor where you validated, this will be the sitting coordinates if the camera coordinates in the demo are 0,0,0.
- Close SteamVR.
- Put the HMD down on the floor facing forwards, towards where you pointed at the "screen".
- Start SteamVR.
- Double click the HTML from the folder and preferably open it in Firefox.
- Click the VR button in the browser to launch it with SteamVR.
- If still not visible in the HMD, move your mouse on the complete right of your screen until you are out of the bounds. Imagine the HMD is a second screen on your computer. Going completely to the right will go into the second screen, click once and now it should display the experience in the headset.
- Because of these steps, you will most likely need to set a delay on when the experience starts.

Side note

If 6DOF is not necessary, the room setup within SteamVR is much simpler. The standing/seated experience can be selected and it only requires the height of the HMD and to indicate which way is forwards.
Adding an arrow on the floor, like in figure 17, helps the participant to position themselves properly.

Figure 17. Screenshot of the arrow placed on the floor to face the right way.

As shown below, the setup of the room used for the user test.

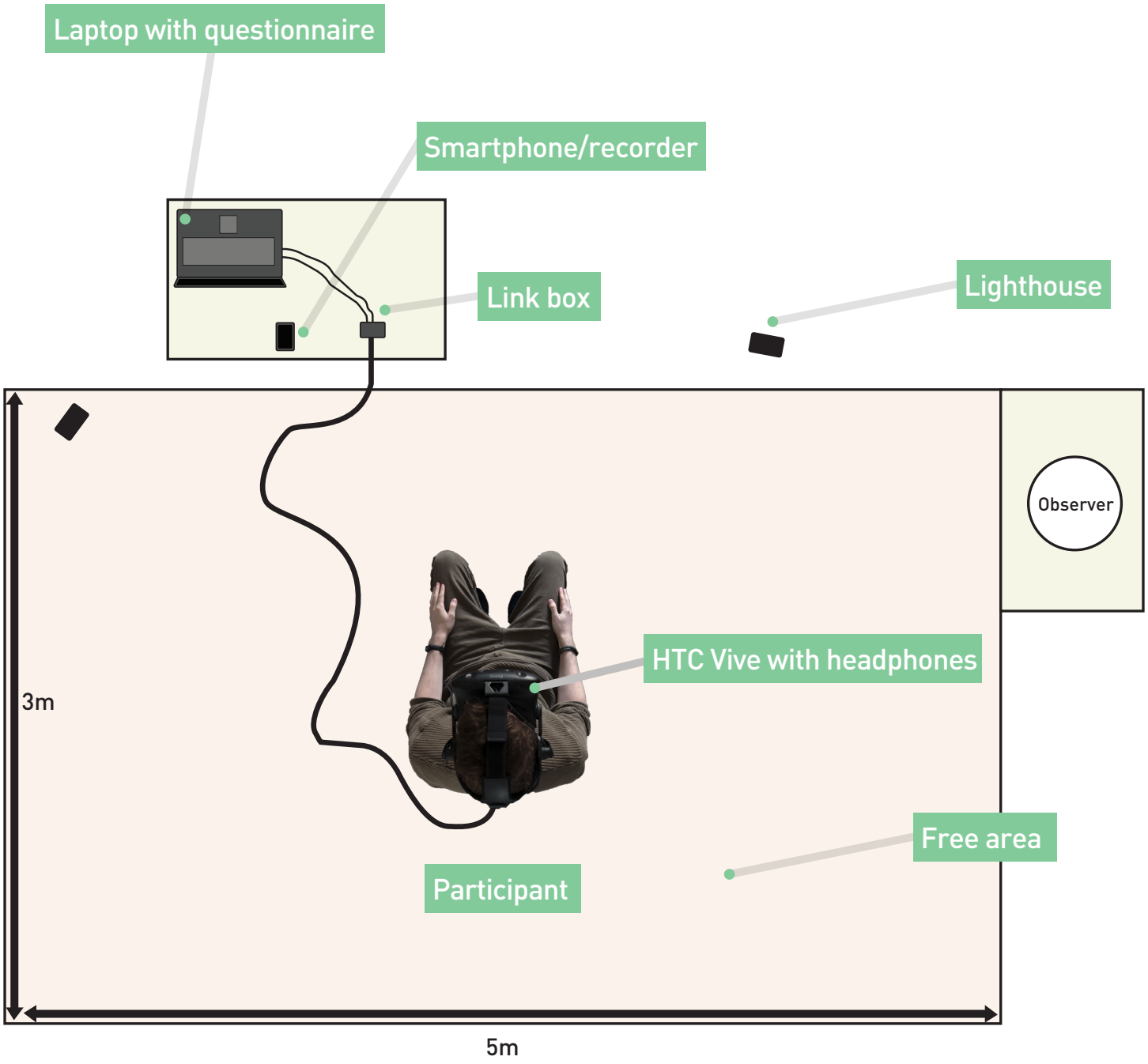


Figure 18. Setup of the room for the user test

USER TEST

Once the preparation and improvements were made to the prototype and setup, it was time for the user tests. It was conducted in a large conference room at TNO, with around 4 by 3 meters of walking space with the HMD. A chair was positioned in the middle of that space facing towards the laptop. The headset was on the chair facing the same direction. The observer was sat on a table noting the remarks or actions of the participant. The laptop had the questionnaire ready to be filled in. The setup can be seen on page 51.

GOALS

This user test had several goals:
- Gain feedback on a basic SocialVR experience
- Compare three different soundscapes to find qualities users prefer
- Determine if and how visuals influence the sound experience
- Determine whether free movement is necessary for the context

SAMPLE SIZE

A total of 5 participants, 2 women and 3 men from 20-65 years old. They were recruited by sending an e-mail to the department at TNO's location in Den Hague.

STRUCTURE

In the user tests, only the first phase (with sound only) had the three environments. The second phase showed the last environment they heard in the first phase. And the third phase also used the same environment. So each participant tested 5 experiences in total. This was a necessary measure to keep the user tests below an hour. The assumption was that the users would not need to experience all three environments again to compare the phases between each other. This would also help the users compare the phases directly since that was the only change, nothing from the soundscape changed.

In figure 19, the phases used during the user test are visualised. With 3 unique soundscapes and 5 participants, the environments were assigned to at least have each unique soundscape be experienced once in each phase.

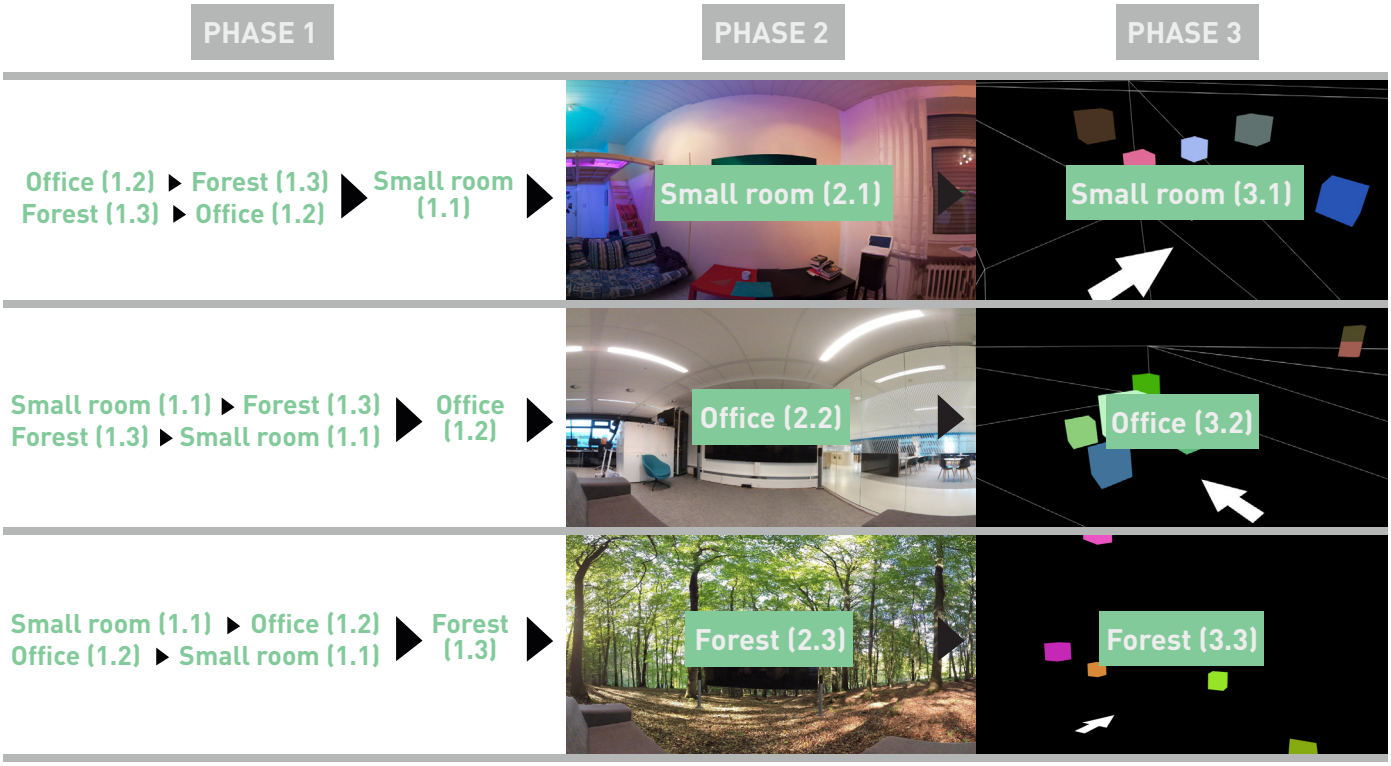


Figure 19. Structure of the user test

METHODS

For this user test, participants were asked to think aloud during the experience. Their actions and spoken words were noted. After each experience, they were asked follow-up questions to either clarify what they said or did during the experiment, or to develop their thoughts while they were still fresh in mind. Then they were asked to fill in a questionnaire rating the different qualities (see Appendix 2.B. for the questionnaire). In total they filled in the questionnaire five times each. All the user tests were also recorded with a microphone, then transcribed and turned into statement cards to list user needs.

QUESTIONNAIRE

The goal of the questionnaire was to rate the experiences based on the three qualities determined in Phase I: Togetherness, immersing and relaxing.

First the definition of these qualities needs to be clarified.

In the context of VR, togetherness is defined by the sense the experience of doing something together in the same shared virtual environment. (Sra, 2018).

The social presence, also known as co-presence, is also part of it. It is the sense of being in a virtual environment with others. Immersion is the feeling of being submerged, in this context, submerged by the environment. As in surrounds or covers the person. The better the presence, the better the immersion.

Presence is defined by the illusion of being in a place despite the knowledge that you are not there. (Cambridge, 2019)

Relaxing is defined as being relaxed, that is feeling happy and comfortable because nothing is worrying you. (Cambridge, 2019).

Togetherness and immersion can be broken up in multiple questions. Based on a paper (Sra, 2018) and my own experience, the following questions were formulated:

Rate togetherness from 1 to 5:

- You felt like you were with another person
- You interacted with another person
- Another person interacted with you
- You felt connected to the other person

Rate presence from 1 to 5:

- You felt like you were physically in that environment
- You remember it as images rather than somewhere you have been
- It felt more like being in a virtual environment than being elsewhere in person

Rate relaxing from 1 to 5:

- I felt relaxed during the experiment

Each of these were rated from 1 to 5, from strongly disagree to strongly agree. A 5 point scale was used rather than a 7 point scale, because John Beerends advised using a shorter scale as during his career he observed people would tend to avoid rating 1 or 7 on a 7 point scale.

QUANTITATIVE DATA ANALYSIS

First the quantitative data was analysed to see if there was any outlier or to find which criteria needed a drastic improvement for the next iteration. Then the qualitative data that was collected by recording the audio of each session combined with the observations was analysed thoroughly.
Considering the low sample size, the qualitative data was leading.

GOALS

The first goal was to analyse whether and how the environments influenced the chosen qualities.
The second goal was to analyse whether and how adding visuals and movement influenced the chosen qualities (dependent variables).
The hypothesis is that both the phase and environment influences the qualities.

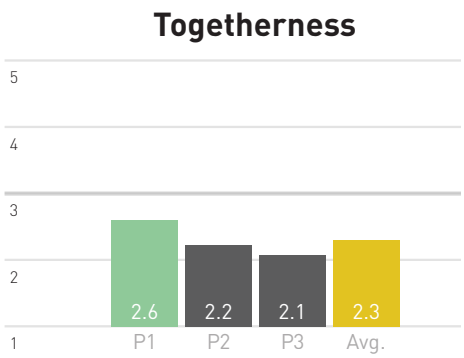
METHODS

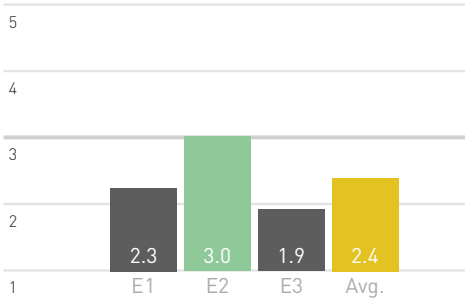
A questionnaire was formulated with 9 questions to rate togetherness, immersion and relaxedness. It was filled in after each experiment. The data was analysed in SPSS.

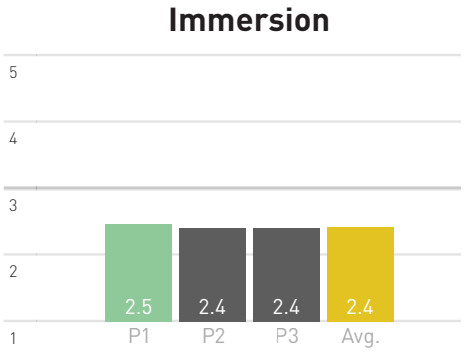
RAW DATA IN APPENDIX 2.C.

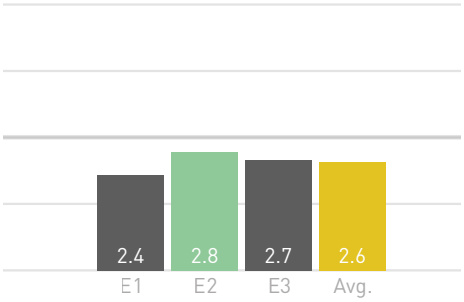
RESULTS

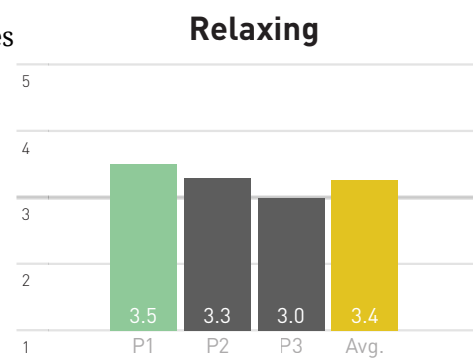
To make the most out of the data, it was split to look into the differences between phase 1, 2 and 3. Then split to look into the differences between the three environments (small room, office, forest).

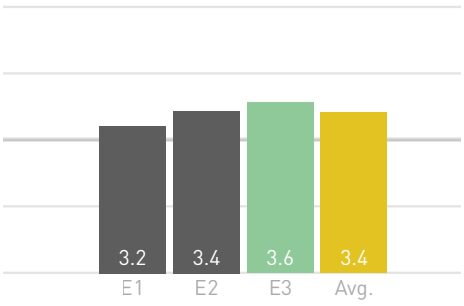












CONCLUSION

Differences between phase 1,2,3

Togetherness scored higher when only the sound was used. This can be explained by being able to imagine the visuals yourself. In both phase 2 and 3 however, the participant could now see whether the visuals match their expectations.
Whether the environment was relaxing or not was better when they only heard sounds. The assumption is that since they only had to focus on hearing, the experience was less overwhelming and thus more relaxing.

Differences between environment 1,2,3

Togetherness had a significant peak for the office and a low point for the forest. One aspect that could explain the office's score is that the reverberation gain was more pronounced while the forest did not have any room effects. It also had a couch that fit in naturally. For the social aspect, room effects seemed to improve the experience.
The forest scoring a little higher than the other environments for relaxing can be explained by the constant sounds in the background. It could have been more significant if the trailer of the movie Venom used for the forest had the same tone as the Ant-Man and Avengers trailers. Although they were all superhero movies with the same genre, the trailer for Venom was more suspenseful.

Differences between the qualities

The best scoring quality was "relaxing", scoring consistently above or equal to 3.0. Immersion and togetherness both scored lower or equal to 3.0. Togetherness was expected to score low because of the lack of interactivity, but immersion ended up also scoring low. The lack of difference between the phases means that the cognitive dissonance did not change the immersion. Or that visuals increased the immersion on some points while decreasing it because of the dissonance.

QUALITATIVE DATA ANALYSIS

GOAL

The goal was to determine user needs based on their feedback during the test.

METHODS

A similar approach as statement cards in context mapping was used. All the recordings of the user tests were transcribed and, the most relevant and interesting quotes were selected and written down on post-its, to be placed on a wall (see figure 20). They were then categorised in clusters, themselves placed within a theme.

Then these statements were translated into user needs for the VR experience of watching content together.

Each participant had their own colour:
Participant 1 was orange
Participant 2 was blue
Participant 3 was yellow
Participant 4 was green
Participant 5 was pink

Each post-it also indicated the environment and phase they were in at the moment of the quote, indicated as (number).(number). The first number is either 1 (sound only), 2 (with visuals) or 3 (blocks). The second number is the environment, either 1 for the small room, 2 for the office, or 3 for the forest.

RAW DATA IN APPENDIX 2.D.

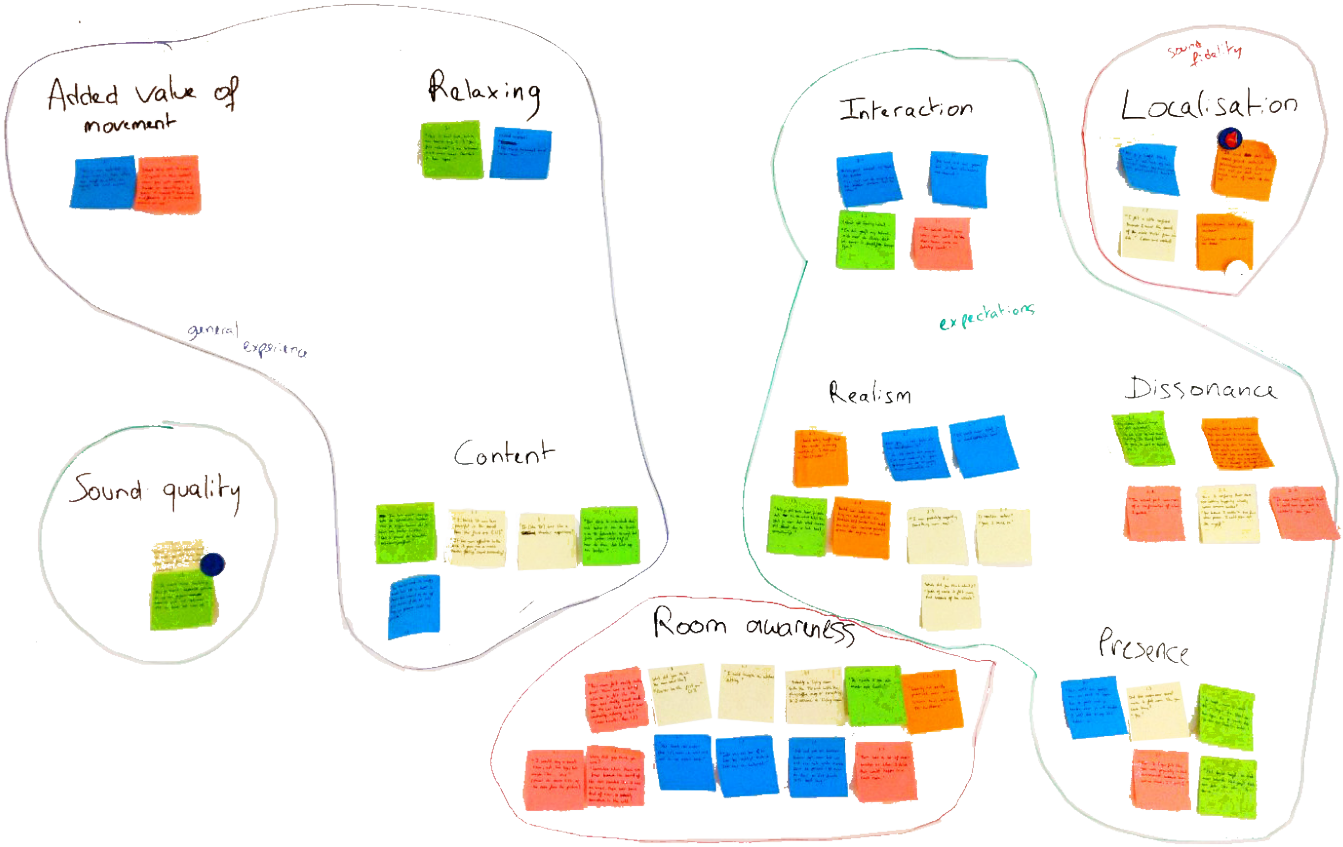


Figure 20. Picture of the wall with statement cards

RESULTS

Added value of movement

Two participants specifically mentioned that for the context of watching TV, it would not matter if they could walk around or not.

Relaxing

One participant mentioned the experience while moving around was fun yet scary, and responded to the question of "you felt relaxed" that it was not true at all anymore because of the walking, and the fact that he could hit anything at any moment. Another participant mentioned one of the trailers as being scary (the Venom trailer) and she expected something to fly at her since it is VR. One of the tests was louder than the others which startled her.

Interaction

Once the participant was able to move around and see 3D blocks float around, most of them reached out and try to interact with it. They also sensed there was no interaction with the other person in the virtual room because they could not hear them in real time, or see them either. They could only hear the person, not talk to them. Also when movement was implied in the scenario, such as walking to the door, the participant expected footsteps.

Localization

The participants were able to accurately tell where sounds were coming from even with the deluxe audio strap's headphones and basic HRTF. If a sound came from above, they would notice that too. No comments about sounds coming from behind or under.

Sound quality

Two participants were especially bothered by the noise switching between the audio clips which was not hidden enough by the background noises. The recording of the voices and sounds in the video all included noise in the background which made it apparent when the sample started and when it stopped.

Content

The content regroups everything the participants had to say about the trailer being shown and/or heard. One of the participants, John Beerends, mentioned the "discolouration effect" with the trailer. According to him, this effect is what we do naturally with rooms we are in. If speakers are playing in a room, the room will affect the sound, our brain will discolour it. Because of this, adding virtual room effects on speakers is too imperfect for our brains to discolour it naturally. However, for a mono audio source such as a voice, it works decently. Another person commented on the TV feeling like a theatre experience, which she associated with the sound feeling powerful.

Room awareness

The participants were able to tell the difference between open and closed rooms based on the sound cues and room acoustics differences. However, there was some confusion about the office sounding smaller than the small room.

Presence

Participants felt less presence when they had to be conscious of external factors such as the cable or knowing that they were walking in a real room with obstacles. Having visuals improved presence. Continuous sounds also helped to make it feel more like they were there.

Dissonance

When the sound moved, but no object moved and vice versa, it created a cognitive dissonance for the participants. Also related to presence, when the users could not see themselves it creates a detachment from their virtual self.

Realism

Three users mentioned finding realism to be better. This is also related to cognitive dissonance between the visuals and the sound. Adding visuals made it more realistic. No user mentioned adding realism would be worse.

CONCLUSION

The conclusions were then combined on a whiteboard (figure 21).

Freedom of movement

It should be a seated experience.

Communication

The recordings/transmissions should be smooth, without clear

DISCUSSION

transitions.

Expectations caused by reality

If an object makes a sound, it should be visible.
If an object is within reach, it is expected to be able to interact with it.
There should be continuous background noises.
Environment should be 3D (not a static image) to avoid dizziness.

Content

The type of content influences the experience.
It should have a movie theatre feeling.
Leave out room effects on TV.

Dynamism

If an object moves, it should be heard.

Sound fidelity

The soundscape should be good enough to understand the environment and where sounds come from.
Although the project focused more on sound, the influence of the visuals on the overall experience was too significant to ignore in the next iterations. Meeting the basic requirements for the visuals would help to improve the sound experience too.

DISCUSSION FROM BOTH DATA

Overall, both quantitative and qualitative data point that there was a lack of interactivity. This needed to be addressed in next prototype.
There needed to be neutral TV content as not to influence the quantitative data as it was assumed it had influence on some answers this user test.
It was also assumed that the addition of basic visuals improved the experience on some aspects while decreasing it on others. It needed to be at the same level as the sound design to improve the experience.
From analysing the scores, there was not a clear indication which type of environment is preferred. But by going through the feedback, it seemed that adding room ambiance to the trailer was detrimental to the intelligibility of it and did not give it a “cinema” experience. Next to that, the assumption was that although people prefer the cinema experience for the content, they prefer the closeness provided by being in a private room. This also corroborates with the comments from the questionnaires from Phase I, about the experience in BigScreen.
Until now, all the sounds were put in a single room with the reverberation and ambiance from that room. But it would be beneficial to separate the TV sounds from the voices.

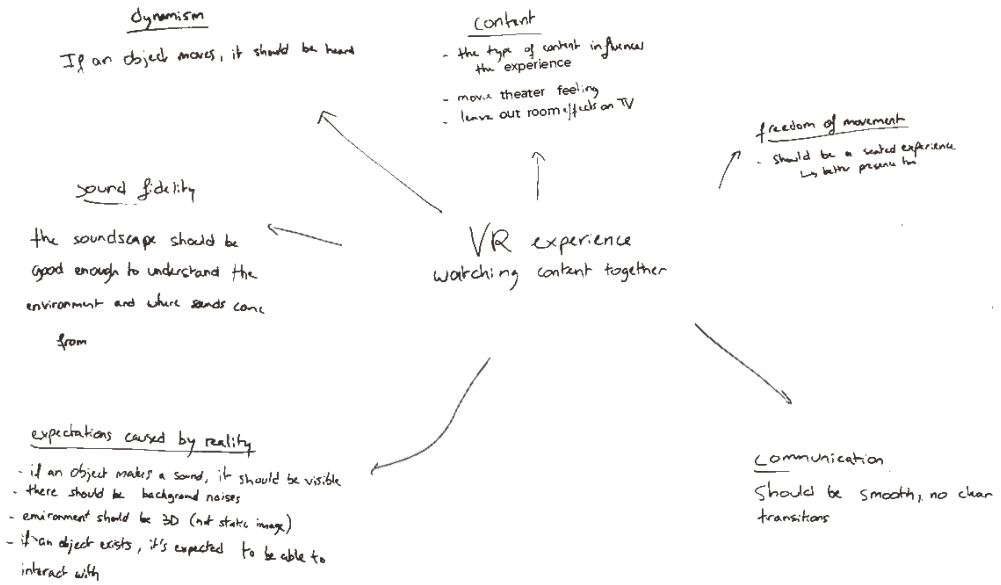


Figure 21. Picture of the wall with user needs

ENVISIONED ENVIRONMENT

This leads to two environments in one, where the sound experience of the content is separated from the rest (figure 22).

The content

Cinema experience. The cinema experience is not only a large number of speakers placed around the user, it is also removing reverberation. The sound coming from the speakers is not mudded by the reflections.

The communication and room ambiance

Home experience. Being in a room with privacy feels cosy and comfortable. Seeing the room would help associate those emotions to the experience and hearing the room effects on the voices would sound natural.

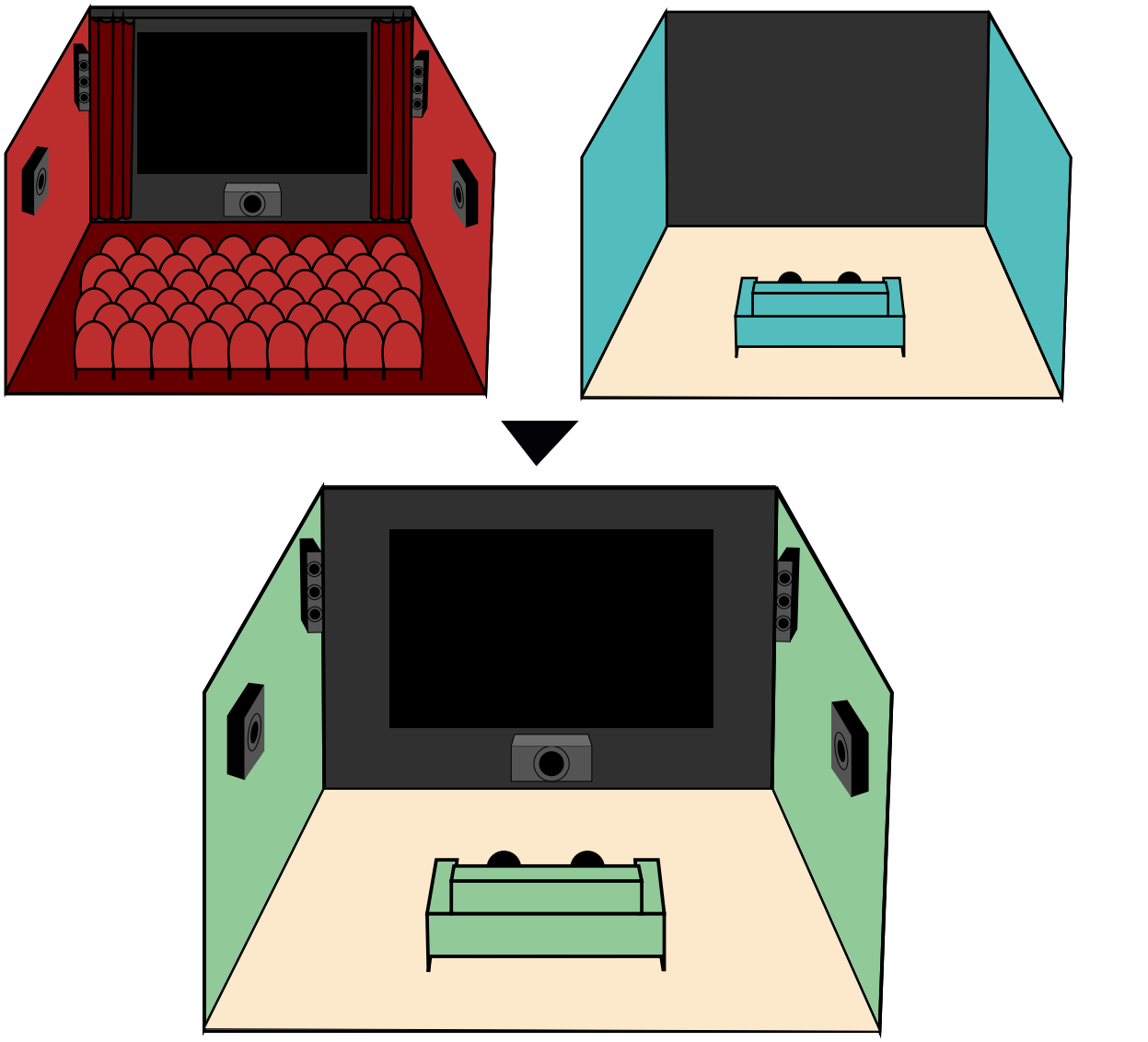


Figure 22. Illustration of the envisioned environment

MAIN FINDINGS

The findings during this phase will be listed here.

SOUND AND VISUALS

In the context of watching TV together in VR, there is no user need to be able to walk around.

The type of content influences the experience.

Noise switching in the sound quality can significantly take the person out of immersion.

Users will try actions they can do in real life such as grabbing objects within reach.

Users expect certain sound cues based on implied movement.

If a sound moves yet cannot be seen, it creates a cognitive dissonance.

Localization with Resonance Audio was acceptable for left/right and down/up.

Reverb on the stereo TV speakers prevents the “discolouration effect”.

Users can pick-up on differences in reverberation between open and closed environments.

The reverberation settings from Resonance Audio are not accurate or realistic for the room sizes.

Presence in VR is partially negatively influenced by real life obstructions such as cables and the headset.

Presence in VR is influenced by the virtual representation of the user.

Adding visuals made it more realistic.

Realism is better.

Testing the sound experience requires visuals.

USER TESTING

These findings are about the user testing itself, and how to improve it. They are either based on observations during the experiment, from setting it up, or based on the participants’ feedback.

It is essential to come up with a list of steps to follow to provide a consistent experience.

Having an arrow on the virtual room to indicate the direction to look at helped to avoid having them disoriented.

The interviewer not being able to see the participant in the environment made it more difficult to understand what they were trying to accomplish.

Asking participants to think out loud while they listen to a sound experience does not work.

Stepping in and out of VR to fill in a questionnaire takes time.

Stepping in and out of VR requires participant to readjust their HMD instead of having the same settings the whole time.

Adding a delay before the experiment starts prevented for it to start too early while the participant was still setting up their HMD.

Not providing a stable 90 frames per second experience can induce dizziness.

It would be more valuable to see the scores given by the participant in real time to ask follow up questions.

After the experiences, asking them to describe how it felt gives them an opportunity to provide a lot of feedback.

Participants noticed when the sound of the trailer was not as loud in another environment.

CONCLUSION

In this phase, a first prototype was made for a basic Social VR experience. The feedback showed what kind of experience is necessary to test the sound experience further, which includes having convincing visuals. It also indicated that the next prototype would need to have real-time interactions between the players to adequately emulate a Social VR setting in which two players watch TV together. Because of the user testing, an envisioned environment could already be made to help with the sound experience requirements for the next prototype.

PHASE III: ADVANCED SOCIAL VR PROTOTYPE

INTRODUCTION

PHASE III: ADVANCED SOCIAL VR PROTOTYPE

Following through on the envisioned environment, in this last phase, the prototype will be improved based on the previous research. By combining all previous feedback, and testing a final prototype, guidelines can be written for the sound experience in Social VR. The goal of these guidelines is to help VR developers think about their sound design in Social VR, and show them recommendations based on user feedback. Aside from recommendations around sound and visuals, there will be its own section about user testing Social VR. Throughout this project, different ways to test Social VR needed to be discovered as of the start of the project, not many resources were about Social VR experiences.

To serve these guidelines, a final user test was conducted. In Phase II, a basic Social VR was created in WebVR. Though for a more complicated experience with live interaction for example, it was too difficult for a beginner to use WebVR for the prototyping. The solution was to transition to Unity, a game engine which can be used for free, depending on the project. Although there is another popular alternative called Unreal Engine, the plugin Resonance Audio was better implemented in Unity which is why it was chosen. Next to that, Unity offers third party plugins and tools on their Unity store. With its popularity with beginning developers, there are also many resources and tutorials to get started with Unity. Compared to WebVR, it is more accessible for beginners, and it gives more options.

That being said, in this case, it required extensive research and coding to be able to make a multiplayer experience with real-time communication and interaction.

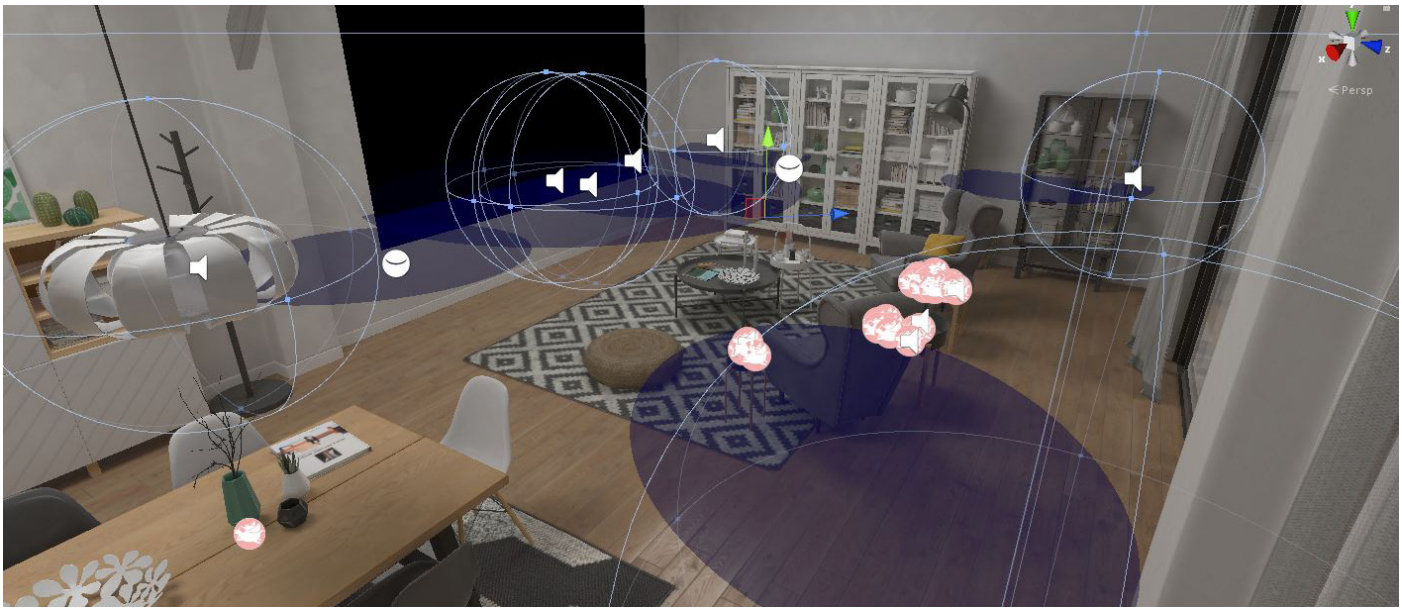


Figure 23. Teaser picture of the prototype

PROTOTYPE V2

The prototype was made in Unity, using a sample project called PlayoVR as a stepping stone. This project introduces basic Social VR features: simple avatars and interactable objects that work through a network. The code was modified significantly to create the wanted experience.

ENVIRONMENT REQUIREMENTS

As established following the user feedback of the first experience, the environment had to be open and familiar for the context of watching television with a friend. It also needed a background noise and a place to display the TV.

From a technical point of view, it also required to be suited for VR performance. This means a relatively low amount of polygons and control over the graphical settings. The former was concluded from a previous VR project where the polygon amount of the model used was too high, and performed poorly in VR. The latter is necessary to create different versions of the experience for laptops and desktops.
After searching for environments that would fit the established requirements, the asset from ArchViz Pro Interior Vol.6 was selected. It can be found on the Unity Asset Store, it performs well in VR, and it provided different rooms that could be used for the experiment. The layout of the house can be seen in figure 24.

ENVIRONMENT

The living room was slightly rearranged and a wooden asset was edited out to provide enough place on the wall for a big screen. Tables were added besides the chairs for the controls. See figure 25, 26 & 27.

The two chairs are facing the wall and the TV can drop down from the wide black bar on the wall.

The two players are meant to sit in the chairs. The left one, for the purposes of the user test, is meant for the participant while the other chair is for the interviewer.

You can experience a demo of this prototype by scanning (or clicking on) the following QR code.

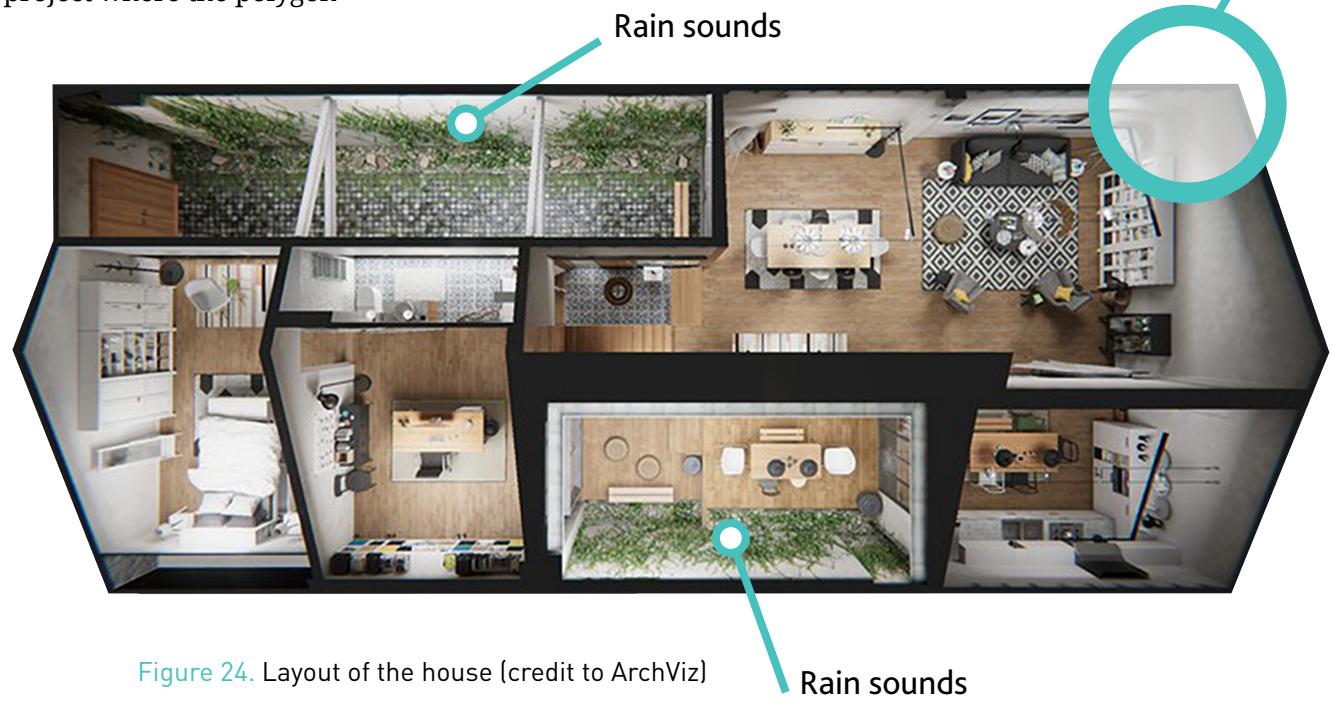


Figure 24. Layout of the house (credit to ArchViz)

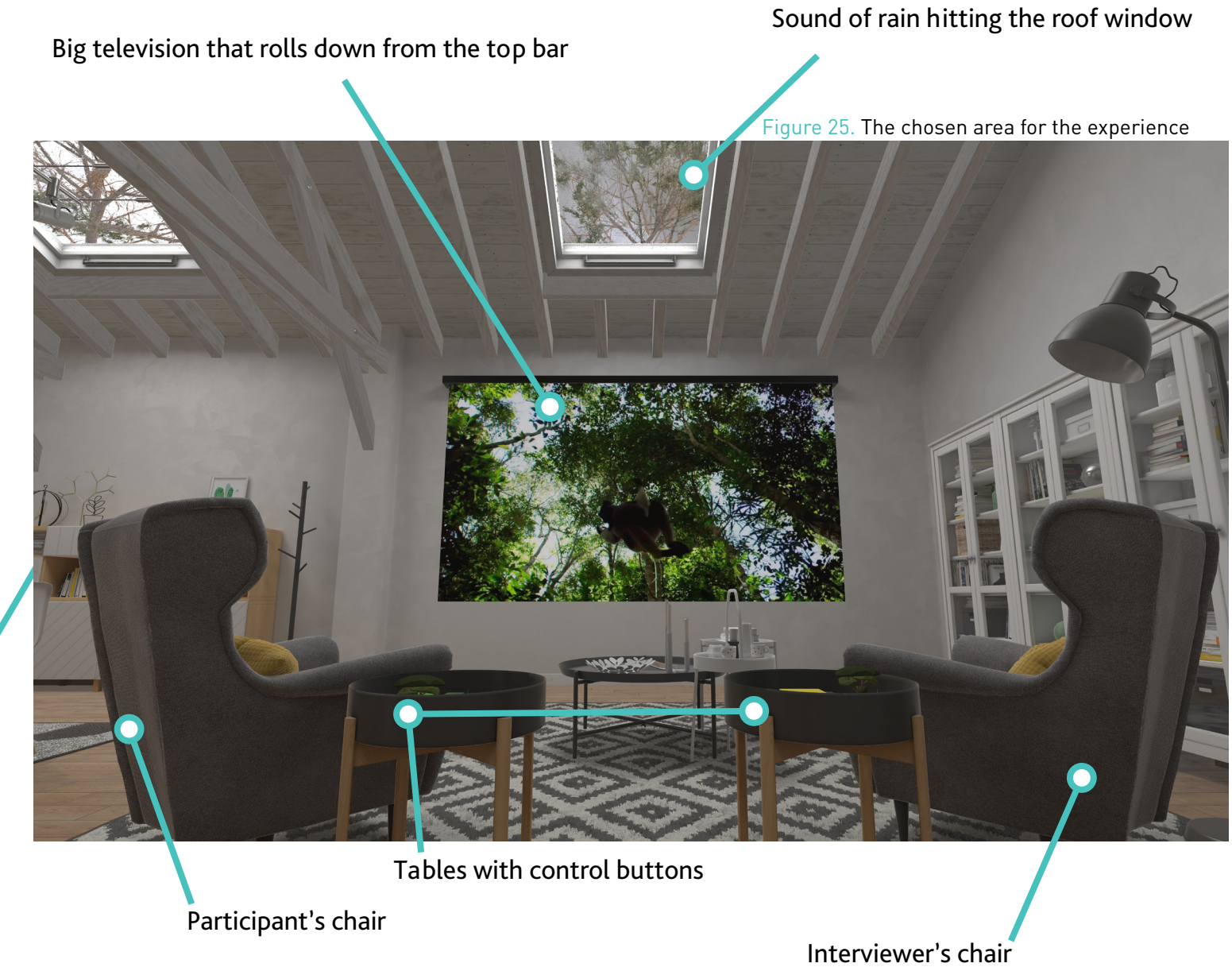


Figure 25. The chosen area for the experience

Figure 26. Television screen rolled up

Figure 27. Front view of the chairs

Figure 28. View of the participant's virtual space with two tables on the chair's sides

CONTROLS

There are a number of buttons on the left and right side of the chairs (see figure 28). The ideal sequence the participant could follow was to first lower the TV, then select a theme by grabbing one of the three 3D models. These models of a submarine, rock, and tree let them choose a video. Each was a different extract of a nature documentary with a theme related to the model. Then the play button could be activated. See figure 29 for all the buttons' explanations. The only differences in available buttons is that the participant has more throwable objects on their left side, and that the interviewer has cubes to control the reverb. These settings change the reverberation on the voice of the interviewer. The default one is the medium setting. The other settings

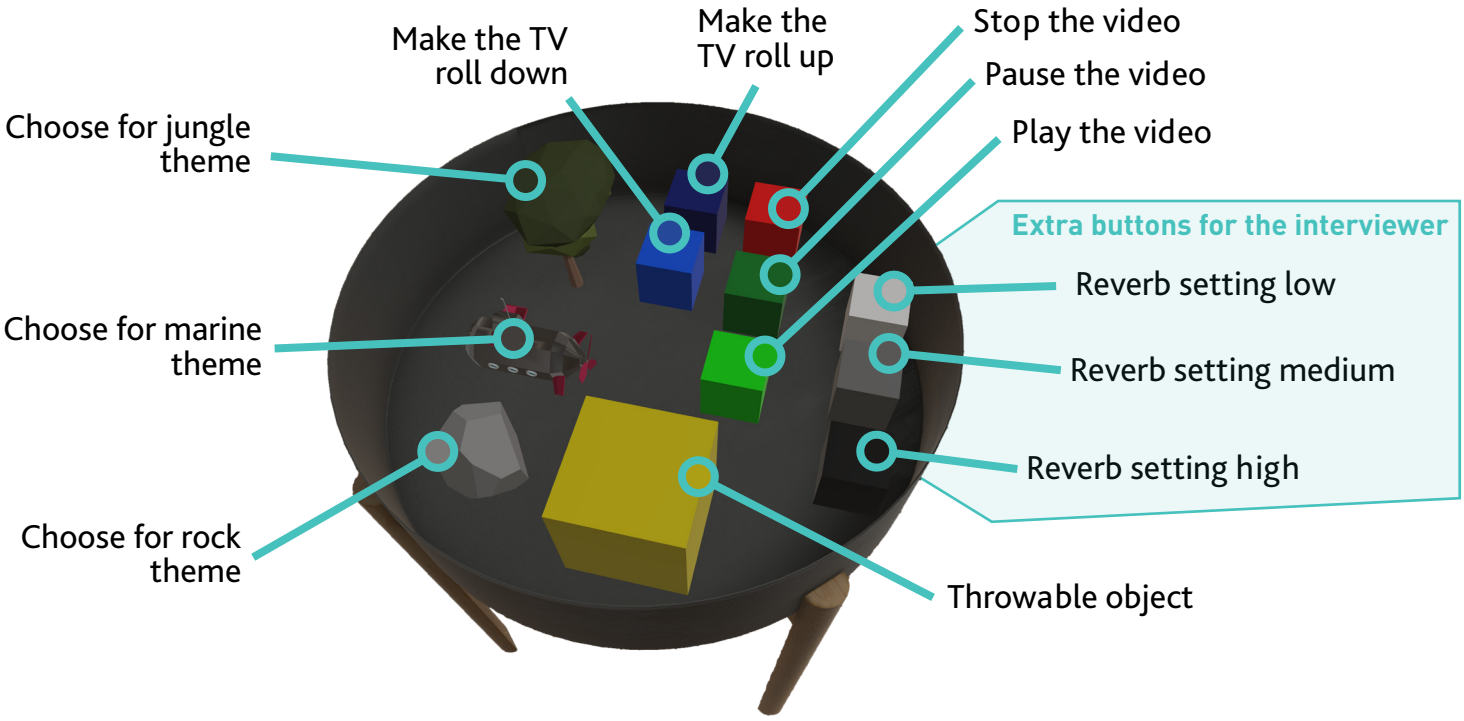


Figure 29. Image of the interviewer's buttons

are exaggerated changes to turn the amount of reverberation up or down. This way the difference is noticeable, and helps to find user preferences.

To grab an item, the user needs to press both buttons on the side of the controller, like grabbing a ping pong racket. To activate the button's action, they need to keep grabbing the sides while pressing the trigger down (see figure 30).

AVATARS

The players were represented by spheres for their hands and a sphere with a headset on for their head (figure 31). All of which follow the player's real movements.
If the player was looking at their own hands, it was represented by the Vive controllers they were holding in real life. This made it easier to see what they were grabbing. And the headset became transparent if it was their own.

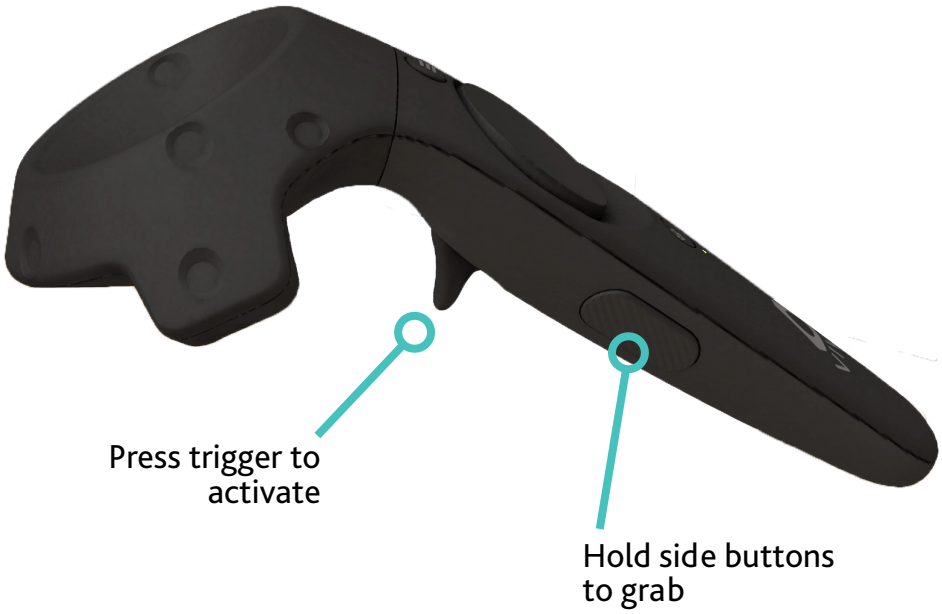


Figure 30. Controller with button explanation

Figure 31. Image of the avatars

SOUNDSCAPE

Figure 32. Image of one of the roof windows

Once the environment was visually set, the soundscape could be created. The background noise was chosen to be rain. The rain was positioned in 3D in two large outside spots as seen in figure 24.
While a specific sound of rain hitting a roof window was added on the two roof windows above the room (figure 32).
This combination resulted in a realistic ambient noise, but also a very familiar sound coming from the windows to enhance immersion.

A sound was also added to the buttons to select a theme in order to indicate if it was properly selected as those were the only buttons that still needed feedback.

The TV

Since previous test results and questionnaires, the wanted sound experience for the TV was that of a cinema. To do so, a 5.1 system was virtually installed in the room. At the front, a left, centre and right speaker with a subwoofer. At the back, a left and right speaker.
These virtual speakers can be seen in figure 33.
Since they are placed in 3D, it creates the illusion there are actually speakers placed around the room when they turn their head.
The video contents for the TV were originally in 5.1. Each channel was extracted to a mono sound and each virtual speaker played its corresponding channel. The speakers were not visualised by an object. They all were assigned a directivity that was close to what speakers do in real life.

The content itself was chosen to be neutral. There were three options for the participants to choose from, all documentaries from BBC's Blue Planet II or Planet Earth II. A short segment of around 3 minutes was edited. The segments were chosen carefully to convey the same emotions.

When the TV went up/down it made a similar sound as a motor.

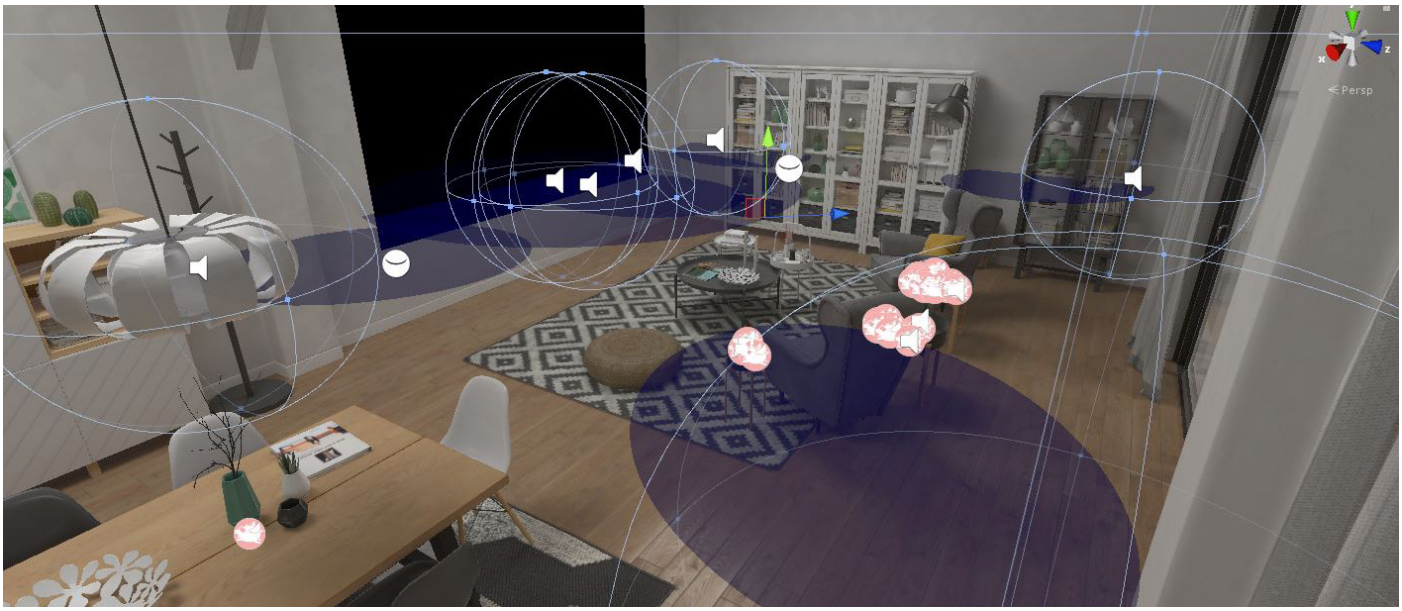
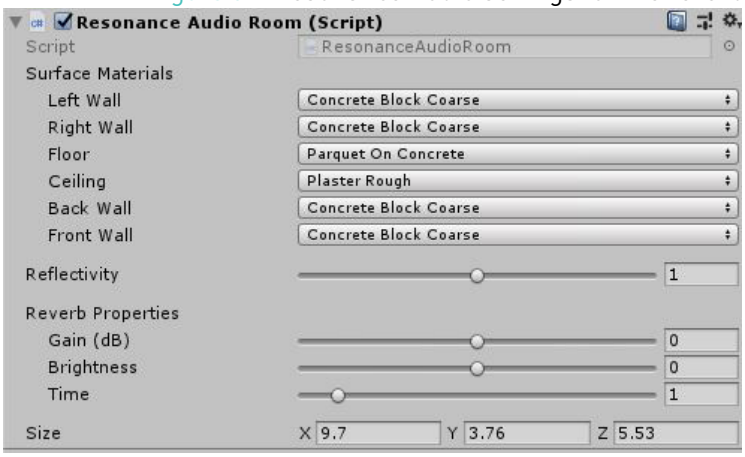


Figure 33. Representation of the virtual 5.1 system with the directivity of the speakers as shown in Unity

Reverb

In the first phase different ways to create a realistic reverberation were found. In this case, the accuracy of the reverberation was not of utmost importance. Resonance Audio provides its own shoebox model to recreate a simple model to create reverberation.
Next to setting the dimensions of the room, different materials can be selected for the walls, ceiling, and floor. Since these surfaces absorb sound and reflect them, it makes a more accurate model. See figure 34 for all the settings.

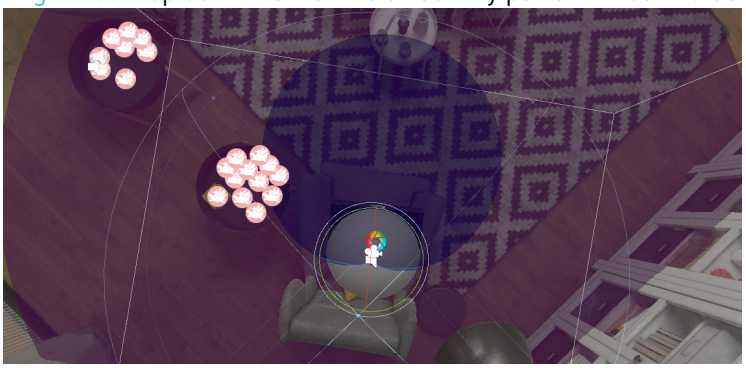
Figure 34. Resonance Audio settings for the reverb



The voices

Each player's voice is directional; its pattern shape is based as closely as possible to the natural pattern of human voices (Chu, 2002). That direction and the source's position is tracked to the HMD's rotation/position. This means that when the person speaking looks away from the other person, it will sound less loud. (See figure 35)
Since it is 3D audio, if the player moves far away, the volume will decrease too. And the sound source matches were the player's head is. Reverb is added to the tester's voice.

Figure 35. Top down view of the directivity pattern in dark blue



Latency

Considering the voices had to go from one user to the other user, different ways were tested: through a network or wired.

Either options result in a latency. In communication, the delay it takes for one sound to go to the other user and then going back is called a round trip latency. It is recommended to have a round trip value of 150ms and not to exceed 250ms (VOIP-INFO, 2018).

To measure the one-way trip, I snapped my fingers near the microphone and let it play back through speakers near the microphone too. By recording it with another microphone, the delay could be measured, as shown in figure 36.
All following measures are one-way, the round trip latency is twice that.

Wired

Using Unity's audio engine is considerably slow. With a simple script, the latency was consistently at 160ms or slightly above. Too high for sidetone, and still slightly too high for communication according to the recommended 125ms.
Using a paid asset from Unity, the same latency was recorded.
And finally, using the trial version of Audiostack, the latency became 50ms since it bypassed Unity's audio engine and made full use of ASIO. This is practical when using an external audio interface to keep input and output latencies short. However, this plugin did not provide directivity and it was only a trial. From then on it was clear that the problem was Unity's audio engine.

Wireless

Transmitting the voice through a network with Photon Voice, the latency hit 300ms and above.

Conclusion

For the test, only a 320ms round-trip latency and above is feasible.

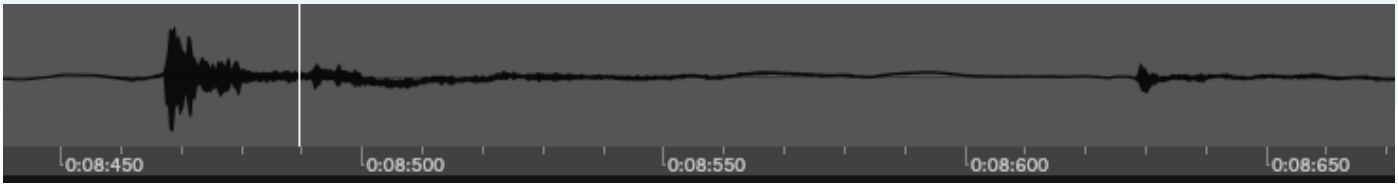


Figure 36. Soundwave of the one-way latency recording in Ableton Live. The time between the peaks is the latency

Figure 37. Front face of the audio interface (credit to progear.ee)

Sidetone

Originally, the sidetone for this prototype was simulated in Ableton Live with EQ'ing (adjusting frequencies), and reverberation, which made it sound more realistic. However, it resulted in latency build-up and crashes when combined with Unity. A more reliable method was to use the built-in monitoring of the Steinberg UR242 which also has reverb effects. See figure 38 for the program to control the settings of the interface. The UR242 is an external audio interface (as pictured above) in which microphones can be plugged in which Unity then uses as input. The headphones can also be plugged in it instead of going through the HMD. Using the built-in monitoring ensured an unperceivable latency and no crashes in Unity. The settings were tweaked based on my own perception of what sounded more natural. This test was done in a large room with the headphones the participants were going to use. With a pilot test they were slightly changed. This solution was acceptable for the goals of the test.

Figure 38. Program to set the reverberation and sidetone of the audio interface

PROTOTYPING IN UNITY

Making the transition from WebVR to Unity, the prototype could not pick up where the previous one left off. It had to be built from the ground up. To shorten that process, a demo was found on GitHub called “PlayoVR” to at least have working networked avatars. And the environment from ArchViz was chosen. From then on, making the experience started.

The first step was to create a way to interact with buttons. The PlayoVR demo already showed grab and activate functions, but not the actions required. Scripts had to be modified to make it possible to activate an action by pressing the trigger of a cube, and sending that action to the server so both players see the same thing happen on their screen. This way a TV remote and reverb buttons could be made. In Unity, the language used for those scripts is C#, as shown on the right. On the right is an image showing how the code looked like to make the TV go up.

Figure 39. Simplified example of the written code for the button to make the TV go up

```
namespace PlayoVR
{
    using UnityEngine;
    using VRTK;
    using UnityEngine.Video;
    using NetBase;

    public class NewMultipleGoneTVSync : MonoBehaviour
    {
        public AudioSource LeftSpeaker1;
        public VideoPlayer Video1;
        public GameObject VideoSystem;
        public Animator TVAnimator;
        public string TVAnimation;
        public AudioSource TVgoingdown;

        [PunRPC]
        void SoundPauseTVGone() {LeftSpeaker1.Stop();}

        [PunRPC]
        void VideoPauseGoneTV() {Video1.Stop();}

        [PunRPC]
        public void DoAnimation()
        {
            if (TVAnimator.GetCurrentAnimatorStateInfo(0).IsName("StayDown"))
            {TVAnimator.Play(TVAnimation);}

            if (TVAnimator.GetCurrentAnimatorStateInfo(0).IsName("Hideblackscreen1"))
            { TVAnimator.Play(TVAnimation); }
        }

        [PunRPC]
        public void SoundPlay()
        {
            if (TVAnimator.GetCurrentAnimatorStateInfo(0).IsName("StayDown"))
            {TVgoingdown.Play();}

            if (TVAnimator.GetCurrentAnimatorStateInfo(0).IsName("Hideblackscreen1"))
            {TVgoingdown.Play();}
        }
    }
}
```

Select a sound source. In this case it had to be done for all 6 speakers for each video.

Lets you choose the video and videosystem

Lets you set the animation for the TV (in this case dropping the TV)

Select a sound source, in this case the sound of the TV going up/down

Stop the speakers' audio if the button is pressed to make the TV go up

Stop the video if the button is pressed

Only let the screen go up if the screen is already down (playing a video or not).

Parts of the code that will be exchanged between players through a network so that the sounds and animations start at the same time on both screens.

Only let the screen make a sound of rolling up if it is already down

Next, creating the soundscape. Resonance Audio was well integrated in Unity which made this easier. The sounds were placed in 3D using Resonance Audio. The speakers and voices both had directivity, and different “reverb rooms” could be made. This way the reverb buttons can basically activate the relevant reverb room and deactivate the others.
The content for the speakers was converted from a 5.1 mp4 video with HandBrake to separate the sound from the video. Then with Audacity, the 5.1 track was split into separate mono tracks so that each virtual speaker could play the appropriate track for its position. Then the communication. This needed a lot of tests to make it work, and to understand the limitations of Unity. The tests determined that a wired solution was better suited, and that the latency would not go under 320ms even wired. The directivity of the voices also had to be connected to the HMD’s position/rotation with a custom script.

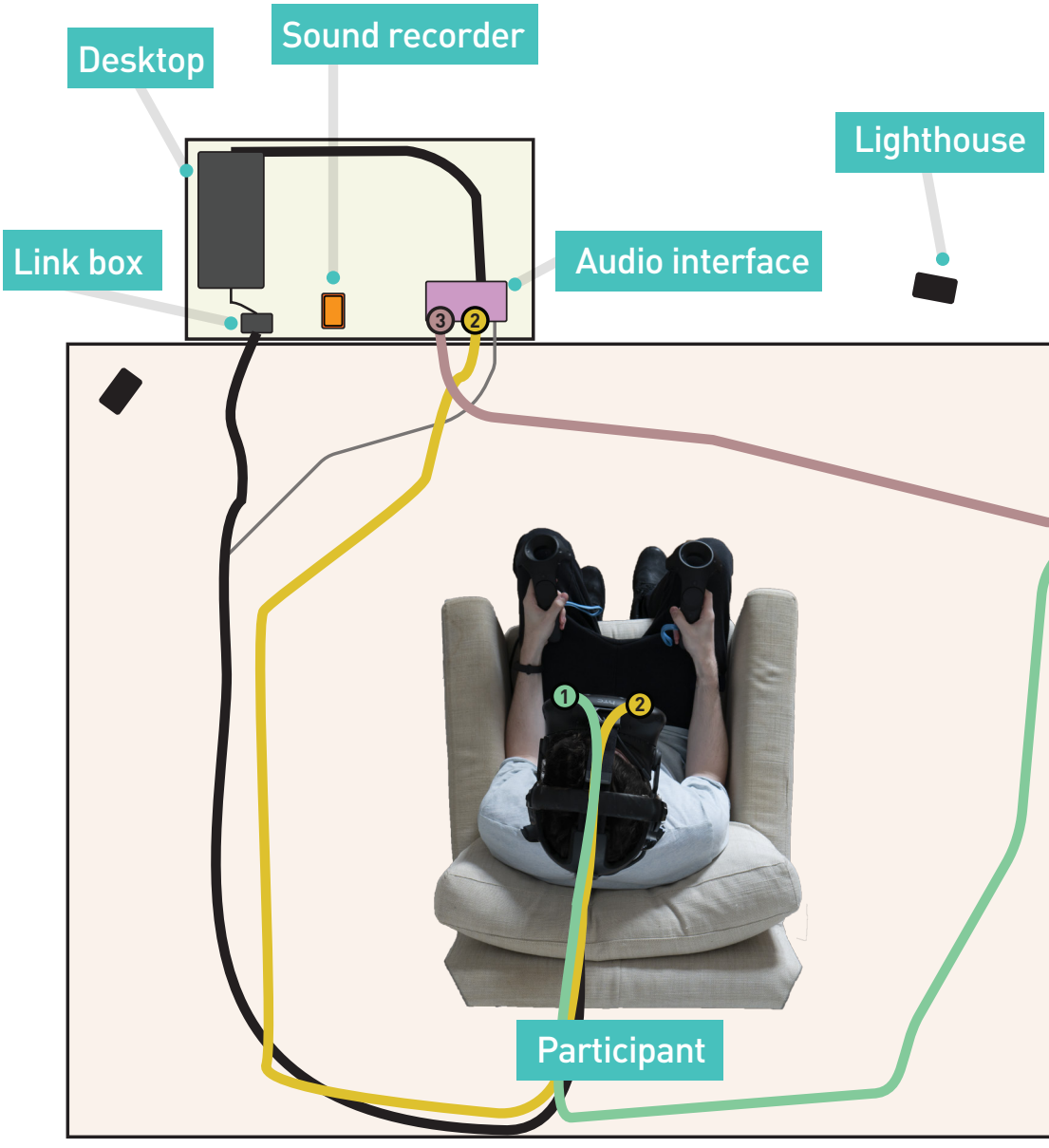
Finally, the optimisation. Since the prototype would run in Unity on a powerful desktop and a less powerful laptop, some optimisation was required. This was done by making two versions. One with all the settings to the max. The other by lowering some graphics settings, and disabling post-processing on the camera which was used to improve the visuals. Since the laptop was for the interviewer those settings were acceptable.

SETUP

The plan of the user test was to have two users join the same virtual room in VR, they could hear each other talk, and sidetone could be introduced. Considering the latency with wireless communication was too high, the setup for the user tests needed to be wired. This meant needing two physical rooms that were in proximity, but far enough that the users could not hear each other. For the user test, I needed to be one of the two people, to explain how it works, which tasks to follow, and ask questions during the test.

Figure 40 is an illustration to clarify what the setup looked like.

User 1 was the participant, and I was user 2, the interviewer. The connections worked as follows. The participant has two microphones, one goes to the interviewer's external sound interface and computer,



the other goes into their own interface for the sidetone. This way the interviewer can hear the participants talk, and the participants can hear themselves through the sidetone. The interviewer only needs one microphone going to the participant's interface which goes to Unity and the sound becomes 3D audio within the virtual environment.
The participant was using a high end desktop with the best visual quality possible, a HTC Vive HMD and controllers, a Steinberg UR242 audio interface, sound recorder, closed-back headphones, and two lavalier microphones. They also had a chair that was similar to the one from the virtual environment.
The interviewer was using a high end laptop, a HTC Vive HMD and controllers, a Focusrite 2i2 audio interface, sound recorder, closed-back headphones, and one lavalier microphone.

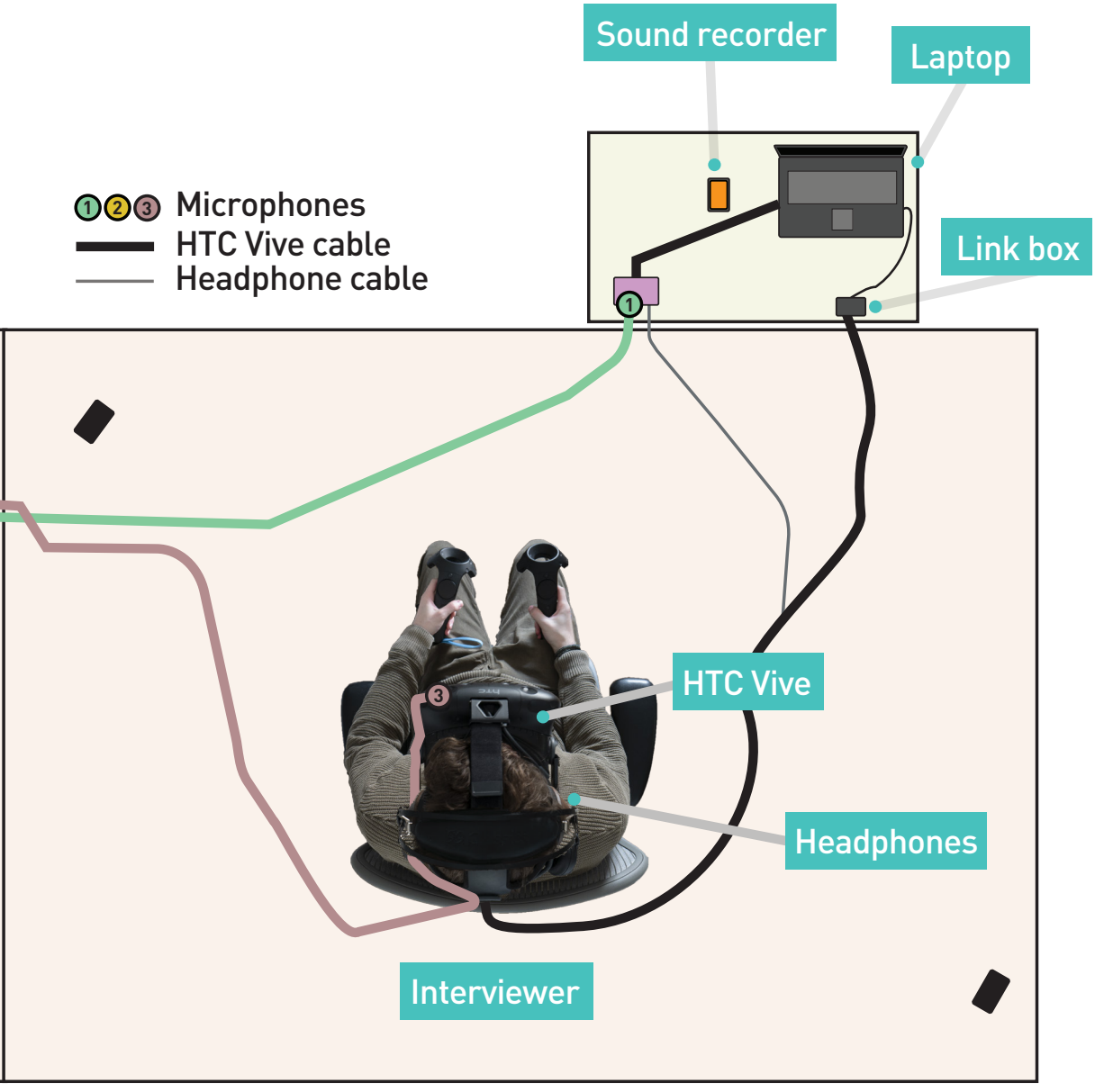


Figure 40. Illustration of the setup for the user test

USER TEST

ASSUMPTIONS

Sidetone sounds better than having none.
The other voice sounds better with reverb than without.
Both voices having reverb improves the sense of being together.
Directivity adds realism.

GOALS

For this user test, the main goal was to end up with recommendations for sound design in Social VR for the TV sound, the communication, and the environment.

To do so, the following sub-goals were formulated:
Evaluate how the prototype scores on relaxing, immersion and togetherness
Determine user preference on reverberation and sidetone
Evaluate environment choice
Evaluate whether the sound quality of the TV was cinematic or not

SAMPLE SIZE

9 people participated in this user test; 4 women and 5 men. Ages from 18-30. None had previous experience in VR, or they used it once or twice before.

STRUCTURE

The user test started by explaining what would happen, and how the controls worked. Then the participant was set up with the headset, headphone, microphones, and controllers. They were left alone for a minute in the virtual environment before the interviewer joined them in VR. They were asked what they thought of the environment, and then they were explained how the test would proceed.
First they lowered the TV, selected a documentary, played it, and watched it together. When it ended, they removed the TV, and they were asked questions about the experience. The participant had full control over whether or not they would pause or stop the video, but it would stop on its own after about 3 minutes.
Next, the reverb settings of the interviewer's voice were changed and compared so the participant could choose what they preferred. Then the sidetone was activated, and they chose whether to keep it or not. They watched another documentary with those preferred settings. The same questions were asked, and finally their HMD was removed. Some follow up questions were asked, and that concluded the user test.

METHODS

To keep the experience more in line with the real context, this time the participants were kept in VR throughout the whole user test. Their actions were observed and asked about when relevant. They were interviewed and asked to rate criteria; see appendix 3.A. for the questions and criteria. All sounds were recorded and later transcribed.

QUALITATIVE DATA ANALYSIS

First the qualitative data was analysed to be able to collect the scores participants gave to the different criteria. The recorded audio of the participants was transcribed, writing down both the given scores and the feedback, and reasoning behind it. Considering the low sample size, the qualitative data was leading once again, and the scores were counted more as indication. Both qualitative and quantitative research were then combined to formulate more specific recommendations for sound design in Social VR.

GOALS

The main goal was to gain feedback about the current prototype so they could be turned into final recommendations for the sound design in Social VR.

METHODS

The same methods were used as in previous user test: a similar approach as statement cards in context mapping. The recordings were transcribed, and the most relevant quotes were collected and clustered. Each participant had its own colour, the quotes were either improvement points or comments about their experience. These were then translated into recommendations for sound design in Social VR.

QUESTIONS & FULL QUOTES IN APPENDIX 3.A. & 3.B.

RESULTS

The interviews were transcribed as in the first user test series, and important quotes were selected and ordered based on their theme. This would organise the user feedback with the goal of finding the user requirements for sound design in social VR.

Sound mixing

"I thought the sound would be louder because at home we also have one of those windows and the rain was a lot louder."

4 participants remarked a sound was not as loud as it should be. Participants noted the voice sounded loud enough, but that the TV did not sound loud enough. One participant said the sidetone was a little bit louder than he would expect.

Interaction

"You helped me with the cubes, almost felt like you were with me in this room."

Two participants would have preferred more conversation during the watching. Two participants thought being helped during the experience made them feel like they were there with someone else.

Localization

"Clearly digital, but clearly from the right and not the same as the sound from the TV."

The sound of the interviewer's voice came from the head, especially with the low reverb setting. With the default reverb it surrounded the participant more.

Other voice

"It's clearly digital. It sounds like a real voice, but it's not as deep as a person sitting next to you."

3 participants noted it sounded digital. As if through a microphone with a filter on it that removed the lower frequency and depth of real voices. The reverberation also made a participant feel like that voice sounded as a godly presence from everywhere.

Reverberation

"Between the white and grey one, not completely gone, but a bit less than the grey one. The reverb is a bit too much, it's more than what I would expect."

2 chose the white one, 3 chose the grey one, 3 wanted something in between and 1 chose the black one. The white one felt more personal, like talking in the ear directly, felt more like a microphone. The grey one had too much reverberation for their expectation, but was nice to listen to. The black one was way too much, but it helped a participant with intelligibility.

Sidetone

"It feels like I'm in the room, more than before."

8 out of 9 preferred to have the sidetone turned on. Four participants felt more like they were in the room because of it. Five participants noted that it felt like a voice recording. When there was no sidetone, four participants noted it sounded muffled, like their own voice, but less loud. It sounded different.

Environment

"It's pretty, it's a nice house. Cool that you can look everywhere, to the kitchen and the stairways."

The environment was seen as comfortable, relaxing, open, realistic, and as a place one would sit at. One participant wanted more dynamism, with a plant or pet. One participant wanted it to be less gloomy.

TV

"It sounds like you're watching in a cinema. The sounds are around you."

All participants thought it sounded more like a cinema than headphones. One mentioned it sounded a little bit like using headphones, but still felt like a cinema. Two mentioned it sounded around them. One mentioned some depth was missing, as if the distance was not right. Another mentioned it would be a 5/5 if there was more bass.

Realism

"Real with a pinch of virtual reality"

The environment was mentioned as feeling real by all participants. One participant mentioned the familiar furniture contributed to that, another mentioned it felt clearly animated, three mentioned it felt real with a bit of virtual.

Presence

"The only way you know you're not in this environment is that when you look down, you only see the chair."

One participant noted they felt like they really sat there. Another that they were aware that they had gear on their head, so it felt like something else than watching TV normally. That could be improved with better comfort and field of view. One felt like they could move around the room to improve presence. One noted that they did not have virtual legs, but for the rest it felt like they were there and that it is more the visuals that gave the sense of being there. One participant noted the avatar and the teleporting were unnatural, but aside from that it felt like physically being there.

Avatar

"I would make the avatar a bit more personal"

One mentioned the avatar was a bit creepy, one noted it could be more personal and display body language better. One mentioned a physical model would probably work better.

CONCLUSION

Presence

All but one preferred having the sidetone on. When they could hear their own voice, it would improve the presence, but it still felt unnatural since it sounded like a recording. It also sounded from both ears instead of like in real life.

Interviewer's voice

Reverberation

The consensus around the reverberation seems to be between the white and grey one. The choice between one or the other was pushed by why they did not like the other one. The grey one had too much reverb for their expectation, while the white one was too dry. When it is too dry it sounds like a microphone or a whisper in their ear. When it is has too much reverb, it feels like another room and not realistic. Interestingly, most were able to estimate what would sound more realistic based on their expectation of the room's acoustics. They could sense if it felt off.

Localization

People were able to locate the interviewer, but using less reverb improved the localization of the voice. Reverberation had an influence on their ability to locate the voice, the more the reverberation, the more the voice surrounded the user.

Quality

The voice also still sounded like a digital voice, through a microphone and not as deep as it sounds in real life.

Latency

The latency was at least 320ms for a round trip. Sometimes it would build up a bit more, which would still take a while before the participants noticed there was any latency. At 320ms it was not noticed.

Co-presence

Hearing both these voices with reverberation made it feel more like they were in the same room as the voice. It improved co-presence. The level of interaction was satisfying, although it would sometimes differ based on what would happen. For example, when two participants were helped with the cubes, it improved their sense of being together. While with some participants there could have been more conversation through the TV watching. Improving the avatar would improve co-presence too by making it less abstract, and give the user more ability to express body language.

Environment

The environment was very realistic, more than expected. It could be improved by having a better quality HMD, less colour grading, and by being less aware of the HMD. It being open and detailed was appreciated, it felt comfortable and looked good. It could be improved with more dynamism or life, and maybe less gloomy.

TV

The cinematic experience of the experience was satisfying. It surrounded the participants as intended and it sounded as though the speakers were in front of them, it surrounded them. It still had room for improvement with more depth and bass.

Sound mixing

Considering how complicated the setup was for the user test, the sound mixing differed a little between the user tests. Sometimes the TV sounded too quiet, or the ambient sounds, or the sidetone was too loud. It had an impact on the participant's responses, and on their impression of how surrounding the sound of the TV was.

General conclusion

This experiment confirmed the assumptions. Adding reverberation, 3D audio and sidetone is the right choice for Social VR. It helped narrow down that people preferred the sounds and reverberation to be natural. And it painted a positive reception on the idea of using this type of setup in the future.

One assumption went unnoticed though: the directivity. Both the speakers and the interviewer's voice had directivity, but it was not experienced for two reasons. Firstly, the participant was sat and did not walk around, which made the speaker's directivity unnoticeable. Secondly, when talking to each other, both users looked at each other. It does not occur naturally in this context to speak without watching the person, unless for a short comment while watching th TV. But not long enough to notice the directivity.

QUANTITATIVE DATA ANALYSIS

Just like in previous series of user tests, the quantitative data was analysed with SPSS, and a qualitative research was done by context mapping important quotes and translating them into improvements and conclusions about how to design sound for Social VR.

GOAL

Rate the influence of the preferred settings on the immersion, relaxing, togetherness (and overall) (i.e. did it make a significant improvement)
Rate the influence of the preferred settings on the localization and naturalism of the sound
Evaluate whether the interaction levels, sound quality, cinema experience were satisfying

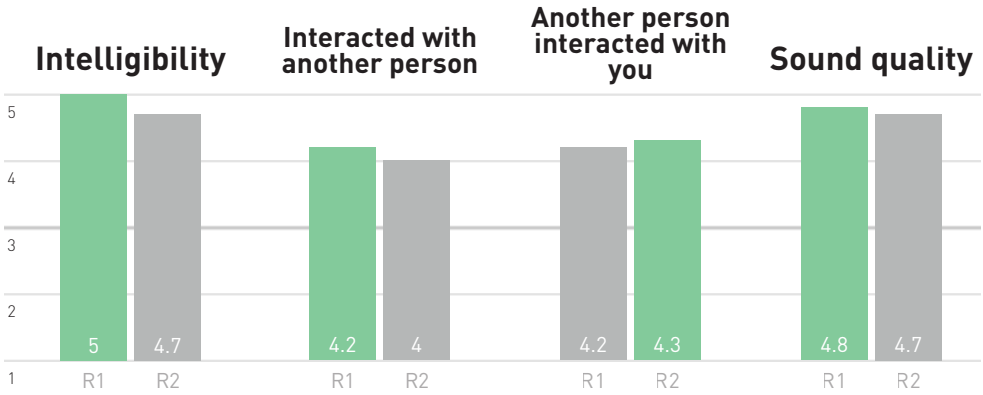
METHODS

T-Test (Immersion/Relaxing/ Togetherness vs. setting S1/2)
T-Test (Localization/Naturalism vs. setting S1/2)
Mean (Interaction, sound quality, cinema, realism)
T-Test (Localization based on reverb setting R1/2)

RAW DATA IN APPENDIX 3.C.

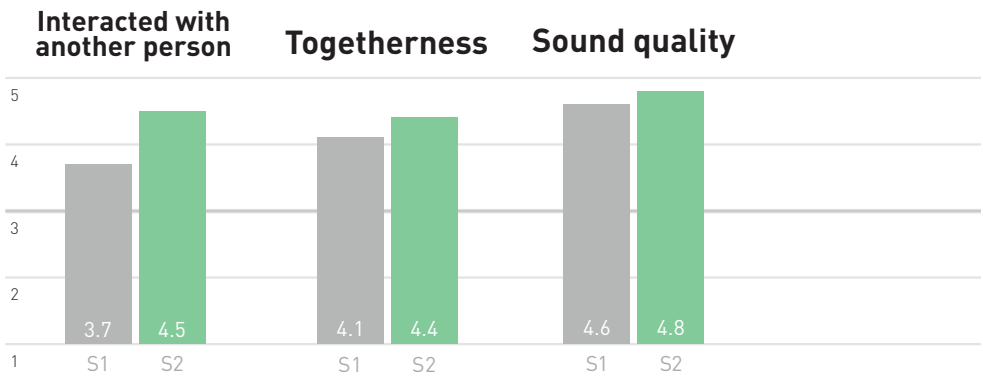
RESULTS

Qualities vs. reverb



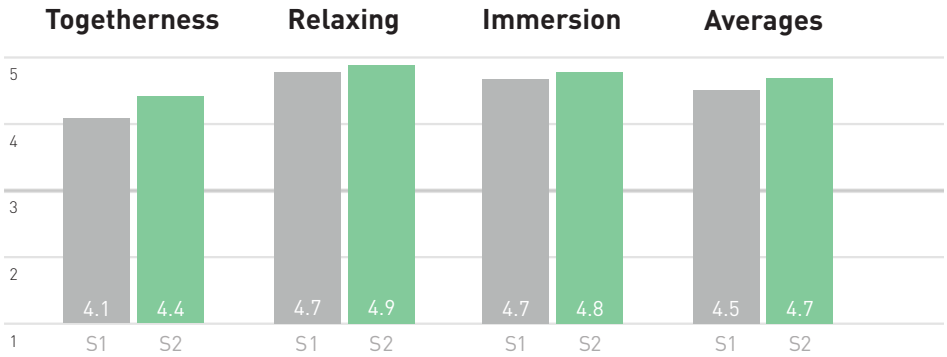
There were a few significant improvements between the low reverb setting (R1) and medium setting (R2). The intelligibility, sense that the participant interacted with another person, and the sound quality. Although the sense that another person interacted with the participant was higher with the medium setting.

Qualities vs. settings

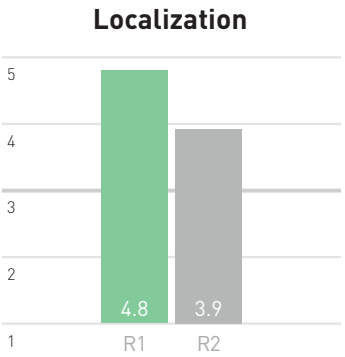


These three characteristics significantly improved by switching from the default setting (S1) to the preferred setting (S2). The default setting had no sidetone, and default reverb (medium). The preferred setting had the choice of using sidetone or not, and to choose which reverb participants preferred.

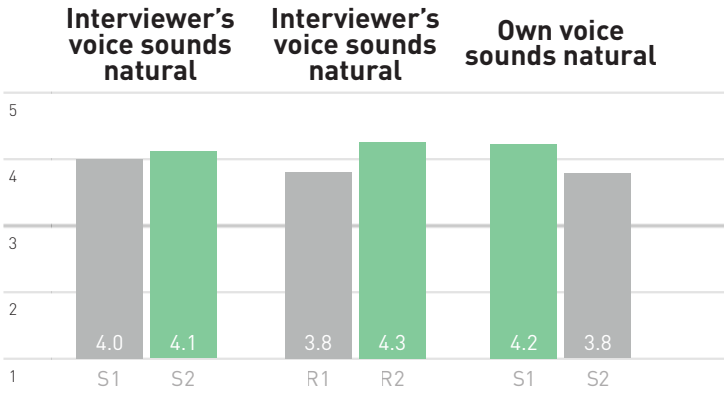
Trends



These averages only serve as a guideline and are meant to be combined with the qualitative data for definite answers.
All settings but one improved when the user chose their preferred reverberation and sidetone. The only one that decreased was the natural quality of the voices. In the preferred setting, togetherness scored well, relaxing scored exceptionally well, immersion too, localization scored well and natural scored the least, but still a 3,9 out of 5.
On average, the preferred setting scored more than the default setting.

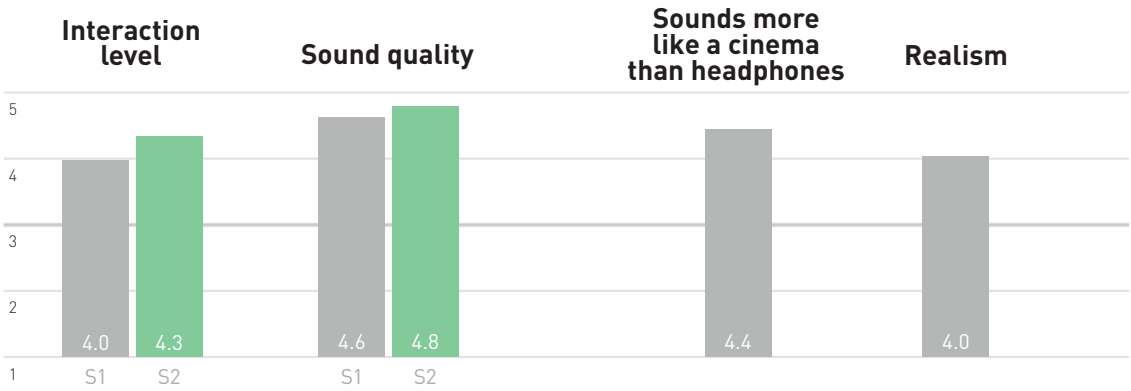


Based on the localization scores according to the reverb setting, it is clear that the lesser the reverberation, the better the localization is of the sound source.



The interviewer's voice sounded slightly more natural in the preferred settings, and more natural with the medium reverb setting.
Hearing their own voice with sidetone sounded less natural than without.

Interaction level, sound quality, cinematic experience and realism



The interaction level, which is the average of the scores given to questions about interaction, is satisfying. These scores indicate the participants were engaged during the experiment.
The sound quality scores a next to perfect score.

The cinematic experience was graded 4,44 out of 5, 1 being a headphone experience and 5 being a cinema. This can still be improved based on their feedback, but it is close to the best experience possible.
The realism hit 4 out of 5. This one had the most feedback for improvement.

CONCLUSION

As expected, having less reverberation improves the intelligibility of speech. The difference was significant, albeit rather small. The low reverb setting also significantly improved the interaction level with another person, and it significantly improved the perceived sound quality. However the medium reverb setting significantly improved the score of “another person interacted with you”.

From this data, it means the reverberation has enough influence to change the perceived quality and intelligibility of speech. It can also change the feeling of interacting with another person.

Next to that, the preferred setting significantly improved a lot of settings: interacted with another person, togetherness, and sound quality. Adding sidetone had influence on the interaction levels, and perceived sound quality.

The sound quality did not stand in the way of the grading of the other qualities, unlike previous test. The dedicated lavalier microphones and audio interfaces worked well.

The following results are not significant because of the low sample size or because of the wide spread of scores. They will be useful to match them with the qualitative data.

Choosing your own reverberation and use of sidetone seems to improve the experience on almost all fronts. Considering how some people chose the lowest reverb or kept the medium reverb, it seems again that providing a setting in between would improve the experience for either group. It also seems that the less reverb, the better the localization is of the sound source.

The only one that decreased was because the sidetone did not sound natural which dragged the natural average down.
Next to that, the medium reverb setting sounded more natural than the low setting.

DISCUSSION

Connections between the quantitative and qualitative data will be established to settle inconclusive or conflicting data.

Reverberation

Reverberation had a significant influence on the intelligibility of speech. Less reverberation being more intelligible. Yet the difference was small. Small enough to not gather any comments about it. Increasing reverberation could become a problem for most users in terms of intelligibility as only one person out of nine preferred the high reverb setting.
Aside from intelligibility it also had a significant influence on the interaction level with another person. Participants felt a dryer sound (less reverberation) was more personal, even too personal for some.
The qualitative data suggests a setting between the low and medium setting would be more realistic and preferable. The quantitative data supports this as the data did not show a significant preference for either low or medium setting. On average, either had their pros and cons.

Preference

Sidetone on was the most preferred option together with either low and medium reverb setting. Giving participants the choice improved the experience on all fronts aside from one. This confirms the settings they picked were indeed a preference.

Localization

There is no significant difference in the ability to locate a sound source between R1/R2 or S1/S2. However it strongly suggests the low reverb improved localization which the feedback confirms multiple times.

Environment

The environment itself was deemed realistic, although there was room for improvement.

Cinematic experience

Both data analysis confirm the sound of the TV was convincingly cinematic, and did not feel like listening to headphones. Although there was still a little room for improvement. They could not tell what would significantly improve the experience though.

Natural

The quantitative data suggests the interviewer's voice sounded more natural with the preferred setting, but that their own voice felt less natural. The feedback confirms this thoroughly. The sidetone should sound more natural.
The medium reverb setting seems to sound more natural than the low one in the quantitative data. If the ideal reverb setting seems to be between the low and medium ones, it should lean more towards the medium one as previous research showed that realism was preferred.

General conclusion

All in all, the results seem consistent with both analyses, or at the very least their differences can be explained. These findings will be regrouped, to clarify what information is concluded.

MAIN FINDINGS

Sidetone
Sidetone is preferred over hearing yourself through closed-back headphones.

Sidetone improves presence.

Sidetone needs to feel natural to further improve the experience.

Reverberation
Users sense when the reverberation does not match the virtual room.

If a sound is too dry for the room, it can influence the sense of distance between the users.

An accurate representation of the reverberation of the virtual room would be preferred.

A dryer sound means better localization.

A dryer sound feels more personal.

Intelligibility improves when there is less reverb.

Reverberation has influence on the perception that another person interacted with you.

Voice
Using a lavalier microphone with an audio interface provided a clear voice.

Using a lavalier microphone sounded like a digital voice.

A latency of around 320ms was acceptable and did not cause problems in communication.

Co-presence
Hearing both these voices with reverberation made it feel more like they were in the same room as the voice.

Environment
The environment was very realistic.

It being large, open and detailed was appreciated.

Using IKEA furniture felt familiar.

TV
A virtual 5.1 setup with 3D audio without reverberation combined with a large screen provides a convincingly cinema experience.

Preference
Choosing settings improved the experience.

FURTHER IMPROVEMENTS
Based on these results, some improvements were directly suggested while some others can be derived from them.

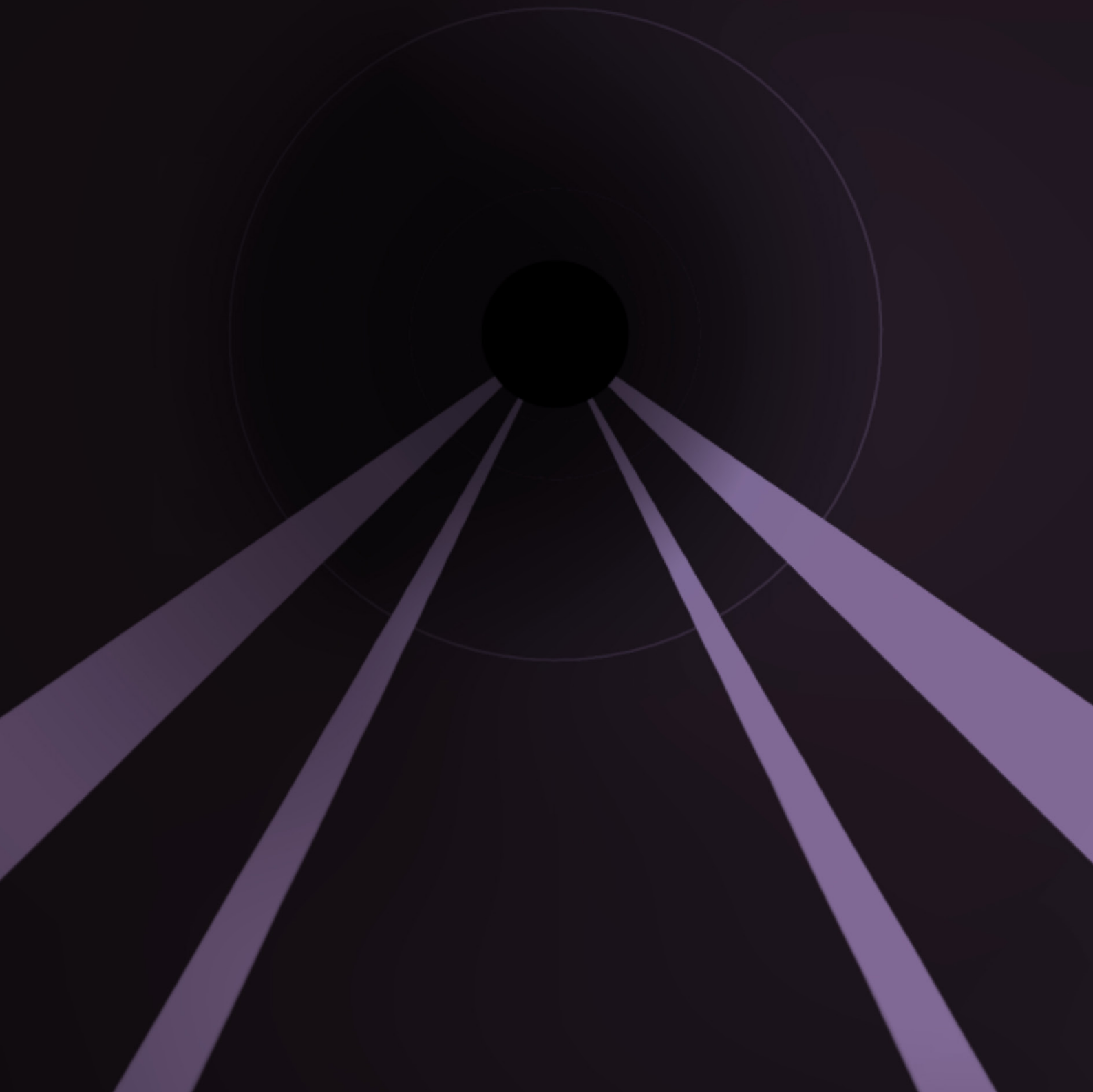
The sidetone still felt unnatural since it sounded like a recording. It also sounded from both ears instead of like in real life.
Improving the avatar could improve co-presence too by making it less abstract, and give the user more ability to express body language.
The environment could be improved by having a better quality HMD, less colour grading, and by being less aware of the HMD. It could be improved with more dynamism or life, and maybe less gloomy.
The TV sound still had room for improvement with more depth and bass.
The sound mixing needs to be consistent for testing.
The reverb setting could be set in between default and low to be more accurate, or another reverberation model could be used.

CONCLUSION

Following this series of user test, it was determined from the feedback that the prototype was successful, and that the goals of this project were reached. No unexpected new aspect of the sound design was found during the testing. There were only suggestions on how to improve the existing elements of the experience.

The findings of all the project's research can now be converted into recommendations.

RECOMMENDATIONS



INTRODUCTION

RECOMMENDATIONS

The findings of all the project's research will be presented in the form of recommendations for developers, researchers, sound designers, and anyone interested in the sound experience in Social VR.
Although the project focused on the sound experience, some feedback about the visuals was also included considering the visuals also had an influence on how the whole experience was perceived.

The recommendations are separated in two:
- Recommendations for Social VR which includes recommendations for the sound and visual experience.
- Recommendations for user testing in Social VR which includes insights from how to prototype for a user test and how to test Social VR with users.

RECOMMENDATIONS FOR SOCIAL VR

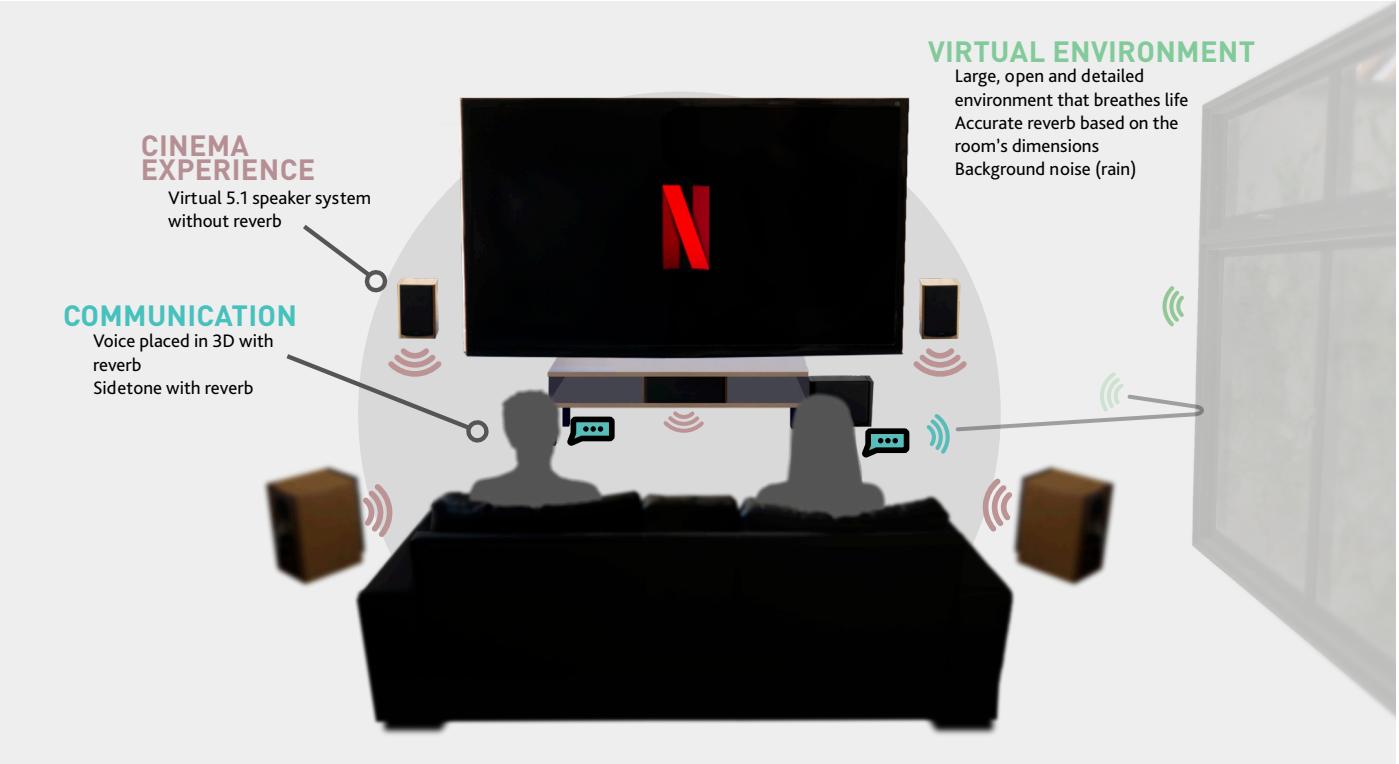


Figure 41. Illustration of the recommended sound experience in the virtual room

This part of the report is a collection of the recommendations for the sound design and visuals in Social VR based on the project's research. Although the focus was on the sound experience, during the testing it became clear the visual aspect influenced the immersion too. The findings for the visuals are highlighted by a purple box.

COMMUNICATION

Interlocutor

Current demos and applications of Social VR have the sound of the other user staying in the left or right ear. Users preferred for it to make use of 3D audio. This means with position, distance effects, and HRTF. The sound is virtually placed in 3D and moves from left to right according to the headset's rotation and position.

The interlocutor's voice should also have reverberation that matches the expected reverb of the environment.
Although introducing directivity to the voice seems appropriate if the users move around, in the context of watching TV together, it seemed to go unnoticed.

Sidetone

An important aspect of talking is being able to hear yourself talk, which is called sidetone in the telecommunication field. Although people who have not tried it think they do not want it, in practice users preferred having sidetone. Providing sidetone improves the experience, however it can sound too digital if no effects are applied to make it sound more natural. This can be improved in different ways:

- Digitally: using EQ filters on the voice to mimic the natural sound better or by placing it in 3D where the mouth is. This requires a plugin that reproduces the head effect well. To avoid the CPU load, it could also be an integrated sidetone in the HMD itself much like old telephones had. Sony's PS4 VR HMD has this type of sidetone although it does not have filters to make it sound more natural.

- Physically: open-back headphones would provide a similar effect to the sidetone since they allow external sound in. Valve, for example, took it a step further and introduced a headphone in their Index HMD that sits a couple of centimetres away from the ears. This is the most natural way to still hear your voice in VR.

Both ways can be interesting depending on the context. If used in a noisy environment, having a digital sidetone provides more control and can be made to only pick up on the voice and not the environment with a directional microphone. But for home use, Valve's implementation seems to be more appropriate as these environments tend to be quiet.

Co-presence

Co-presence is defined as the feeling of being in the same environment together.
Hearing all voices, including your own, with the same reverberation made it feel more like they were in the same room together.
Avatars also had an influence on co-presence. For the user tests, basic avatars were used with a spherical head and hands. Users would prefer a more physical avatar that can express more body language, and that feels more personal. Not having virtual legs while sitting takes away from the sense of being there.

Presence

Presence is the feeling of being there in the virtual environment. It is negatively affected by real life obstructions such as cables and the headset.
Even knowing that the real room has obstructions will be kept in mind when moving around in VR. If seated, those obstructions are ignored.
Providing a chair that had approximately the same dimensions as the virtual chair participants were sitting on, improved their presence.

Sound quality

The sound quality of the microphones matters. For example, the HTC Vive's microphone can result in a bad experience because of its quality. Although this is mostly bound to the user's choice of hardware, it is preferred to use a lavalier microphone connected to an audio interface to improve the recording quality, clarity, and intelligibility. A clean recording also removes unwanted background noise that sounds unnatural, and takes away from the immersion.
Hearing when the voice goes on and off because of the microphone's noise especially broke the immersion.

As a means to listen to the sounds, it is recommended to use closed-back headphones through an external audio interface, if the environment is noisy. Otherwise an open-back headphone which would then allow a natural sidetone. The reverb on the sidetone will probably not match the one of the virtual environment though, but this could be solved by still routing the microphone to the environment and playing back only the reverb of the virtual room.

Latency

In communication, the delay it takes for one sound to go to the other user and then going back is called a round trip latency. It is recommended to have a round trip value of 150ms and not to exceed 250ms (VOIP-INFO, 2018). Unity's Audio Engine provides around 320ms round trip latency for a wired solution. In practice, that latency was not noticed. It can however be improved by using middleware such as Wwise or FMOD and to transmit sounds over a network.

Localization

Several effects are needed for good spatialization: Interaural Time Difference (ITD), Interaural Intensity Difference (IID), and pinnae effects.
The use of HRTF is crucial for 3D spatialization. Custom HRTF would provide better accuracy, but generic are acceptable.
More reverberation means less accurate directional localization.

CINEMA EXPERIENCE

The TV experience users wanted was that of a home cinema. This effect was achieved by creating a virtual 5.1 system placed in 3D. This creates a satisfying cinematic experience. The only improvement would be to improve the amount of bass. For example by using a real subwoofer to provide the shaking feeling of theatres which cannot be emulated by headphones.
Although reverb is needed for the voices, it is best to leave room effects out for the TV speakers. Mono sources sound realistic enough with reverb, but applying reverb to a virtual 5.1 speaker system muddies the sound and is distracting.

ENVIRONMENTAL INFLUENCE

Environment

Users prefer a large, open and detailed environment. Some organic elements should be added such as moving trees, wind, plants, and candles. A 3D environment works better than a 360 image and avoids dizziness with head movements.

Using appropriate background noise helped to relax the users while covering up any unwanted noise from the microphones. For example, a large open living room with rain hitting the windows. If using rain, it should also visually look like it is raining.

Reverb

Users can sense the difference between open-air and closed environments based on reverb. They are even able to sense when the reverberation does not match the virtual room. They prefer it to be realistic.
The shoebox model is not accurate enough for the reverberation of a room if it is only based on its size and materials of the walls.
A room filled with furniture, decorations, and carpets will have less reverb than if it were empty, which is not taken into account with the shoebox model. The settings can be tweaked based on user feedback of what they expect the reverb to sound like. Or a more accurate reverb could be done with convolution reverb of a similar room, at the cost of more computational power.
Next to influencing directional localization, reverb also influences intelligibility negatively: the more reverb, the less intelligible the sound becomes.

MOVEMENT

In the context of watching TV together in VR, users found there was no need to be able to walk around.
Users expect certain sound cues to happen based on implied movement. Such as walking to a door.

SOUND MIXING

It is important to provide volume controls to users to make sure they are using the most natural settings possible for the experience.
Providing control to the users improves their experience.

THE REAL WORLD

The outside also had influence on the VR experience as we have seen with the cables and HMD weight decreasing presence.
Another factor is the visual fidelity of the hardware. The field of view available in most HMD's is currently too low to see the other person while watching TV in VR without a specially adapted virtual room.
The screen-door effect takes away from the immersion if it is noticeable.
The viewing of a movie in VR also requires a better resolution since people are used to seeing movies with a certain resolution in real life. They would rather have the same level of resolution in VR too.
All these aspects are bound to the hardware and is improving by the generation, however it is still important to keep in mind during the development phase for Social VR. The FOV for example is improving as shown by the Pimax 5K and Valve Index which have 170 and 135 horizontal FOV respectively compared to the HTC Vive's 110. Both also have better resolution and a less noticeable screen-door effect.

SIDE NOTE

I would suggest looking into the Valve Index and comparing it to closed-back or even open-back headphones to find out if it improves the experience as suggested by this report. In theory it would introduce better comfort, more immersion, and natural sidetone. The Index could be combined with an external subwoofer for a true cinematic experience too.

RECOMMENDATIONS FOR USER TESTING IN SOCIAL VR

During this project, testing Social VR was a learning experience on its own. Hereby recommendations for any developer wanting to delve into developing for Social VR and the challenges they can expect for user testing. One of the first lesson was that for user testing sound, it was necessary to provide satisfying visuals too, considering how much influence they had on sound and vice versa.

PLATFORM CHOICE

The most important choice that can be made is the platform. Each platform has its perks and downsides. Each platform has different amount of resources available, requires different levels of expertise, and provides a different set of add-ons, third party plugins, and assets. The choice also depends on the complexity of the project.

For example, using WebVR without prior programming knowledge for Social VR is not practical. For a beginner using available resources, it is only possible to create a basic audio-visual experience such as listening to music and displaying a 360 image. Anything more advanced than that will require much more experience.
Using an open source platform such as Unity and Unreal Engine for Social VR is more accessible for beginners. There are much more resources available online with tutorials, there are assets available that can be purchased, and there are many third party plugins available for advanced features. For Social VR, I would recommend the PlayoVR project on GitHub to understand how it works in Unity, and especially how to network objects and sounds.

LIMITATIONS WITH UNITY

Although there are many resources available for Unity, in this project some limitations became clear as soon as more advanced features were required. Using a sound interface through ASIO with Unity's Audio Engine is not possible. A middleware is necessary. The latency Unity provides for the audio is also barely acceptable for Social VR purposes, especially if it is recorded and transmitted through a network. Since the plugins available on the Unity Store use Unity's Audio Engine, expertise on using middleware is required to improve the latency. This may be important for user tests, as a round trip latency above 250ms could influence the results. Programming knowledge is required for any advanced features and for networking. Making a single player demo of Social VR can be done in minutes, but the hardest part is understanding how to network objects so that they move in multiplayer, and so both players can interact with it.

PRE-TEST RECOMMENDATIONS

The type of content that is shown during the test influences the experience. This is especially important when comparing environments between each other. Using a short fragment of an animal documentary without violence or chase was neutral enough.

Users will try actions they can do in real life such as grabbing objects within reach. In a pilot test, a user grabbed an important button and threw it away. It is suggested to make throwable items users can practice with.

It is essential to come up with a list of steps to follow to provide a consistent experience.
If the experience is seated, making sure the participant is facing the right way can be done by having an arrow on the virtual floor.

If the experience is timed to start at a certain time, adding a delay before the experiment starts prevented it from starting too early while the participant was still setting up their HMD.

Not providing a stable 90 frames per second experience can induce dizziness, and will hurt the grading and feedback.

Providing enough possibilities to interact is crucial. It should feel like an active experience, they can interact with the environment, with the content, with the other player. Otherwise it feels like they are being experimented on rather than they are experiencing it.

TESTING SOCIAL VR

It is recommended to use a fixed VR setup to keep the testing consistent and avoid wasting time on troubleshooting unexpected errors in the setup. In case of testing the sound experience, it is therefore also important to keep the sound mixing consistent.

Stepping in and out of VR to fill in a questionnaire takes time. It also requires the participant to readjust their HMD instead of having the same settings the whole time. This again introduces inconsistency as sometimes they position the headset differently than before.

Asking participants to think out loud while they listen to a sound experience does not work.

Being able to compare different settings directly in VR without taking the headset off helped to grade and compare those settings. Participants were able to keep the headset on for around 30 minutes without issues.

If the interviewer cannot see the participant in the virtual environment, it makes it more difficult to understand what the participant is trying to accomplish. Having the interviewer as one of the two people in the virtual room considerably helped. If the user did something unexpected, it could quickly be noted and asked about. If they needed an explanation, it was easy to provide without leaving VR or breaking the immersion.

By asking participants to grade settings verbally, the interviewer knows what they respond and can ask them to justify or develop their grade. This is helpful to understand what can be improved.

After the experience, asking them to describe how it felt gives participants an opportunity to reflect on it and provide more details.

Having a 5 minute experience was long enough for users to get used to an environment's soundscape and be able to grade it on different criteria.

RESOURCES

The following resources were valuable during this project. Other resources can be found in the list of references, but these resources helped to clarify information or ask for support.

Oculus Documentation which includes thorough and clear information about Virtual Reality Audio amongst others.

Facebook group Spatial Audio in VR/AR/MR which includes experts and researchers in this field.

Subreddit "LearnVRdev" which includes tutorials and support.

StackOverflow for programming questions.

PlayoVR for a sample of a SocialVR project in Unity to help understand how it works.

VRTK which is a VR toolkit that helps to quickly set up interactions in VR such as buttons.

CONCLUSION

The purpose of this project was to analyse and improve the sound experience in Social VR. To do so, the project was divided in three phases.
The first phase focused on understanding the current state of Social VR and the theory behind hearing sound in real life and in VR.
Based on user requirements, a basic Social VR prototype was developed to further understand what users expect and want.
With an iterative approach, a more advanced prototype was made which represented an ideal environment based on the accumulated feedback. The prototype was well received, with few minor suggestions on how to improve the experience even further.
Because of the efforts put into prototyping the sound experience in high fidelity, the conclusions and feedback could be converted into recommendations. Anyone interested in sound design for Social VR has now access to knowledge which they can explore further and apply for their context.
Since prototyping Social VR and user testing it is also a fairly new development, some guidelines were assembled for researchers. The purpose was to help them get started quicker, avoid pitfalls, and to understand how much efforts it requires to test in VR.

Hopefully this report will show developers the value of putting more effort into the sound experience, so that their users can not only be satisfied by the visuals, but also by the sound.

REFLECTION

This project was much more challenging than what I anticipated at the start of it. Going through the first phase resulted in interesting insights, but the value of the project essentially lied in testing the theory in practice.
On one hand, I had prior experience with VR in Unreal Engine, coding for Arduino, and experience in Sound design and music composing. On the other hand, the requirements of the prototypes were beyond my experience.
My personal goal was to either learn those skills while at TNO or to get support from them. But WebVR was too challenging for me to learn on the go. It was mostly coding only, barely any visual feedback.
Using TNO's implementation would have yielded a more complex "avatar" than the spheres I used, which is why I stuck with WebVR for months in the hopes of having high fidelity avatars.
But the knowledge gap was too great, the WebVR platform was not the right choice for me.
At that point I started looking at Unity, which felt more familiar and visual. It required many weeks of going through resources to understand how to make the prototype, but it felt possible. The biggest struggle was how to turn the experience from single player to multiplayer. This requires networking, and specific code to transfer actions from one user to the other. Actions such as body language or simply playing content on the TV at the same time. That part of the prototype was eventually figured out by going through the PlayoVR demo, analysing the script of each networked object, and figuring out what each line of the code did. Then changing those lines to fit my needs.
Other difficulties were with binding the directivity of the sound source to the headset. A custom script had to be written to make that possible.

Aside from prototyping in VR, testing in VR was also a new sight. There are so many aspects that influence the experience, so many elements that can go wrong. The sound mixing for example, even with the same mixing in Unity, in reality it depends on the volume set in Windows, the physical buttons on the audio interface, the settings in the software of the interface. It even depends on how well the headphone is put on.
Even with a list of steps to follow, there were inconsistencies. Most could be filtered out by asking the participants some questions, but some slipped through and found their way in the feedback.
The setup of the last prototype was difficult to figure out. Because of the latency, using wireless communication would introduce more inconsistencies such as sound quality issues.
It would also not provide the recommended maximum latency for communication. Setting it up with wires made more sense, latency wise, and it gave me more control over the experience.
However, even with that setup I could not test different latencies and verify what value seems to be the maximum acceptable latency in Social VR. From the feedback, I assume that latency can actually be higher than 250ms for VR.
Another aspect I wish I could have tested more is the directivity. In the context of watching TV with a friend, it might not be noticeable because you turn your head to talk to each other. In other contexts in Social VR where players are more active, it could make more difference, especially when there are more than two users.

CONCLUSION

Although transitioning from the company to working on my own with a new platform took time, I am satisfied with how the prototype paid off and resulted in conclusions I could not have reached without it. For example at the start of the project I found out about sidetone. Then from a questionnaire people said they would not expect to want sidetone in VR. While when they heard it in the prototype the large majority wanted to keep it.
The project was challenging in every step of the way, but taught me how to work with a company, how to approach unexpected situations, and how to adapt to them. I was able to apply the knowledge accumulated during my master to lead this project to a satisfying closure.

REFERENCES

Alary, B. (2017, 02 07). Reviewing the immersive potential of classic reverberation methods. Retrieved from Audiokinetic Blog: https://blog.audiokinetic.com/reviewing-the-immersive-potential-of-classic-reverberation-methods/

AltspaceVR. (2015, 10 22). Try spatialized audio in AltspaceVR. Retrieved from AltspaceVR Blog: https://altvr.com/try-spatialized-audio-in-altspacevr/

Altunian, G. (2017, 11 29). The Basics of Monaural, Stereo, Multichannel, and Surround Sound. Retrieved from Lifewire: https://www.lifewire.com/monaural-stereo-multichannel-surround-sound-3134860

Atwood, J. (2012, 06 27). Concluding the Great MP3 Bitrate Experiment. Retrieved from Coding Horror: https://blog.codinghorror.com/concluding-the-great-mp3-bitrate-experiment/

Audio Ease. (2018). Software. Retrieved from audioease: https://www.audioease.com/index.php#software

Bates, E. (2017, 06 19). Comparing Ambisonic Microphones. Retrieved from endabates: https://endabates.wordpress.com/2017/06/19/comparing-ambisonic-microphones/

Beerends, J. (2002, 04 15). On the quality of hearing one's own voice. Journal Of The AES.

Beerends, J. (2015). On the Assessment of High-Quality Voice Recordings including Voice Postprocessing. In Drossos, Evaluating the Impact of Sound Events' Rhythm Characteristics to Listener's Valence (pp. 174-183). Corfu: Audio Engineering Society.

Begault, D. R. (2000). 3-D Sound for Virtual Reality and Multimedia. Moffett Field: Ames Research Center.

Boyd, D. (2014, 06 04). Using a Shotgun Microphone for Video Interviewing. Retrieved from DigitalOmnium: : http://digitalomnium.com/using-a-shotgun-microphone-for-video-interviewing/

Cambridge. (2018). Relaxed. Retrieved from Cambridge: https://dictionary.cambridge.org/dictionary/english/relaxed

Cambridge. (2018). Immerse. Retrieved from Cambridge: https://dictionary.cambridge.org/dictionary/english/immerse

Casman, J. (2018, 01 25). Effectively Using Spatial Audio - Interview with Michael Wohl. Retrieved from theta360: http://lists.theta360.guide/t/effectively-using-spatial-audio-interview-with-michael-wohl/2265

CATT. (2018). Company Info. Retrieved from Catt: http://www.catt.se/

Chu, W.T. (2002, 01 12). Detailed Directivity of Sound Fields Around Human Talkers. Retrieved from NRC Publications Archive: https://nrc-publications.canada.ca/eng/view/fulltext/?id=c1449aba-ee5b-48de-8312-ec325d31ef37

Dolby. (2018). 5.1 Speaker Placement. Retrieved from Dolby: https://www.dolby.com/us/en/guide/surround-sound-speaker-setup/5-1-setup.html)

Earick, B. (2017, 07 24). Listening to interactive, three dimensional audio in VR. Retrieved from uploadvr: https://uploadvr.com/world-vr-audio-perspective/

Ferron, E. (2017, 01 10). Hands on: 3D audio with Ossic X headphones. Retrieved from Newatlas: https://newatlas.com/ossic-x-3d-headphones-hands-on/47290/

Franco, M. G. (2018). Generic HRTFs may be good enough in Virtual Reality. Improving source localization through cross-modal plasticity. Frontiers in Neuroscience, 21.

Gamespot. (2014). True 5.1/7.1 headset vs virtual surround sound. Retrieved from Gamespot: https://www.gamespot.com/forums/pc-mac-linux-society-1000004/true-5-1-7-1-headset-vs-virtual-surround-sound-31030702/

Glynn, L. (2016, 04 16). The Different Types Of Microphones Explained. Retrieved from Dawsons: https://www.dawsons.co.uk/blog/the-different-types-of-microphones-explained

Hollerweger, F. (2008, 10). An Introductin to Higher Order Ambisonic. Retrieved from Mur: http://flo.mur.at/writings/HOA-intro.pdf

Howard, D. M. (2017). Acoustics and Psychoacoustics. Taylor & Francis.

Kawamura, S. (2016). Virtual Reality (VR). Greenville: IEEE.

Lau, P. (2008). The Lombard Effect as a Communicative Phenomenon. Retrieved from Berkeley: http://linguistics.berkeley.edu/phonlab/documents/2008/lau.pdf

Lokki, T. (2011). Lateral reflections are favorable in concert halls due to binaural loudness. In A. S. America, The Journal of the Acoustical Society of America Volume 130, Issue 5. Acoustical Society of America.

Mallery, S. (2013). Using Lavalier Microphones with Recorders. Retrieved from Bhphotovideo: (https://www.bhphotovideo.com/c/find/newsLetter/Lavalier-Microphones.jsp

Mantione, P. (2017, 07 16). The Basics of Synth Envelope Parameters, Functions and Uses. Retrieved from Pro Audio Files: https://theproaudiofiles.com/synthesis-101-envelope-parameters-uses

McDermott, R. J. (2007). Journal of School Health. American School Health Association.

Mitsuownes. (2017, 01 12). PSVR - How to Access The SECRET MENU! (Sidetone). Retrieved from YouTube: https://www.youtube.com/watch?v=oRw562EcRf8

Morrison, G. (2013, 11 2). Open-Back vs. Closed-Back Headphones. Retrieved from Forbes: https://www.forbes.com/sites/geoffreymorrison/2013/11/02/open-back-vs-closed-back-headphones/#3c1c26d316a5

Nufire, T. (2018). Crosstalk Cancellation. Retrieved from ibink: www.ibink.com/tnufire/docs/XTalkCancelation.pdf

Oculus. (2015). Ambisonic Microphones. Retrieved from OculusVR: https://forums.oculusvr.com/developer/discussion/20409/ambisonic-microphones

Oculus. (2017, 10 26). Oculus Connect 4 | Breakthroughs in Spatial Audio Technologies. Retrieved from YouTube: https://www.youtube.com/watch?v=l7mhXRB9PA4

Oculus. (2018). Audio SDK. Retrieved from Oculus developer: http://developer.oculus.com/documentation/audiosdk/latest

Ossic. (2017, 08 18). The differences between stereo, virtual surround, and 3D-audio headphones. Retrieved from Ossic: https://www.ossic.com/blog/2017/8/18/the-differences-between-stereo-virtual-surround-and-3d-audio-headphones

Picinali, L. (2014). Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. In E. Motta, Human-Computer Studies (pp. 393-407). Elsevier.

Poirier-Quinot, D. (2016, 09 11). Augmented auralization: Complementing auralizations with immersive virtual reality technologies. Retrieved from ica2016: http://ica2016.org.ar/ismra2016proceedings/ismra2016/ISMRA2016-14.pdf

Prasad. (2016, 06 27). Editorial: Headphone jack vs. USB audio. Which is better? Retrieved from GSMArena: https://www.gsmarena.com/headphone_jack_vs_usb_audio_which_is_better_-blog-19060.php

Pyzdek, A. (2018). The World Through Sound: Refraction. Retrieved from Acoustics Today: http://acousticstoday.org/10-world-sound-reflection-refraction-principle-least-time/

Reality Check R. (2017, 12 28). BIGSCREEN VR MOVIE NIGHTS! TOP GUN 3D! | Round Table Chat w/Founder Darshan Shankar. Retrieved from YouTube: https://www.youtube.com/watch?v=20Jy_z0ghmQ)

Reddit. (2017). Side effect of using the Sidetone option. Retrieved from Reddit: https://www.reddit.com/r/PSVR/comments/5ul117/side_effect_of_using_the_sidetone_option/

Robjohns, H. (2004, 09). Can I remove the vocals from a track using phase? Retrieved from Sound on Sound: https://www.soundonsound.com/sound-advice/q-can-remove-vocals-track-using-phase

Rozario, H. (2017, 06 3). 10 Best 7.1 Surround Sound Headsets You Can Buy. Retrieved from Beebom: https://beebom.com/best-7-1-surround-sound-headsets/

SantoPietro, D. (2013, 12 28). Production of sound. Retrieved from Khanacademy: https://www.khanacademy.org/science/physics/mechanical-waves-and-sound/sound-topic/v/production-of-sound

Shinn-Cunningham, B. (2001). Localizing sound in rooms. Beacon St.: Boston University.

Silva, R. (2018, 03 05). Headphone Surround Sound - The Basics. Retrieved from Lifewire: https://www.lifewire.com/headphone-surround-sound-1846756

Sra, M., Mottelson, A., & Maes, P. (2018). Your Place and Mine: Designing a Shared VR Experience for Remotely Located Users. Conference on Designing Interactive Systems.

Telme2. (2017). What is APT X and latency. Retrieved from Telme2: https://www.telme2.com/blog/what-is-aptx-and-latency

VOIP-INFO. (2018). VOIP QoS Requirements. Retrieved from VOIP-INFO: https://www.voip-info.org/qos/


A

APPENDIX 1.A.

SOFTWARE LIST THAT COULD BE USED FOR PROTOTYPING

Oculus Audio SDK

Oculus has gone through many breakthroughs in Virtual Reality Audio. Its features include HRTFs, volumetric sources, point sources, near-field and far-field rendering, ambisonics and environmental modelling. It however, does not represent occlusion. Their SDK is available for Unity, Audiokinetic Wwise and DAWs (as a VST).

Resonance Audio (Google)

Resonance Audio includes support for Ambisonic files, room ambiance (early reflections and reverb), occlusion, ITD, IID, HRTF and even sound source directivity. This SDK is available for Unity, Unreal, Audiokinetic Wwise, Web and DAW.
https://www.youtube.com/watch?v=IYdx9cnHN8I

Steam Audio

Steam Audio has realtime sound propagation, occlusion, reflection, HTRF-based binaural rendering and ambisonics support. Available for Unity and Unreal Engine 4.
https://www.youtube.com/watch?v=e4L8Kc1oma8

DearVR

DearVR is a 3D audio reality engine plugin. It provides HRTF, ambisonic, elevation, azimuth and distance, reverb, auralization with early reflections and occlusion. It can export the sound as binaural or ambisonic.
DEMO: https://www.youtube.com/watch?v=ln61FNG3bB4)

AmbiX

AmbiX decoder plugin is a program that can work with ambisonic files and do post production on them. It is a VST and therefore compatible with DAWs or MaxMSP.
Demo: https://vimeo.com/81073470
BlenderVR, MAX/MSP (Poirier-Quinot, 2016)

Facebook 360: Spatial workstation

This free software suite is available for designing spatial audio in 360 videos and VR. It is available as a VST and SDK. It can also track and automate the position of a source, manually though.

Reverberation and reflections

To give a good idea of the room ambiance, you need reverb and reflections. This can either be simulated by Oculus Audio SDK or the Google Resonance API however, these use a simplistic “shoebox model” which does not represent the true geometry of the environment. This room ambiance is also called auralization, which 3rd party plugins can achieve in much more details than previously named SDK’s. For more specific purposes such as room acoustics predictions and auralization with geometrical acoustics, Catt (CATT, 2018) and ODEON are available. However, next to simulating the reverb, it is more realistic to make use of convolution reverb which is a way to capture and apply the real room ambiance of a real-world location. For 360 photos and videos this means that you could record the real room ambiance and apply that to the post processing on the voice of the people interacting with each other.
Audiokinetic has been keeping track on different technologies available for VR and the immersive potential of reverberation methods. (Alary, 2017)

1R1 Convolution reverb (Waves)

This plugin captured the sonic characteristics of over 100 places and the settings of the impulse responses of spaces and devices and be tweaked. It also recreates the early reflection buildup.
Demo: (Waves, 2018)

360pan suite (Audioease)

This plugin is more advanced as it can use the ambisonics recording and apply the reverb and room ambiance on a moving sound source in 360. However, it does not seem to support live head tracking and real-time application of the room ambiance.

https://youtu.be/WNOC3K9Civk

APPENDIX 1.B.

COMPARISON BETWEEN DIFFERENT SOCIAL VR APLICATIONS

Typical applications of virtual reality do not require to record the users unless the application is Social VR.

A good example is VRChat, which is a free program that allows people with or without a HMD to join a chat room in which they are represented by an avatar. The users can interact by talking to each other, move around and participate in activities such as watching a movie on the moon. VRChat uses 3D audio for the sound and adds distance effects on the sounds made by users to fade the sound out if they walk away. The sound source also follows the avatar's position. The players typically use (gaming) headset microphones or small dedicated microphones. (Steam, 2018)

AltspaceVR is a similar Social VR app which puts users in virtual spacesin which they can walk around, they can attend live events or play interactive games. AltspaceVR uses the Oculus Audio SDK which is more spatialized audio than what Unity provides on its own as it provides better reverb and room echo on the voices. (AltspaceVR, 2015)

vTime is a social VR experience too which is more similar to what TNO is working on as the user sits into a seat in an environment and they can invite friends or family and discuss together. The users cannot move in the environment, but they can still move around with their avatar that also synchronizes the mouth movements with the sound the user makes. Here too, the sounds are represented in 3D audio by panning left to right and do not seem to be influenced by the environment at all. (vTime, 2018)
There are also multiplayer games in which users can play and communicate together such as Star Trek: Bridge Crew (Ubisoft North America, 2016). Different users are sat in a space ship and have different roles assigned. They can communicate together to give orders and collaborate on solutions. In this game too, the voices are unaltered and only panned to the users location.

These recent examples indicate that Social VR exists already in different shapes, but that sound is rather primitive with the exception of AltspaceVR which at least used a different SDK to develop better audio for the environment, while still leaving the live voices raw. Another major shortcoming is that the users have cheap microphones which are often badly setup and saturate at times. Since the sound quality provided by it is poor, adding effects to it would render it incomprehensible and thus defeat the purpose of social VR.

APPENDIX 1.C.

TESTING THE TNO DEMO

The experience is that we are watching a presentation that slides through on its own while in a 360 picture of an office. If you move around, the scaling is wrong in the virtual environment. The isolation does not work well on black, which is why there are some patches disappearing on the headsets since they are black. There were a lot of sound issues though, one microphone would saturate, two microphones were only played on the left ear and none of it was dependent on the head orientation. During the experience, I had no idea how loud We are all watching a presentation, in a 360 photo, the scaling is a bit weird, the isolation of the person doesn't completely work properly, sound is coming from one side for two users (to the right), no idea how loud I am talking, the noise-cancelling headphone is working amazingly except that you can hear other people talking too, background noise is gone though. The talking seems pretty slow, not sure if people hear you, no effects from the environment, scale seems wrong with the table though, do you want controllers in your hands the whole time? Sometimes people freeze. The placement of the table is different per person, also the placing of the users.

- What kind of interactions are possible? Can you pause the video? No
- Is there a noticeable delay (visually, audio)? (Audio… I'm not sure, it felt like it, like a skype meeting)
- Does talking to the other person feel natural? (not really, can't hear myself, don't know if they hear me, couldn't really tell where the sound was coming from?)
- Is the sound matching when the person talks? (yes)
- How does the talking/own voice work? Is that distracting? Can you hear yourself over the headphones? (No own voice which was confusing and made people talk loudly, you can hear yourself a little bit, pretty well filtered)
- What's the goal of the demonstration? (watch TV/video content together in the same environment)
- What can you do? (watch Tv/ video content, talk to each other)
- What type of interactions are possible? (pause people apparently/mute. Not really interact with the content. Move forward? No. Change content on the fly? No. You can also not really move around, you have to stay in one spot otherwise the scaling is wrong.)
- Does the environment interact with you too or no? Is the sound stereo? Are the people placed in the environment tracked in position or not? (no, sound is barely even stereo for some people)

(when not me testing, some freezes, there is an echo? Maybe too close to the other person next to them? Not supposed to be a problem though for my project as you're not supposed to be in the same room. Talking pretty loudly though, almost shouting, because they don't hear themselves? Because of the noise-cancelling?)

There was another demo too, which used Google Resonance and it was already much better. Spatial sound worked well, the sound quality was also much better, but it does not include people in the environment.

APPENDIX 1.D.

QUESTIONS QUESTIONNAIRE 1

1. Why do you want to watch a show/movie with friends?
2. Do you:
    - choose in advance what you want to watch together?
    - browse through channels?
    - pick something together on the spot?

3. What is more important?
Choose between 1, 2, 3, 4, 5, 6, 7.
The social aspect (1) ----- The content (7)
4. In what specific words would you describe the experience?
5. What is the tone of the situation?
6. What interactions do you expect?
7. What kind of environment is it?
    - Loud
    - Quiet
    - Busy
    - Outside
    - Inside
8. Do you experience any problems related to sound in that setting? If so, which ones?
9. How often do you ask friends to repeat themselves after not hearing them properly?
    - Never
    - Sometimes
    - Often
    - Always
10. In the context of watching TV with friends, what sound quality would you prefer coming from the TV?
    - Mono
    - Stereo
    - Surround/Home cinema
    - Cinema
11. If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?
12. More specifically, what would you want the sound experience to feel like?
13. What is your age group?
    - 15-24
    - 25-44
    - 45-64
    - 65+

APPENDIX 1.E.

RAW DATA QUESTIONNAIRE 1

There are 38 responses on 13 questions.

14 apr. 2018 10:10

1 Why do you want to watch a show/movie with friends?

It' more fun than on your own.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

fun and cozy

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

sing-along, talking, eating, drinking, laughing, being scared and pulling the blanket over the head.

7 What kind of environment is it?

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

Friends might comment on a scene. Neighbours might be noisy

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

It should feel real, I want to feel their presence. I want to look to the side and see them.

12 More specifically, what would you want the sound experience to feel like?

the movie should be louder than my friends' voices.

13 What is your age group?

25-44

13 apr. 2018 04:33

1 Why do you want to watch a show/movie with friends?

Because I like spending time with friends

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Watching a movie with friends is relaxing and pleasant

5 What is the tone of the situation?

Fun, nostalgic, exciting

6 What interactions do you expect?

Talking, discussing, laughing, eating

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Sometimes can't hear content

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Friends avatars present in VR. Body language is important. Feel like sitting on couch watching content, not theater

12 More specifically, what would you want the sound experience to feel like?

Cozy but with a focus on the content

13 What is your age group?

15-24

12 apr. 2018 15:47

1 Why do you want to watch a show/movie with friends?

because i just like movies and sharing experience of it

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

it can be brainless fun with friends or an extatic trusted debate with friends

5 What is the tone of the situation?

relaxed and open and either serious or joyfull

6 What interactions do you expect?

all of them, there nothing to expect of it, the situation and the people involved

7 What kind of environment is it?

Loud

Outside

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

i felt like the quality isn't always there, but i'm not sure if it's hardware or software

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

full virtuality thing

12 More specifically, what would you want the sound experience to feel like?

natural and smooth, something discrete but latent

13 What is your age group?

25-44

11 apr. 2018 14:34

1 Why do you want to watch a show/movie with friends?

We are lazy to go out

2 Do you:

pick something together on the spot?

3 What is more important?

5 out of 7

4 In what specific words would you describe the experience?

Cozy and fun

5 What is the tone of the situation?

Relaxed and joyful

6 What interactions do you expect?

All

7 What kind of environment is it?

Inside

Quiet

Loud

8 Do you experience any problems related to sound in that setting? If so, which ones?

Maybe we bother my mom. Maybe we cannot understand the movie well without subtitles if someone talks

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Super cool out of this world but we are all sharing connected to each other. Its not that each of us is immersed in their own experience.

12 More specifically, what would you want the sound experience to feel like?

Like super clean

13 What is your age group?

15-24

11 apr. 2018 09:03

1 Why do you want to watch a show/movie with friends?

Share experience, comment on it, spend time

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Share, enjoy, fun

5 What is the tone of the situation?

Depending on the movie. It might be tense if it is scary. Usually joyful in either case. I think the movie sets the tone, though

6 What interactions do you expect?

All that you mention. Probably less singing than the rest

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

When someone speaks and you don't hear what they said. When you don't understand the dialogue, because your English is not up to that level.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Never

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

I think the nice part of watching a movie with friends is that you are actually together. Also before and after the movie

12 More specifically, what would you want the sound experience to feel like?

If this is more like a vr game. I expect it sounds like we are also same location. Maybe that I can hear direction of my friend or if they are walking closer or further

13 What is your age group?

25-44

11 apr. 2018 07:03

1 Why do you want to watch a show/movie with friends?

Doing something together, with another human, is usually just more enjoyable - no matter what it is (so long as you are in good terms with the other person). Watching a show/movie together fulfills this and gives you something to potentially talk about afterwards.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

Relaxing, entertaining, stimulating, companionship

5 What is the tone of the situation?

Relaxed, stress-free

6 What interactions do you expect?

A bit of talking (about what's going on in show/movie), a bit of general conversation if it is a casual show but not too much to distract from content, some laughing (if comedic content), light eating of snacks or drinks, silence/focus if it is more serious movie or show

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Surround sound not being set up properly or utilized correctly, loud A.C., people doing other things in the background, restless animals (if you have any; e.g. squawking bird, barking dog, etc.)

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Baseline hopes are that the image quality is good, screen size sufficient for the type of content being watched (and could be changed), warm/organic/homey environment, properly emulated surround sound that comes naturally from sound source(s) - doesn't simply sound like you're wearing headphones. Beyond that: ability to chose from a variety of environments (realistic, fantasy, and beyond) that are still catered to such a movie/show watching environment. For example, I don't care to be in the middle of the Arizona desert with my friends, simply with a floating screen in front of us, but a cool environment with the same setting (in some little hut, or outside but with comfy couches/setting - environments, no matter how crazy, that are still catered to watching something with friends or alone) is more appealing/inviting. The ability to customize/tweak sound (spatial/sonic tweaks) and picture settings.

Overall, no different from watching in real life, but you get to do it in cool, exciting, and unique environments.

12 More specifically, what would you want the sound experience to feel like?

As mentioned, natural/organic, like it comes from a real audio source, and not just sound directed to my headphones. For sound to sound stereo or mono if it is set as such, and surround as well, coming from emulated speakers.

13 What is your age group?

15-24

10 apr. 2018 15:49

1 Why do you want to watch a show/movie with friends?

tall about the content. Feel together

2 Do you:

browse through channels?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

comfortable, interesting, fun

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

laughing, talking, eating

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

A bit not clear

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

immersive content

12 More specifically, what would you want the sound experience to feel like?

real, surrounding

13 What is your age group?

25-44

10 apr. 2018 14:12

1 Why do you want to watch a show/movie with friends?

Learn each other about good movies, having a nice chat, drink some beers, prepare for an evening out

2 Do you:

pick something together on the spot?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

Fun but also a struggle for I have a strong opinion on movie watching

5 What is the tone of the situation?

fun, serious, slow

6 What interactions do you expect?

discussion, opinions, learning about eachothers likes and dislikes, laughing, storytelling (about seen movies)

7 What kind of environment is it?

Loud

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

No

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Like I could talk to them/ see whether they are actively watching the movie or not

12 More specifically, what would you want the sound experience to feel like?

Surround with the focus on the movie but also some capabilities of conversation

13 What is your age group?

15-24

10 apr. 2018 08:23

1 Why do you want to watch a show/movie with friends?

Memories

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

as stated before , I think memories are very important when hanging out with friends. When you watch a movie together you can later make jokes using some of the movie lines

5 What is the tone of the situation?

Chillll

6 What interactions do you expect?

Eating , laughing , crying from laughter

7 What kind of environment is it?

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

None

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Like we are all I'm the same
Room watching the movie .

12 More specifically, what would you want the sound experience to feel like?

13 What is your age group?

15-24

10 apr. 2018 01:48

1 Why do you want to watch a show/movie with friends?

Companionship

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Movies with commentary

5 What is the tone of the situation?

joyful, relaxed

6 What interactions do you expect?

talking, laughing, eating, drinking

7 What kind of environment is it?

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Not really

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

At least in sync

12 More specifically, what would you want the sound experience to feel like?

realistic

13 What is your age group?

15-24

9 apr. 2018 23:39

1 Why do you want to watch a show/movie with friends?

To share opinion on a movie, comment on it, laugh or discuss it

2 Do you:

A & C

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

It really depends on the genre of the movie

5 What is the tone of the situation?

Open, joyful

6 What interactions do you expect?

Chilling on the sofa, discussing

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

No

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

I would want it to look as much as possible like reality, because that's what I like about watching movies with friends - the real aspect of physical proximity and hanging together.

12 More specifically, what would you want the sound experience to feel like?

Sound from the friends avatars to come from where they sit. Sound from the movie to come from virtual sprakers or from the virtual screen.

13 What is your age group?

25-44

9 apr. 2018 21:57

1 Why do you want to watch a show/movie with friends?

To have something to do, that we share a common interest

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Secluded, intimate, enjoyable

5 What is the tone of the situation?

Same as above I suppose

6 What interactions do you expect?

Talking, laughing, eating, pausing, talking

7 What kind of environment is it?

Inside

Quiet

Spacious

8 Do you experience any problems related to sound in that setting? If so, which ones?

Yes, my computer makes a ticking noise sometimes that annoys when there is a silent scene.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Very similar to the real experience, as if he/she is right next to me, eating the same snacks

12 More specifically, what would you want the sound experience to feel like?

As real as possible, so that the balance between me and my friend is similar. But when I skype and play games at the same time for example, I often prioritize one over the other, since we are doing different contents. This should not be like this when we watch a movie/serie.

13 What is your age group?

15-24

9 apr. 2018 21:11

1 Why do you want to watch a show/movie with friends?

For fun/quality time

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Fun, bonding and shared interests

5 What is the tone of the situation?

relaxed, chill, cosy

6 What interactions do you expect?

eating and drinking, commenting on the show/movie, laughing

7 What kind of environment is it?

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Sometimes, especially when no subtitles are available, it can be hard to understand what they are saying in the movie/show, due to the talking of friends

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Like you are part of the show, it would also be nice to be able to see your friend maybe, so you have the extra feeling of them being their besides their voice

12 More specifically, what would you want the sound experience to feel like?

All round sound, it would be a pity if you have the experience to be in the movie/show yourself, but only visually and not through sound. On the other hand, I would still like to be able to hear my friend since I enjoy being able to comment on what I watch and talk about it when watching

13 What is your age group?

15-24

9 apr. 2018 20:39

1 Why do you want to watch a show/movie with friends?

share experience

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

fun

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

talking
drinking
Screaming
Laughing

7 What kind of environment is it?

Loud

8 Do you experience any problems related to sound in that setting? If so, which ones?

echo

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

no interest in that

12 More specifically, what would you want the sound experience to feel like?

only tv sound

13 What is your age group?

25-44

9 apr. 2018 15:15

1 Why do you want to watch a show/movie with friends?

For fun, conviviality and love of good entertainment and/or films

2 Do you:

Browse through Netflix

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Fun
Convivial
Epic (sometimes)

5 What is the tone of the situation?

Playful and open

6 What interactions do you expect?

Laughing
Drinking
Eating
Talking

7 What kind of environment is it?

Quiet

Inside

Shared flat

8 Do you experience any problems related to sound in that setting? If so, which ones?

Noise from the side road, friend often pause during movie (it's poopy time !), neighbours hate Dolby surround and noise.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

I would like it to feel like a simple screen, with waves indicating my friends talking, and I maintain a button to participate and hear, a bit like a talkie-walkie

Or we are seating in a movie theatre, and we can throw each over stuff like popcorn, bend and say crap, etc etc

12 More specifically, what would you want the sound experience to feel like?

Like in a theatre, or like the sound is really near me, like if I was in the middle of the action, with 3D orientation of the sound

13 What is your age group?

15-24

9 apr. 2018 14:50

1 Why do you want to watch a show/movie with friends?

Partager un bon moment; partager quelque chose que l'on a en commun.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

5 out of 7

4 In what specific words would you describe the experience?

apprécier, découvrir, rire, commenter, critiquer

5 What is the tone of the situation?

décontractée, tranquille, sympa, détendue

6 What interactions do you expect?

rire, commenter, s'étonner, ne pas être indifférent, going to the toilet ? really ?

7 What kind of environment is it?

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

Faire attention aux voisins

9 How often do you ask friends to repeat themselves after not hearing them properly?

Never

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Voir exactement la même chose, au même moment et pouvoir communiquer (pas de décalage, pas de latence)

12 More specifically, what would you want the sound experience to feel like?

il faudrait pouvoir toujours s'entendre, que le son ne prenne pas le dessus

13 What is your age group?

25-44

9 apr. 2018 12:14

1 Why do you want to watch a show/movie with friends?

To experience the movie together and discuss it while watching and afterwards.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

Social, entertaining

5 What is the tone of the situation?

Social, relaxed and fun

6 What interactions do you expect?

Talking, eating, drinking, toilet pause, laughing

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

A good sound system is important for movies. The lack of bass for example decreases the experience

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

As if my friends are at my place.

12 More specifically, what would you want the sound experience to feel like?

As if my friends are at my place, but a mute function or volume control would be handy.

13 What is your age group?

25-44

9 apr. 2018 12:14

1 Why do you want to watch a show/movie with friends?

To experience the movie together and discuss it while watching and afterwards.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

Social, entertaining

5 What is the tone of the situation?

Social, relaxed and fun

6 What interactions do you expect?

Talking, eating, drinking, toilet pause, laughing

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

A good sound system is important for movies. The lack of bass for example decreases the experience

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

As if my friends are at my place.

12 More specifically, what would you want the sound experience to feel like?

As if my friends are at my place, but a mute function or volume control would be handy.

13 What is your age group?

25-44

9 apr. 2018 11:51

1 Why do you want to watch a show/movie with friends?

Spend time with them, see a movie in a chill environment where you can comment in between, laugh too loudly and get food whenever you like (as opposed to the cinema)

2 Do you:

pick something together on the spot?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Fun
Valuable
Happy
Gezellig

5 What is the tone of the situation?

Relaxed and joyful

6 What interactions do you expect?

Chatting throughout the movie, singing if Disney, laughing, eating all types of snacks, drinking and bathroom breaks
I usually do not watch sad/scary films with friends. You want to leave the house in a happy vibe

7 What kind of environment is it?

Busy

Inside

Friendly, everything is ok (If you want to concentrate on the movie, cool, if you want to comment om everything during the movie, cool)

8 Do you experience any problems related to sound in that setting? If so, which ones?

The neighbours drilling
Sound quality of the downloaded movie, lag, bad speakers

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

All around me
As if they were here, live, without time difference

12 More specifically, what would you want the sound experience to feel like?

Similsr to reality, not like when you're gaming and it's like you're on the phone

13 What is your age group?

25-44

9 apr. 2018 11:51

1 Why do you want to watch a show/movie with friends?

To share opinion and have discussion regarding a shared experience.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

5 out of 7

4 In what specific words would you describe the experience?

Wholesome, complete, greater ability to absorb content

5 What is the tone of the situation?

Playful, relaxed

6 What interactions do you expect?

Cross-talking, eating, drinking, all of the above

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

Outside noises, not good enough sound proofing.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Real time, there s no ouldnt be any delay between what we are indiviudually experiencing.

12 More specifically, what would you want the sound experience to feel like?

Wholly absorbing, as to block out outside influences

13 What is your age group?

15-24

9 apr. 2018 11:33

1 Why do you want to watch a show/movie with friends?

To share the experience and also augment it with our commentary/reactions.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

5 out of 7

4 In what specific words would you describe the experience?

Enhanced
Shared
Relatable

5 What is the tone of the situation?

Variable

6 What interactions do you expect?

Talking
Laughing
Eating
Toilet Breaks
Snack Prep Breaks
Fidgeting
Throwing popcorn/pillows at each other

7 What kind of environment is it?

Loud

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

Cars driving by outside
Weather

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Similar, but at the same time enhanced.

12 More specifically, what would you want the sound experience to feel like?

I want it to feel more convenient, like as in being able to hear someone even if they're on the other side of the room. Also, individuals being able to turn subtitles on or off.

13 What is your age group?

15-24

9 apr. 2018 11:32

1 Why do you want to watch a show/movie with friends?

to make fun of characters and comment on clothing and other nonsense

2 Do you:

pick something together on the spot?

3 What is more important?

1 out of 7

4 In what specific words would you describe the experience?

fun(ny)

5 What is the tone of the situation?

relaxed, but playful/joyful

6 What interactions do you expect?

talking laughing eating drinking

7 What kind of environment is it?

Inside

quiet enough to understand, but also loud because we'd be talking every now and then

8 Do you experience any problems related to sound in that setting? If so, which ones?

when you eat crisps and you can't hear well enough anymore :\\\ also when you talk over something and you miss a bit, that's just annoying

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

i'd want to be able to see them and feel close to them. i want the whole experience to feel warm and funny at the same time

12 More specifically, what would you want the sound experience to feel like?

i think i'd want it to be exactly as it would be if they were sat next to me. So, i would not want the sound of the show to get quieter or louder when we're talking. it should feel as real as possible

13 What is your age group?

15-24

9 apr. 2018 11:10

1 Why do you want to watch a show/movie with friends?

Entertainment such as a movie, is a nice way to spend time and be happy.

2 Do you:

I only watch it on my laptop or download it and stream it to the TV

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Just a moment to enjoy together but at the same it isn't really as interactive as for example playing cards.

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

laughing, eating, drinking, talking

7 What kind of environment is it?

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

hmm, maybe at some crazy moments in the movie you want total silence to focus on it haah.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

interesting idea. I think that i would like to hear their reactions during exciting moments of the movie. It actually speaks what I said before of complete silence during crazy moments, but I think this is one of the things you need even if you think its undesired.

12 More specifically, what would you want the sound experience to feel like?

Like still being with friends, socially

13 What is your age group?

25-44

9 apr. 2018 10:30

1 Why do you want to watch a show/movie with friends?

To chill out

2 Do you:

pick something together on the spot?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

Excuse to get together

5 What is the tone of the situation?

Relaxed

6 What interactions do you expect?

Not too much conversation, but a feeling of being close.

7 What kind of environment is it?

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

None

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

I want to see their facial expression

12 More specifically, what would you want the sound experience to feel like?

I want to hear the small gasps and oww's from my friend. Not be alone in the VR chamber

13 What is your age group?

15-24

9 apr. 2018 10:05

1 Why do you want to watch a show/movie with friends?

share the experience

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

2 out of 7

4 In what specific words would you describe the experience?

-

5 What is the tone of the situation?

fun

6 What interactions do you expect?

talking

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

-

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

bigscreen is almost perfect for this so..... like that

12 More specifically, what would you want the sound experience to feel like?

-

13 What is your age group?

25-44

9 apr. 2018 10:05

1 Why do you want to watch a show/movie with friends?

create cinema/going out feeling

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

interesting, and allow to talk during experience

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

talking comments (cross talking) laughing

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

weak audio set up

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Overwhelming in vision AND sound

12 More specifically, what would you want the sound experience to feel like?

natural related to the space we are in (VR)

13 What is your age group?

45-64

9 apr. 2018 09:20

1 Why do you want to watch a show/movie with friends?

For the same reason I go into a cinema: it can sometimes a better experience, when there are some funny comments on movie scenes or I can see the reaction of my friends.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Fun, private watching, socialising

5 What is the tone of the situation?

Relaxed, joyful, open

6 What interactions do you expect?

Talking, laughing, discussing, looking at friends

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Not really

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

A fun evening, where we can talk an gesticulate like we are sitting beside each other. Movie should be good quality, but I think the social aspect weights more in this scenario

12 More specifically, what would you want the sound experience to feel like?

Good ;) definitely 3D sounds, so I am able to locate who of my friends is speaking right now. Maybe it would be a feature, if the movie sound will be reduced a bit when someone is speaking? But could be really annoying... Just an idea

13 What is your age group?

25-44

9 apr. 2018 08:26

1 Why do you want to watch a show/movie with friends?

to share the experience

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

??

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

going to toilet, making cups of tea

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

mild background noise from road outside

9 How often do you ask friends to repeat themselves after not hearing them properly?

Never

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

like being at the cinema with great sound in a dark environment

12 More specifically, what would you want the sound experience to feel like?

THX quality with substantial base

13 What is your age group?

25-44

9 apr. 2018 08:14

1 Why do you want to watch a show/movie with friends?

to spend time together and share the same topics

2 Do you:

pick something together on the spot?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

time consuming

5 What is the tone of the situation?

depends on the atmosphere of the movies

6 What interactions do you expect?

talking murmuring snacking drinking

7 What kind of environment is it?

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

the movie sound being to loud to bother our parents or other roommates

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Friends would text each other to enter the same virtual room at a specific time. Once everyone is ready, the host will start playing the movie.
During movie time, participants are allow to chat and murmur. But other participants have the right to mute the speaking sound or adjust the volume.

12 More specifically, what would you want the sound experience to feel like?

360 sourrounding

13 What is your age group?

15-24

9 apr. 2018 08:05

1 Why do you want to watch a show/movie with friends?

Everything is better with other people.

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Not sure what you're asking here.

5 What is the tone of the situation?

Playful

6 What interactions do you expect?

Depends on the group of people and the type of content being watched.

7 What kind of environment is it?

All of the above? Depends on people/content.

8 Do you experience any problems related to sound in that setting? If so, which ones?

You can't gauge how loud the other people have their movie sound vs voices vs ambient room. In a normal environment you are sharing the space so you know if you need to raise your voice to be heard but it's not quite the same in VR.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Not really sure... I don't think it has to necessarily emulate a TV or a theater.

12 More specifically, what would you want the sound experience to feel like?

13 What is your age group?

25-44

9 apr. 2018 07:25

1 Why do you want to watch a show/movie with friends?

Rigoler ensemble, discuter du film en meme temps

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

C'est rigolo, on peut débattre du sujet aussi

5 What is the tone of the situation?

Détendu, marrant, blague

6 What interactions do you expect?

Parler, rire, manger et boire

7 What kind of environment is it?

Loud

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

Pas vraiment. Son parfois pas assez fort du film ?

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Un moment qu'on partage en privé juste entre nous, plutot ambiance posée (ou l'on parle peu) ou rigolote (ou genre on fait les doublages en meme temps, etc)

12 More specifically, what would you want the sound experience to feel like?

Rien de particulier. Pouvoir moduler le son du film et des amis

13 What is your age group?

25-44

9 apr. 2018 06:07

1 Why do you want to watch a show/movie with friends?

To laugh

2 Do you:

pick something together on the spot?

3 What is more important?

4 out of 7

4 In what specific words would you describe the experience?

Chillin

5 What is the tone of the situation?

Relaxed playful

6 What interactions do you expect?

Talking laughing

7 What kind of environment is it?

Inside

Loud

8 Do you experience any problems related to sound in that setting? If so, which ones?

No

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Stereo

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Social

12 More specifically, what would you want the sound experience to feel like?

Like it would in a normal living room

13 What is your age group?

15-24

9 apr. 2018 05:51

1 Why do you want to watch a show/movie with friends?

as a celebration after work

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

immersive, and engaging

5 What is the tone of the situation?

anticipating

6 What interactions do you expect?

eating, shouting

7 What kind of environment is it?

Loud

Busy

8 Do you experience any problems related to sound in that setting? If so, which ones?

yes other laugher

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

able to see his emotions or how he feels like even though it is a virtual

12 More specifically, what would you want the sound experience to feel like?

individualize

13 What is your age group?

25-44

9 apr. 2018 05:36

1 Why do you want to watch a show/movie with friends?

For the social interaction and the post discussion

2 Do you:

pick something together on the spot?

3 What is more important?

3 out of 7

4 In what specific words would you describe the experience?

Engaging

5 What is the tone of the situation?

playful

6 What interactions do you expect?

laughing, talking

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

problems with the surround sound

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

like they are there and not in a different place. maybe good microphone to movie balance.

12 More specifically, what would you want the sound experience to feel like?

Balanced between movie and discussion

13 What is your age group?

15-24

9 apr. 2018 05:18

1 Why do you want to watch a show/movie with friends?

long distance

2 Do you:

choose in advance what you want to watch together?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

better than txting or talking on the phone

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

like you're on the couch together, but not

7 What kind of environment is it?

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

not really

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Like you're watching together (dunno how else to say it)

12 More specifically, what would you want the sound experience to feel like?

Talk to each other and still hear the movie

13 What is your age group?

25-44

9 apr. 2018 05:04

1 Why do you want to watch a show/movie with friends?

Sharing the viewing experience with others is more enjoyable

2 Do you:

pick something together on the spot?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

Depends on the content. A movie (especially a new one) is usually quiet, dark, movie theaterish... our tv viewing is usually more social

5 What is the tone of the situation?

Relaxed is pretty key. I think that's why most people pick this kind of pass time

6 What interactions do you expect?

Once again, it depends on the content. I'd hope a new movie would have very few distractions and slim to none as far as chit chat. On the other hand, watching an episode of family guy could be pretty free-reign.

7 What kind of environment is it?

Quiet

Inside

8 Do you experience any problems related to sound in that setting? If so, which ones?

I love movies, and I'm also kind of an audiophile myself, so when I'm limited to my apartment living room (and only the tv to produce sound), the viewing experience can be kind of underwhelming. For this reason I avoid watching a lot of my favorites here (imagine how disrespectful it'd be to Hans Zimmer if I flipped on Intersteller and limited his art to my 10 year old Vizio plasma). Also, my neighbors throw a lot of parties, so watching tv has to work with their schedule too... I would pay A LOT of money for a kickass VR movie experience.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Often

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

Authentic... yet tailored just enough to eliminate the problem you brought up in the previous question (asking them to repeat themselves).

12 More specifically, what would you want the sound experience to feel like?

I'd like the sound to be immersive. High quality sound in a vr headset is a non-negotiable for me.

13 What is your age group?

15-24

9 apr. 2018 04:56

1 Why do you want to watch a show/movie with friends?

An attempt to spread in an experience one of us enjoys. Or share something we both enjoy.

2 Do you:

pick something together on the spot?

3 What is more important?

6 out of 7

4 In what specific words would you describe the experience?

Light-hearted critisicm and discussion.

5 What is the tone of the situation?

Casual critiquing.

6 What interactions do you expect?

Talking, cross-talking, theme discussion, comparison, rating, errors, plot devices, mistakes.

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

Other than occasionally talking over one another, no.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Sometimes

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

I would want it to be a rough analogy. Distractions are a must. Watching the screen isn't as fun if you can't break away for a moment. It wouldn't feel 'real' enough.

12 More specifically, what would you want the sound experience to feel like?

Having a 'simulated sound' option where everyone hears it the same no matter where they are would probably work best, and then a placeable speaker option would be neat for contrast. It would be fun, and frustrating to have to try to set up a VR surround system.

13 What is your age group?

25-44

9 apr. 2018 04:21

1 Why do you want to watch a show/movie with friends?

To fill in the void. I get bored watching stuff alone sometimes.

2 Do you:

All of the above.

3 What is more important?

5 out of 7

4 In what specific words would you describe the experience?

Chill

5 What is the tone of the situation?

Relaxed & Open

6 What interactions do you expect?

Everything including jokes but for the love of God, don't try to over talk an already going conversation.

7 What kind of environment is it?

Inside

Quiet

8 Do you experience any problems related to sound in that setting? If so, which ones?

I prefer loud volume while others say it's too loud.

9 How often do you ask friends to repeat themselves after not hearing them properly?

Always

10 In the context of watching TV with friends, what sound quality would you prefer coming from the TV?

Surround/Home cinema

11 If you were to watch a show/movie with friends in virtual reality, what would you want the experience to feel like?

IDK. I literally use Bigscreen every single day. Whatever happens, happens. It's just nice to know that you can mute people and kick/ban then from your room.

12 More specifically, what would you want the sound experience to feel like?

Like I'm in the center regardless of where I'm sitting .

13 What is your age group?

15-24

APPENDIX 1.F.

QUESTIONS QUESTIONNAIRE 2

1. Why do you want to watch a show/movie with friends in VR?
2. Compared to real life, is the experience:
    - As good
    - Better
    - Worse
3. Which elements are better than in real life?
4. Which elements are worse than in real life?
5. What is the tone of the situation?
6. What interactions do you expect?
7. What are you looking for in terms of qualities when hearing a sound in VR?
8. Which map do you usually use?
    - Grand Theatre
    - Big Cinema
    - Home Theatre
    - Living Room
    - Lobby
    - The Void
9. Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)
    - Yes
    - No
10. Do you have any problems comprehending what your friends are saying?
11. Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?
    - Yes
    - No
12. What is more important?
Choose between 1, 2, 3, 4, 5, 6, 7.
Sound immersion (1) ----- Intelligibility (7)
13. What is more important?
Choose between 1, 2, 3, 4, 5, 6, 7.
The social aspect (1) ----- The content (7)
14. What would your recommendations be to improve the sound in BigscreenVR?
15. What is your age group?
    - 15-24
    - 25-44
    - 45-64
    - 65+

APPENDIX 1.G.

RAW DATA QUESTIONNAIRE 2

There were 12 filled in questionnaires.

18 apr. 2018 14:14

1 Why do you want to watch a show/movie with friends in VR?

The big cinema screen makes it great. Less effort of organizing it, it's free (don't have to pay for cinema tickets), we can talk together and use hand-gestures, and afterwards we have the freedom to do other things in bigscreen after the movie (like youtube).
A big bonus is also that other strangers can join in and watch with us if we choose to have an open room.

2 Compared to real life, is the experience:

better

3 Which elements are better than in real life?

The general social aspect of it, as I'm in control of the movie, we can pause the movie, talk more, take breaks, do other things.
In general the freedom you have when you're in control of the entire cinema, and the futuristic aspect of being in virtual reality (in a VR cinema you kinda forget about real-life more than you do in a normal cinema, as you don't see your real body) as well as the aspect of joining in on strangers watching a movie, instantly getting together well.
In VR it's also more social than a normal cinema, in my experience, perhaps because there are less people, it feels less formal, the movie can be paused, and it's free (also you can teleport to a seat further away if you don't want to listen to talking). It feels a lot more personal than a massive cinema too, last time i watched a movie with some strangers, we became friends quickly, and as an example, we would pause the movie if someone needed to go get something or go to the toilet, something you cant normally do in a cinema.

4 Which elements are worse than in real life?

The comfort, the HMD gets a bit uncomfortable for many, this will likely improve as VR gets more popular and hardware gets better, however.

5 What is the tone of the situation?

Playful, relaxed and open.

6 What interactions do you expect?

Talking, laughing, gesturing to each others, moving about in one's seat, some people going to the toilet every now and again, drinking some beverage.

7 What are you looking for in terms of qualities when hearing a sound in VR?

Bineural audio, so that it sounds real and can be positioned. If you're talking about added sound-effects that aren't part of the movie, like fake "crowd chatter" in the background, it should be realistic but un-interruptive, preferrably you should be able to disable such extra sound.

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

No.

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

5 out of 7

13 What is more important?

2 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Implement bineural audio if it isn't already.

15 What is your age group?

15-24

17 apr. 2018 02:07

1 Why do you want to watch a show/movie with friends in VR?

Companionship. I fall asleep if I watch a movie alone.

2 Compared to real life, is the experience:

better

3 Which elements are better than in real life?

I can leave any time I want.

4 Which elements are worse than in real life?

Not really as "social" as IRL but close. Also, my real life friends are not in VR.

5 What is the tone of the situation?

What situation? I don't know what this question means.

6 What interactions do you expect?

Laughing, talking, taking a break / intermission.

7 What are you looking for in terms of qualities when hearing a sound in VR?

Good volume and audio quality, directionality.

8 Which map do you usually use?

Campfire or cosmos.

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

Not unless something is set up wrong.

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

No

12 What is more important?

7 out of 7

13 What is more important?

4 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

It sounds good to me. If there were a way to emulate more channels (5, 7, atmos) through stereo headphones that would be cool. I would not spend more money to get it.

15 What is your age group?

45-64

13 apr. 2018 19:59

1 Why do you want to watch a show/movie with friends in VR?

commentary

2 Compared to real life, is the experience:

as good

3 Which elements are better than in real life?

comfort of being at home

4 Which elements are worse than in real life?

Quality of Video and Sound

5 What is the tone of the situation?

relaxed

6 What interactions do you expect?

All

7 What are you looking for in terms of qualities when hearing a sound in VR?

7.1

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

Sometimes the movie drowns out talking

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

5 out of 7

13 What is more important?

3 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Make it so you can set preference for talk over movie sound or the other way around

15 What is your age group?

25-44

12 apr. 2018 23:20

1 Why do you want to watch a show/movie with friends in VR?

Big screen and an amazingly realistic online social experience.

2 Compared to real life, is the experience:

worse

3 Which elements are better than in real life?

I can be sitting in my own home and "leave" the social experience whenever I want.

4 Which elements are worse than in real life?

Not quite as immersive, things in the "real world" still force you back to "reality".

5 What is the tone of the situation?

It can be all of the above, though I generally use it in a relaxed, playful way.

6 What interactions do you expect?

talking, laughing, singing, moving, working, all of this stuff would be amazing.

7 What are you looking for in terms of qualities when hearing a sound in VR?

To be able to know where it's coming from is important. Also, it should sound lifelike, and I've noticed a lot of VR experiences don't spend too much time on that.

8 Which map do you usually use?

Grand Theater

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

Most of the time it works well.

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

4 out of 7

13 What is more important?

3 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Appropriate ambiance noises would be nice, such as crackling of fire. Also, weather might add an interesting aspect to the experience, but these should be able to be adjusted (for volume or turned off all together) to fit different situations.

15 What is your age group?

25-44

12 apr. 2018 20:44

1 Why do you want to watch a show/movie with friends in VR?

I enjoy the out of home experience alot however the social aspect makes it really worth it. You meet some awesome people and can enjoy a movie together. Feels like going to an actual movie theater except with the addition of the social aspect which is rewarding. When you goto a movie theater solo you cant socialize, you dont know the person next to you, its akward. On Vr in my case BIGSCREEN, its ok to meet others socialize and enjoy a flick.

2 Compared to real life, is the experience:

better

3 Which elements are better than in real life?

Social for sure, not having to worry about children or anyone acting stupid, though in vr there's plenty of trolling. For the most part I thoroughly enjoy it though.

4 Which elements are worse than in real life?

Sound glitches sometimes, for me when I host a room people can hear my mic and we can socialize however video audio is non-existent. That's an issue I will try to trouble shoot myself but really there is nothing worse about it. Its better in just about every way!

5 What is the tone of the situation?

depends what room your in and who is in it. A-lot of the time its playful and relaxed, and definitely open. I've had some really relaxed experiences where we watched an entire flick occasionally conversing and afterwards discussing the movie. it truly is enjoyable.

6 What interactions do you expect?

talking, cross-talking, laughing. All of which I find enjoyable.

7 What are you looking for in terms of qualities when hearing a sound in VR?

Crisp sound, and 3d audio would be awesome. I haven't paid as much attention to if that was already a thing but hearing surround sound in it would be awesome. It has that for the mics, someone on your left talks you hear it in your left ear, same with the right. If your sitting far from someone it mimics that really well too.

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

None whats so ever.

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

3 out of 7

13 What is more important?

3 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Video audio feeling like its coming from speakers all over like in a movie theater. They have big speakers on the walls and in the corners. Really makes you feel like your in it... except not even close to the level achieved by VR

15 What is your age group?

25-44

9 apr. 2018 17:55

1 Why do you want to watch a show/movie with friends in VR?

To comment silly and trivial things like we were sitting on a couch watching tv together

2 Compared to real life, is the experience:

worse

3 Which elements are better than in real life?

more privacy, I don't have to leave my house, we can stay together even if we're far

4 Which elements are worse than in real life?

resolution, comfort, audio, lag, video quality

5 What is the tone of the situation?

relaxed and joyful

6 What interactions do you expect?

talking, laughing, seeing each other reactions

7 What are you looking for in terms of qualities when hearing a sound in VR?

I really need it to be 3D and high quality

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

A lot because of the lag and the varying audio quality depending on the person's connection

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

No

12 What is more important?

6 out of 7

13 What is more important?

3 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Quality improvement, lag reduction

15 What is your age group?

25-44

9 apr. 2018 12:37

1 Why do you want to watch a show/movie with friends in VR?

Because it is nicer to watch a movie with others and VR is a great way of meeting others, who are in a different physical location

2 Compared to real life, is the experience:

as good

3 Which elements are better than in real life?

I'm still at home, the screen is bigger

4 Which elements are worse than in real life?

Resolution, reduced nonverbal communication

5 What is the tone of the situation?

fun, relaxed

6 What interactions do you expect?

talking, laughing, sharing stuff

7 What are you looking for in terms of qualities when hearing a sound in VR?

proper 3D Sound

8 Which map do you usually use?


Home Theater

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

The microphone quality of the Vive is pretty bad, but it's ok and most of the times everything is comprehendable

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

5 out of 7

13 What is more important?

4 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

better spatial sound

15 What is your age group?

25-44

9 apr. 2018 10:10

1 Why do you want to watch a show/movie with friends in VR?

share the experience

2 Compared to real life, is the experience:

worse

3 Which elements are better than in real life?

ease of not being there in person

4 Which elements are worse than in real life?

not bring able to see the person, share food, etc. hassle of headset and controls.

5 What is the tone of the situation?

fun

6 What interactions do you expect?

talking

7 What are you looking for in terms of qualities when hearing a sound in VR?

positional audio for people talking but don't need it from media content.

8 Which map do you usually use?

Grand Theater

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

Yes

10 Do you have any problems comprehending what your friends are saying?

sometimes - hard to find right audio balance

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

7 out of 7

13 What is more important?

2 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

-

15 What is your age group?

25-44

9 apr. 2018 09:50

1 Why do you want to watch a show/movie with friends in VR?

Friends live in different city

2 Compared to real life, is the experience:

as good

3 Which elements are better than in real life?

Less travel, less expensive, greater control over the experience, greater amount of Privacy, security issues are a non issue.

4 Which elements are worse than in real life?

Specifically the 3 D graphics in bigscreen are not the greatest, check out cinevr to compare the graphics

5 What is the tone of the situation?

For entertainment purposes

6 What interactions do you expect?

Talking

7 What are you looking for in terms of qualities when hearing a sound in VR?

I utilize the Dolby atmos app on my PC, so current 3d sound presentation is improving

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

No

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

2 out of 7

13 What is more important?

7 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

The best 3d sound through headphones utilize hrtf.

15 What is your age group?

45-64

9 apr. 2018 05:29

1 Why do you want to watch a show/movie with friends in VR?

Social chat about the movie

2 Compared to real life, is the experience:

worse

3 Which elements are better than in real life?

People far away. Immersion. Big screen.

4 Which elements are worse than in real life?

Visual fidelity, comfort, natural emotion.

5 What is the tone of the situation?

Relaxed

6 What interactions do you expect?

All of the above

7 What are you looking for in terms of qualities when hearing a sound in VR?

Spatialized realistic sound

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

Background noise of the person's room

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

1 out of 7

13 What is more important?

4 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

positional audio of the screen (5.1 / 7.1)

15 What is your age group?

25-44

9 apr. 2018 04:28

1 Why do you want to watch a show/movie with friends in VR?

Cause I don't have friends IRL.

2 Compared to real life, is the experience:

better

3 Which elements are better than in real life?

Freedom, space, comfortability.

4 Which elements are worse than in real life?

Cleaning up afterwards.

5 What is the tone of the situation?

Playful, relaxed, open.

6 What interactions do you expect?

Everything including weirdos jacking off. I hate it when they do that. KICK!

7 What are you looking for in terms of qualities when hearing a sound in VR?

Binaural Audio but that sorely depends on the source of the audio.

8 Which map do you usually use?

Friends = Home Theater. I Host = Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

Yes

10 Do you have any problems comprehending what your friends are saying?

Currently, yes as Bigscreen has it's own set of issues on-top of people's background noise. But with the people I talk to on a daily basis, no. We all come through crystal clear to each other.

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

5 out of 7

13 What is more important?

5 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

For audio to be played as if you were sitting in the center of the cinema regardless of your actual position.

15 What is your age group?

15-24

9 apr. 2018 02:49

1 Why do you want to watch a show/movie with friends in VR?

Share the experience

2 Compared to real life, is the experience:

better

3 Which elements are better than in real life?

No rude people in the theater, no malpractices on the owner's part, complete control over playback and additional content, comfortable as your home. Also, crosstall free stereo 3D and no crappy dubbing (here in Italy they only release dubbed content) !

4 Which elements are worse than in real life?

Resolution for now, but it'll change with time (I backed the piMax 8k on kickstarter and should receive it in the next months)

5 What is the tone of the situation?

It depends on the content, Eg. discussing breaching tactics and weapon systems professionally if watching a training video or making jokes if watching a comedy

6 What interactions do you expect?

Talking without disrupting the dialogues, also having 3D props to play with, possibly from the feature that is being watched

7 What are you looking for in terms of qualities when hearing a sound in VR?

As said in the ither questionnaire, optimizing the current hrtf to emulate a full 49.4 dolby atmos

8 Which map do you usually use?

Big Cinema

9 Would you like to hear yourself through the headphones/earbuds while talking? (sidetone)

No

10 Do you have any problems comprehending what your friends are saying?

Only happens when people don't use the hmds integrated mic

11 Would you like to be more immersed through better room ambiance effects (reverberation, early reflections, background noise)?

Yes

12 What is more important?

7 out of 7

13 What is more important?

7 out of 7

14 What would your recommendations be to improve the sound in BigscreenVR?

Optimize hrtf and support 49.4 dolby atmos emulation

15 What is your age group?

25-44

APPENDIX 2.A.

PROTOTYPE CHANGES

To improve the user test and prototype, it was shown to my supervisors at TU Delft, and modified afterwards. The other parts are detailing what the prototype ended up being for the user tests with the participant.

Changes made:
- Clarify the user test better to the participants from the start. What will they test, what the focus of this project is on, how the test is divided.
- There needs to be a neutral environment at the start of the experiment to let the participant adapt to VR and make sure the HMD and headphones are positioned properly.
- The sounds initially started immediately, it did not take into account the time required to put the HMD on. Added a delay and a voice introducing the start of the experiment/when to take the HMD off.
- The room at this point still launched in seemingly random positions. This needed to be fixed to provide a consistent experience.
- To learn about how they feel about the experience, ask them to speak aloud during the experience. This later on during the actual user tests did not work, participants were too distracted and busy to think aloud. They were still asked to think aloud, but after taking their HMD, additionnal questions were asked to let them tell what they experienced.
- The version they tested was rather short, around 1 minute for each environment. This needed to be increased to about 5 minutes so the participants have enough time to experience it.
- Suggestion to take notes with times already written on the paper to make it easier to know what happened when.
- Bigger variety of sounds, with, for example, triggers that sound high pitched since those are easier to localise than low pitch sounds.
- Remove the tripods from the images when you look down.

APPENDIX 2.B.

QUESTIONNAIRE

1→ Which test was this (for the interviewer)? *

- A 1.1
- B 1.2
- C 1.3
- D 2.1
- E 2.2
- F 2.3
- G 3.1
- H 3.2
- I 3.3

2→ You felt like you were with another person *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

3→ You interacted with another person *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

4→ Another person interacted with you *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

5→ You felt connected to the other person *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

6→ It felt like the other person was close to you (in terms of distance) *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

7→ You felt like you were physically in that environment *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

8→ Thinking back to the experience, you remember it as images rather than as somewhere you have been *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

9→ It felt more like being in a virtual environment than being elsewhere in person *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

10→ You felt relaxed in that environment *

1	2	3	4	5

Strongly disagree — Neutral — Strongly agree

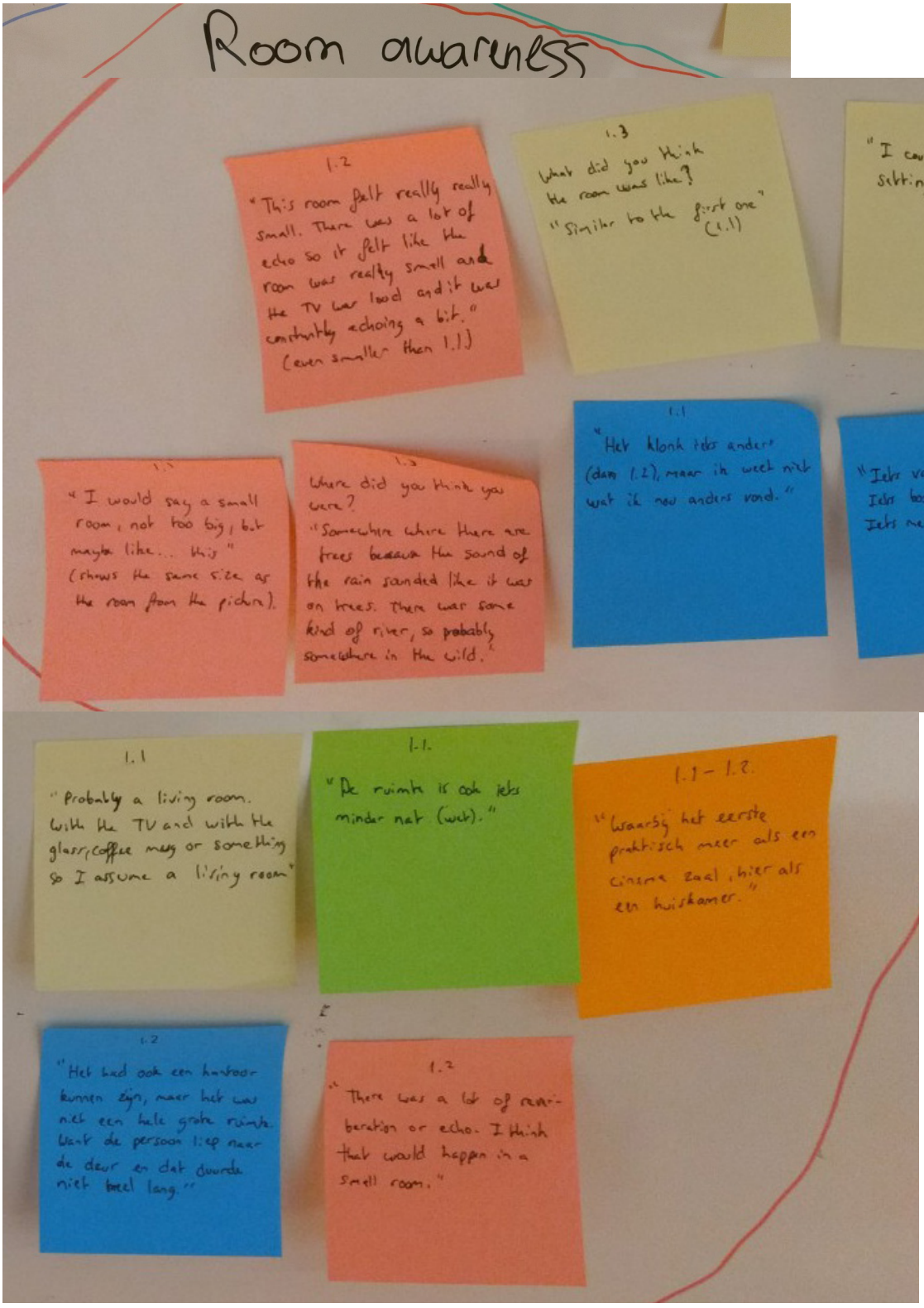
APPENDIX 2.C.

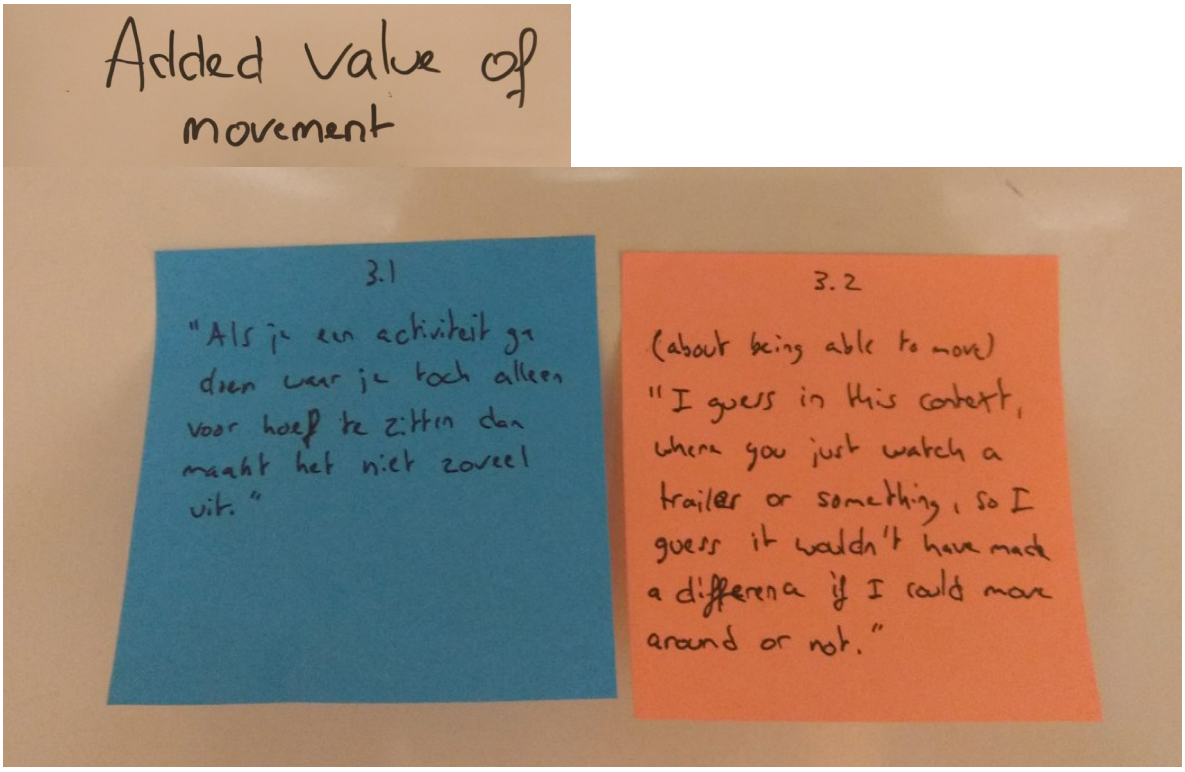
RAW DATA QUESTIONNAIRE

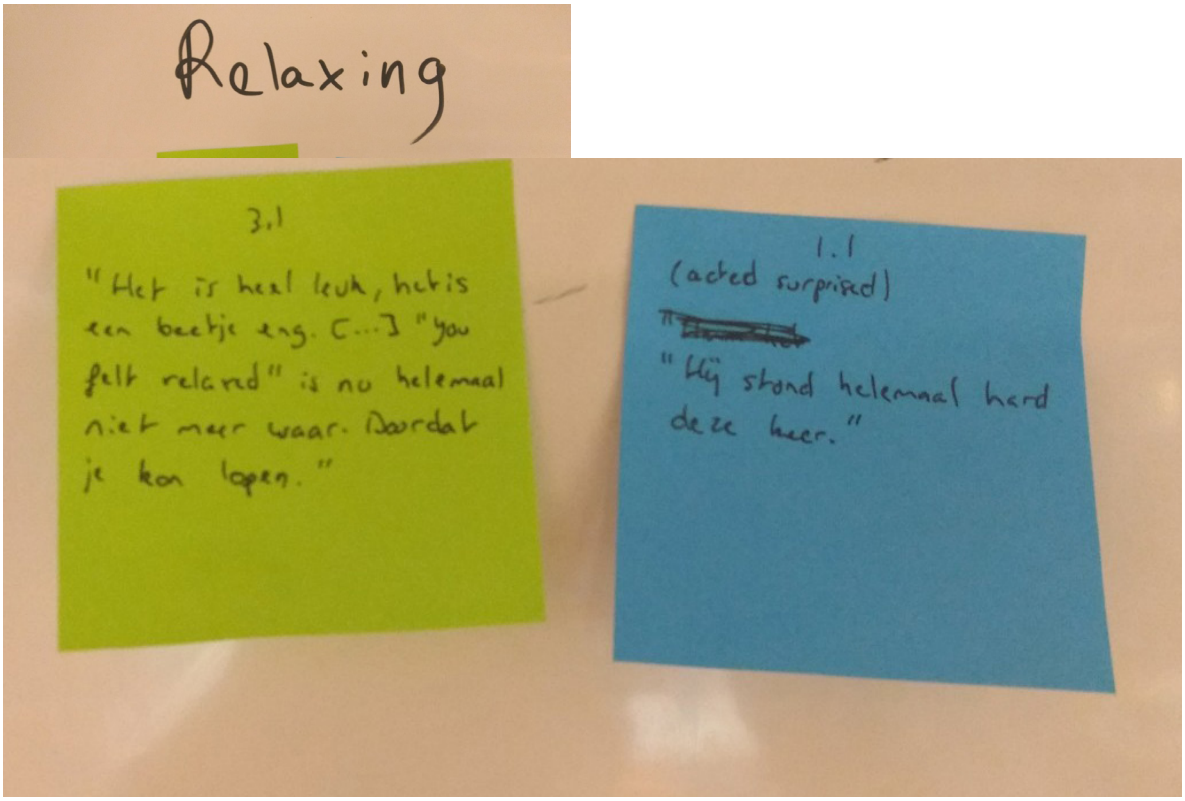
#	Which test was this (for the inter-viewer)?	You felt like you were with another person	You inter-acted with another person	Another person interact-ed with you	You felt connect-ed to the other person	It felt like the other person was close to you (in terms of distance)	You felt like you were physically in that environ-ment	You remember it as images rather than as some-where you have been	It felt more like being in a virtual environment than being elsewhere in person	You felt relaxed in that environ-ment	Submit Date (UTC)
Participant 5	3.2	2	3	4	3	4	2	3	5	3	2018-08-07 14:27:57
Participant 5	2.2	4	3	4	3	4	4	3	3	4	2018-08-07 14:22:32
Participant 5	1.2	4	3	4	3	4	2	2	2	3	2018-08-07 14:14:33
Participant 5	1.3	4	3	4	3	4	3	2	4	3	2018-08-07 14:08:13
Participant 5	1.1	5	2	4	4	5	1	1	2	3	2018-08-07 14:00:24
Participant 4	3.1	1	1	1	1	2	1	5	5	1	2018-08-07 13:10:02
Participant 4	2.1	2	2	2	2	2	2	5	5	3	2018-08-07 13:01:08
Participant 4	1.1	1	1	1	1	2	1	5	5	3	2018-08-07 12:54:19
Participant 4	1.2	1	1	1	1	3	1	5	5	3	2018-08-07 12:42:54
Participant 4	1.3	3	1	1	3	5	1	3	5	3	2018-08-07 08:55:25
Participant 3	3.2	4	4	3	4	4	4	3	4	2	2018-07-30 15:20:24
Participant 3	2.2	4	5	4	4	5	5	2	3	4	2018-07-30 15:10:36
Participant 3	1.2	4	3	4	4	4	4	3	5	4	2018-07-30 15:01:19
Participant 3	1.3	4	2	4	3	4	4	3	4	4	2018-07-30 14:54:05
Participant 3	1.1	3	1	3	4	4	3	5	4	4	2018-07-30 14:46:09
Participant 2	3.1	2	2	3	4	4	3	2	2	4	2018-07-30 13:58:37
Participant 2	2.1	2	2	3	3	4	4	2	3	4	2018-07-30 13:43:53
Participant 2	1.1	4	2	2	4	4	4	2	3	4	2018-07-30 13:31:52
Participant 2	1.2	3	2	3	3	4	4	4	2	4	2018-07-30 13:23:29
Participant 2	1.3	4	2	3	4	4	1	5	4	3	2018-07-30 13:12:50
Participant 1	3.3	1	1	1	1	1	3	4	4	4	2018-07-30 12:53:51
Participant 1	2.3	2	1	2	2	4	4	2	2	4	2018-07-30 12:45:39
Participant 1	1.3	2	1	2	3	4	4	4	2	4	2018-07-30 12:36:28
Participant 1	1.2	2	1	2	2	4	3	4	4	4	2018-07-30 12:28:03
Participant 1	1.1	2	1	2	2	3	3	4	4	3	2018-07-30 12:13:31

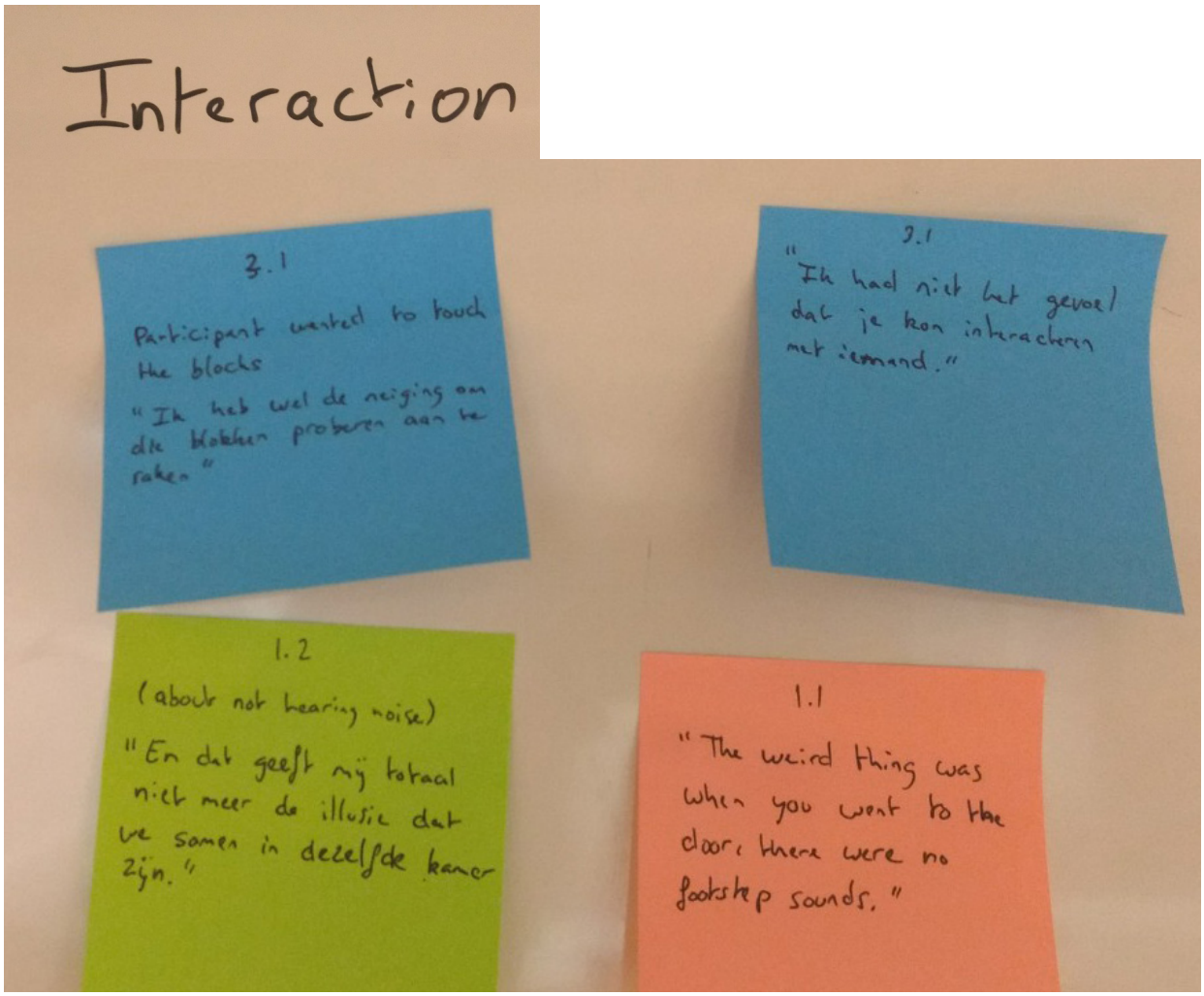
APPENDIX 2.D.

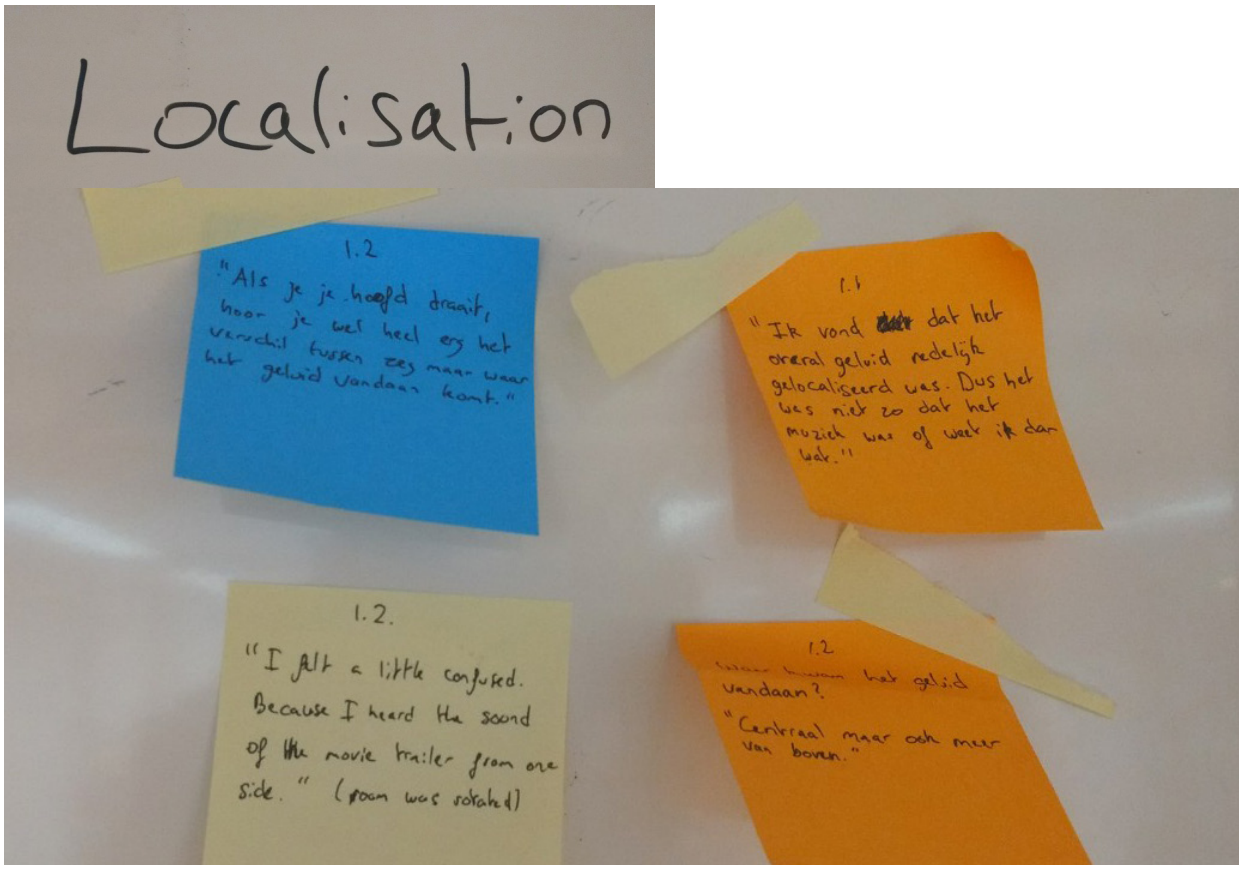
STATEMENTS

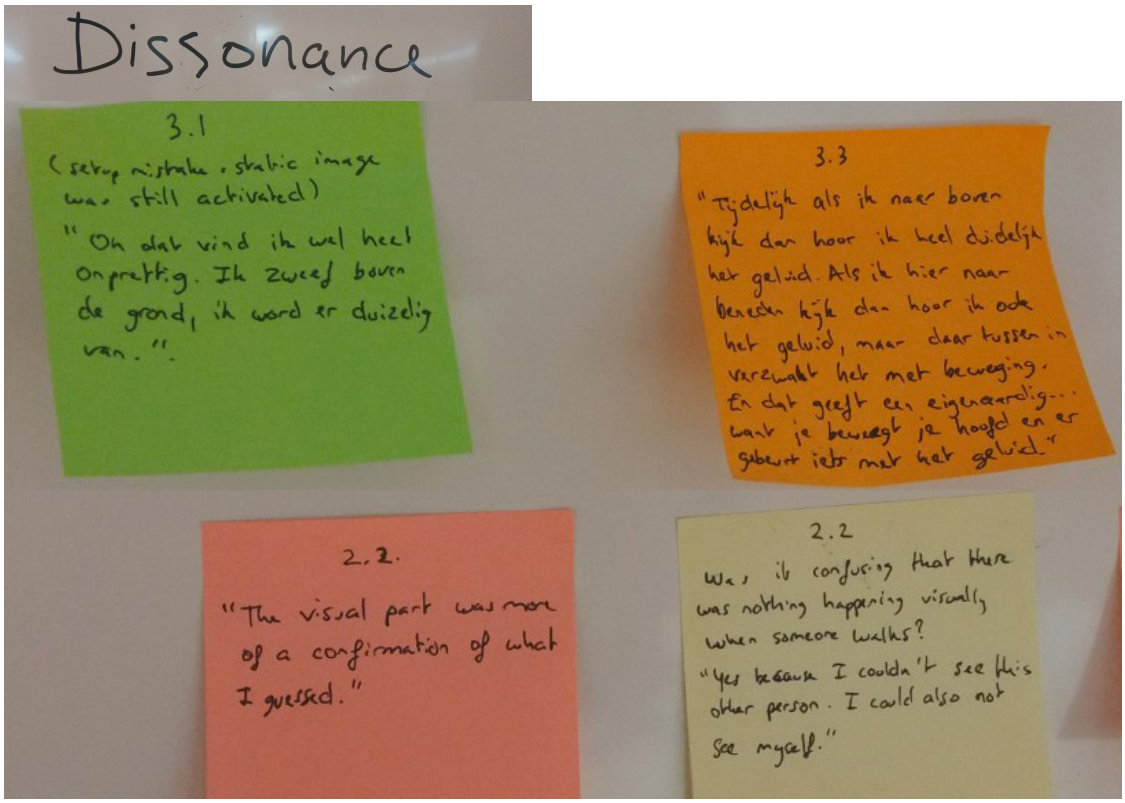


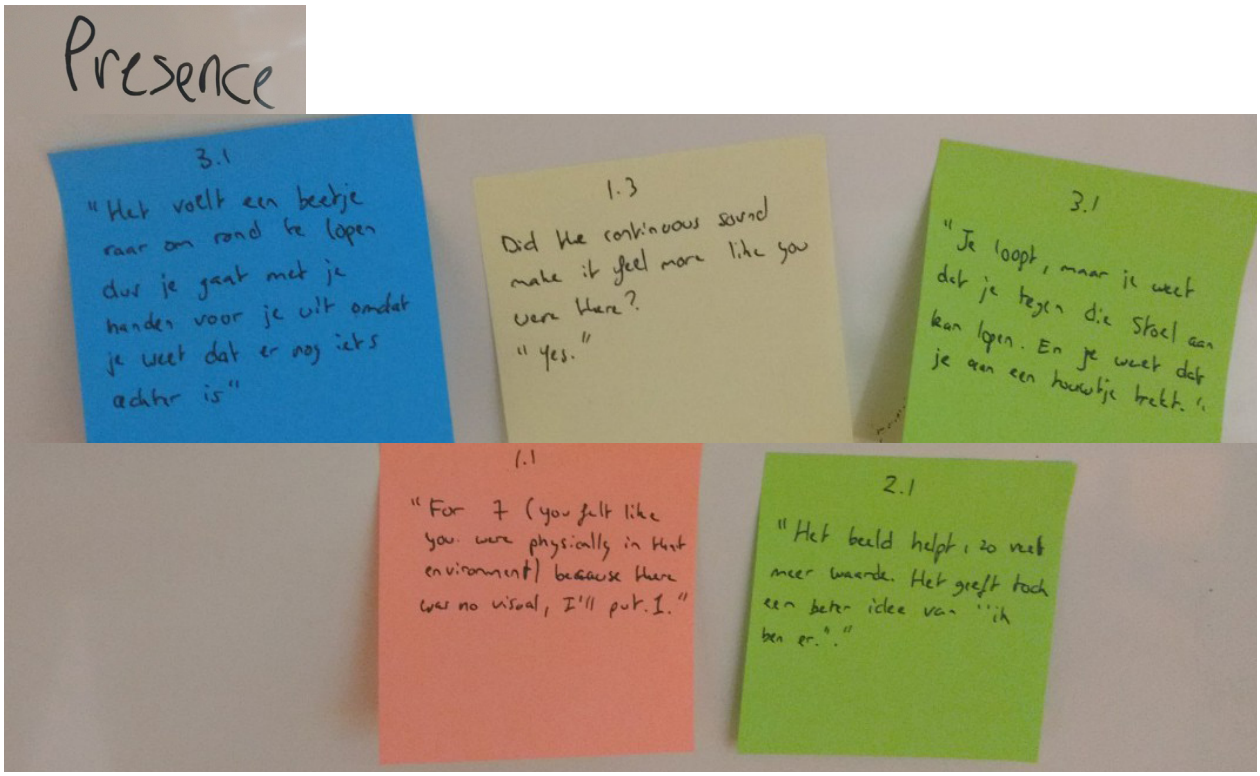

Added value of movement
3.1
"Als je een activiteit ga doen waar je toch alleen voor hoef te zitten dan maakt het niet zoveel uit."
3.2
(about being able to move)
"I guess in this context, where you just watch a trailer or something, so I guess it wouldn't have made a difference if I could move around or not."

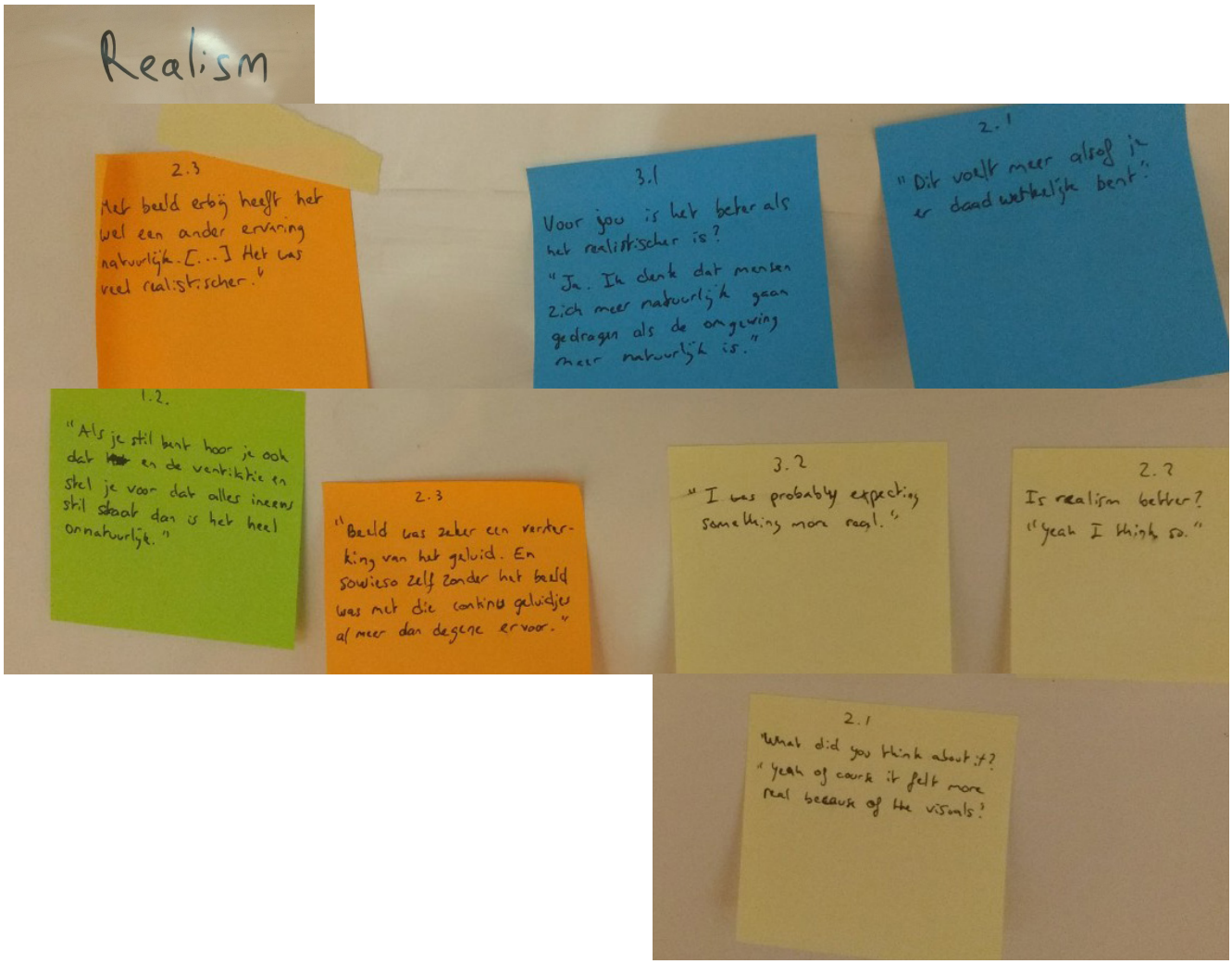

Relaxing
3.1
"Het is heel leuk, het is een beetje eng. [...] "you felt relaxed" is nu helemaal niet meer waar. Doordat je kon lopen."
1.1
(acted surprised)
"Hij stond helemaal hard deze keer."

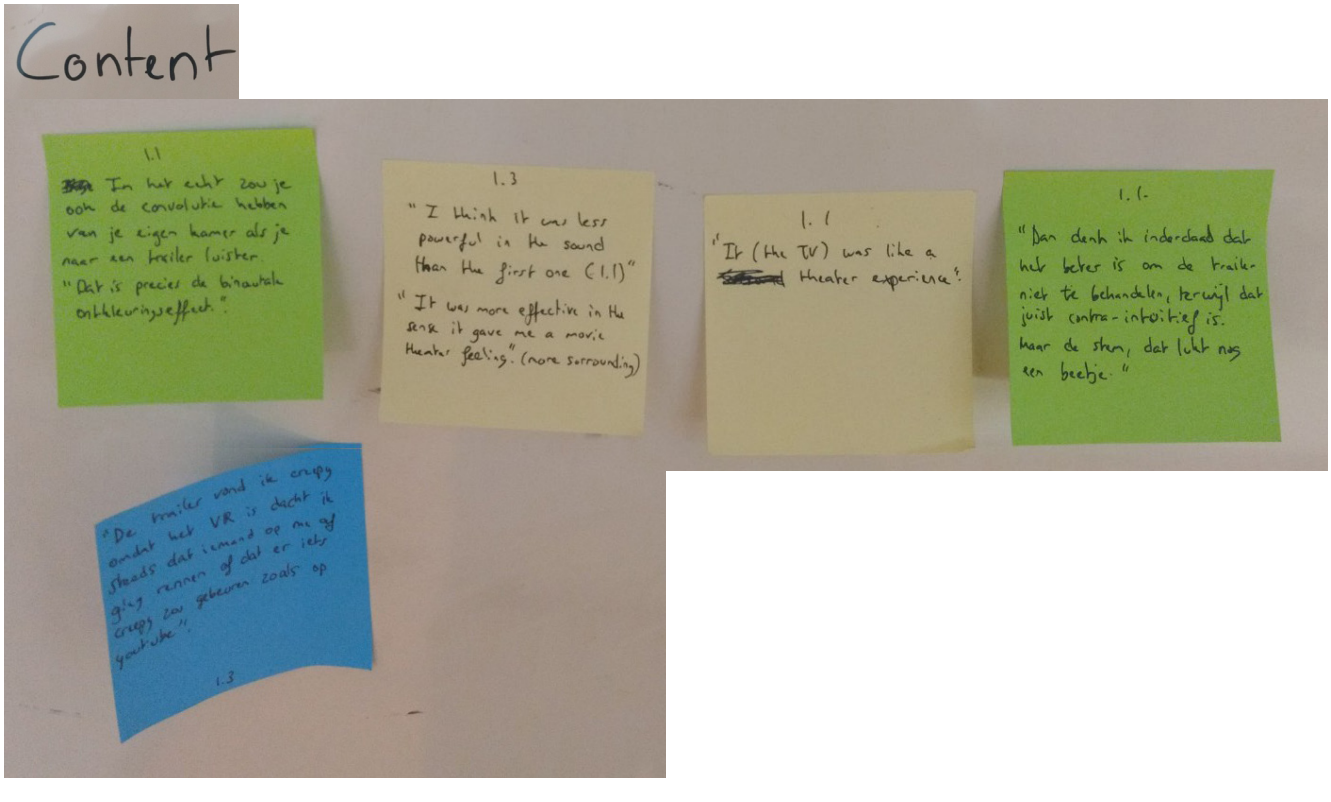

Interaction
3.1
Participant wanted to touch the blocks
"Ik heb wel de neiging om die blokken proberen aan te raken"
2.1
"Ik had niet het gevoel dat je kon interacteren met iemand."
1.2
(about not hearing noise)
"En dat geeft mij totaal niet meer de illusie dat we samen in dezelfde kamer zijn."
1.1
"The weird thing was when you went to the door, there were no footstep sounds."

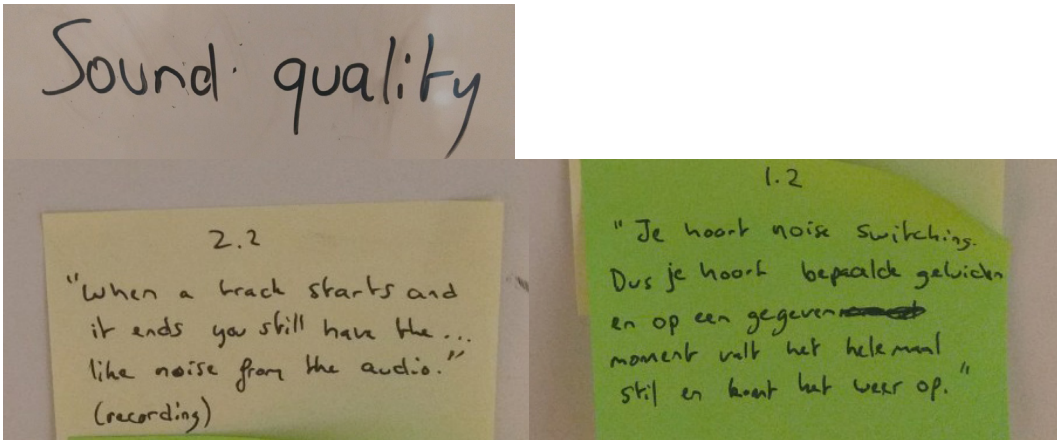

Localisation
1.2
"Als je je hoofd draait, hoor je wel heel erg het verschil tussen zes maar waar het geluid vandaan komt."
1.1
"Ik vond dat het overal geluid redelijk gelocaliseerd was. Dus het was niet zo dat het muziek was of weet ik dan wat."
1.2.
"I felt a little confused. Because I heard the sound of the movie trailer from one side." (room was rotated)
1.2
vandaan?
"Centraal maar ook meer van boven."


Dissonance
3.1
(setup mistake, static image was still activated)
" Oh dat vind ik wel heel onprettig. Ik zweef boven de grond, ik word er duizelig van."
3.3
" Tijdelijk als ik naar boven kijk dan hoor ik heel duidelijk het geluid. Als ik hier naar beneden kijk dan hoor ik ook het geluid, maar daar tussen in verzwakt het met beweging. En dat geeft een eigenaardig... want je beweegt je hoofd en er gebeurt iets met het geluid."
2.2.
" The visual part was more of a confirmation of what I guessed."
2.2
Was it confusing that there was nothing happening visually when someone walks?
" Yes because I couldn't see this other person. I could also not see myself."


Presence
3.1
" Het voelt een beetje raar om rond te lopen dus je gaat met je handen voor je uit omdat je weet dat er nog iets achter is"
1.3
Did the continuous sound make it feel more like you were there?
" yes."
3.1
" Je loopt, maar je weet dat je tegen die stoel aan kan lopen. En je weet dat je aan een hoekje trekt."
1.1
" For 7 (you felt like you were physically in that environment) because there was no visual, I'll put 1."
2.1
" Het beeld helpt, zo veel meer waarde. Het geeft toch een beter idee van "ik ben er."."


Realism
2.3
Met beeld erbij heeft het wel een ander ervaring natuurlijk. [...] Het was veel realistischer."
3.1
Voor jou is het beter als het realistischer is?
"Ja. Ik denk dat mensen zich meer natuurlijk gaan gedragen als de omgeving meer natuurlijk is."
2.1
"Dit voelt meer alsof je er daadwerkelijk bent."
1.2.
"Als je stil bent hoor je ook dat [redacted] en de ventriktie en stel je voor dat alles ineens stil staat dan is het heel onnatuurlijk."
2.3
"Beeld was zeker een versterking van het geluid. En sowieso zelf zonder het beeld was met die context geluidjes al meer dan degene ervoor."
3.2
"I was probably expecting something more real."
2.?
Is realism better?
"yeah I think so."
2.1
What did you think about it?
"yeah of course it felt more real because of the visuals"


Content
1.1
[redacted] In het echt zou je ook de convolutie hebben van je eigen kamer als je naar een trailer luistert.
"Dat is precies de binaurale ontkleuringseffect."
1.3
"I think it was less powerful in the sound than the first one (1.1)"
"It was more effective in the sense it gave me a movie theater feeling." (more surrounding)
1.1
"It (the TV) was like a [redacted] theater experience"
1.1-
"Dan denk ik inderdaad dat het beter is om de trailer niet te behandelen, terwijl dat juist contra-intuïtief is. Haar de stem, dat lukt nog een beetje."
"De trailer vond ik creepy omdat het VR is dacht ik steeds dat iemand op me af ging rennen of dat er iets creeps zou gebeuren zoals op youtube."
1.3


Sound quality
2.2
"When a track starts and it ends you still have the... like noise from the audio." (recording)
1.2
"Je hoort noise switching. Dus je hoort bepaalde geluiden en op een gegeven moment valt het helemaal stil en komt het weer op."

APPENDIX 3.A.

QUESTIONS USER TEST 2

Rate from 1 (strongly disagree) to 5 (strongly agree), and explain why.

You felt like you were with another person.

You interacted with another person.

Another person interacted with you.

You felt like you were physically in that environment

You felt relaxed in that environment

Was my voice clear?

Can you understand me?

Does it feel natural?

Do you find the sound to be localised?

TV sounds like headphones or cinema?

Was your voice clear?

Can you understand yourself?

Is it natural?

Thinking back to the experience, you remember it as images rather than as somewhere you have been.

It felt more like being in a virtual environment than being somewhere in person.

Open questions:
Have you had previous experience with virtual reality?

What would you want to improve in this experience (environment, content, communication)?

APPENDIX 3.B.

RAW QUALITATIVE DATA USER TEST 2

Each quote of the same colour is from the same participant.

TV

I really had the impression that I could hear it around me.

Felt like a cinema, but a little bit with headphones.

It sounds like you're watching in a cinema. The sounds are around you.

More like a cinema

When I talk to you, I hear the depth as if it comes from you, but with the cinema I don't have that depth, it feels like it comes from both ears, not as if the distance is right.

Proper theatrical experience I would say.

Pretty convincing, like a cinema. I think I'm biased because I was never as optimistic about VR so I'm surprised.

Reverberation

Why did you not like the white one? - It feels like i'm listening to a recording more than like a natural recording. So it's like someone talking in my ear directly.

This (white one) sounds more realistic, it suits the room more than the echo (grey one). It feels more personal.

The grey one is nice to listen to, but the white one is the most natural. I still think the white one is better, without reverb. I don't feel like I hear it around me.

Between the white and grey one, not completely gone, but a bit less than the grey one. The reverb is a bit too much, it's more than what I would expect.

The black one gives me more time to understand what you are saying. It helps with understandability, but it's less natural. Prefer understandability over it sounding natural.

White one is less accurate, feels more like a microphone. I prefer the grey one.

Grey looks a bit better, the white one felt like you were whispering in my ear. So it is not natural I'd say.

(White one) This sounds much more realistic. I think the original reverb sounds much more logical for a small room and for this type of house, the white one makes more sense. Because the reverb is gone now you sound much more from the sphere so the interaction improves from that.

There's a weird echo. Like we're in a cathedral. Maybe I'm not used to being in such a big room, but I have the impression that a similar house would have less. Somewhere between the white one and the grey one would be more what I'd expect.

Sidetone

I like this, more engaging. Sounds more like this voice goes directly in my ears.

It feels like I'm in the room, more than before.

It feels more like I'm in the room because I hear myself like you. But it doesn't completely feel like I would hear myself normally. Sounds within head.

It's digital. Doesn't feel very natural because I'm not used to hearing my voice. It sounds different because you always hear your voice differently than a recording. (Prefered without)

Better than it was. It seems less natural though. Not used to hearing my own voice and I think there is some latency. I understand myself better though so I prefer it.

First I used my voice more to hear myself at that volume and now I hear myself louder than that and there's a bit of reverb. It feels like I'm in a bigger room. Feels more like I'm being here. It's nice to hear my own voice placed in that environment.

When I can hear my voice it feels a bit better. This is better. When I go to a movie with a friend I have this kind of experience, I can hear my voice and you. This feels more natural than the previous one.

This is weird. Because your voice recorded sounds different than the sound you're used to. It's like hearing yourself in a recording. There's a bit too much reverb for the room, but it feels natural. It should be a bit less reverb based on my expectations. It's louder and clearer, but it has that reverb. I don't really have a preference.

When I hear my voice and yours with the same effect it makes it feel more like we are together in the same place.

Without sidetone

You hear yourself a bit muffled.

Pretty clear through the headphone, a bit muffled though. It's my own voice except less loud.

Kind of confusing. I feel like sometimes I say something and it's not really what I wanted to say. I hear myself a bit differently, I feel this is not my natural setting.

You have a bit of reverberation, I guess it would feel more real to have it on my voice too.

I can understand myself pretty clearly. A bit muffled.

I can hear myself, but it is muffled thourgh the headphones.

Sound mixing

I think it wasn't loud enough (cinema)
It's a little bit louder than in a normal conversation. If the volume was better it would be a 5. (side-tone)

I thought the sound would be louder because at home we also have one of those windows and the rain was a lot louder.

(Is it loud enough?) Could be increased slightly. (Better?). Yes.

Your voice is loud enough, the TV isn't.

Environment

I feel like this place is comfortable and I feel present in this place.

I find it light and with the grey colours it's relaxing, but maybe a plant somewhere in the corner or a candle.

It's pretty, it's a nice house. Cool that you can look everywhere, to the kitchen and the stairways.

The environment is awesome, through the windows you can see the trees move and it looks like it's raining. It was on such a high level that it looked very realistic.

The setting is a bit unfamiliar, maybe a more calm setting would be nice, less gloomy. More bright.

It's a place where you would sit at home. Good chair, that helps. As if you're home.

Realism

I think the environment felt real, but I was conscious it was virtual. Probably because I didn't know the place.

Still felt a little virtual, but it felt more real than I expected.
The fun part was that I recognized IKEA furniture so I thought it was funny. This makes it more real.

It's clearly animated, it looks very different in this real room than there. Especially the display quality. When I was there it definitely felt more real than now that I can see this real room.

It looked like a real environment, would be better with a better resolution of the headset.

It felt like somewhere I've been. To the point I tried to grab something or wanted to walk, but I knew there were cables.

The light could be softer. It's pretty but also not really how color would really be. (Color grading). Felt like I was really there.

Real with a pinch of virtual reality.

Feels like a real place. I can imagine if you have a small house that you put this on your head and suddenly you have a big house with a big TV, that's fun. You can still clearly see it's a display (SDE), but it doesn't break the immersion. The HMD could be improved. It's big and bulky, you feel the weight. For the sound, maybe take the headphone off so I can hear myself directly.

When I got back home I thought it was a real place I went to. Sometimes would be aware of the cables.

Avatar

The only thing is that you are a little sphere.

Sphere worked pretty well. Maybe an avatar would also work.

It's like a floating sphere so it's not the same feeling as someone else on the couch, but the rest feels exactly the same. I would make the avatar a bit more personal. The only thing I missed at interaction level is that normally you do a lot of body language.

(How to improve) You and me as a physical person probably.

Your avatar is a bit creepy, I see a floating sphere with a VR headset on.

APPENDIX 3.C.

RAW QUANTITATIVE DATA USER TEST 2

Test	Participant	VR experience	You felt like you were with another person	You interacted with another person	Another person interacted with you	You felt like you were physically in that environment	You felt relaxed in that environment	Was my voice clear?	Can you understand me?	Does it feel natural?
1	2749	none	5	3	3	5	5	5	5	5
2	2749		5	5	4	5	5	5	5	5
1	3350	novice	4	3	3	4	4	5	4	4
2	3350		5	4	4	4	4	4	5	4
1	3538	none	4	4	4	5	5	5	5	4
2	3538		4	4	4	4	5	5	5	4
1	4214	novice	4	3	3	4	5	5	5	4
2	4214		4	4	4	5	5	5	5	3
1	5119	novice	5	3,5	5	5	4	5	5	5
2	5119		5	4,5	2,5	5	5	5	5	3
1	5155	novice	4	5	5	4	5	3	3	4
2	5155		4	5	5	5	5	4	4	5
1	5519	novice	4,5	5	5	5	5	5	5	5
2	5519		5	5	5	5	5	5	5	5
1	10208	novice	3	2	5	5	5	5	5	2
2	10208		4	4	4	5	5	5	5	4
1	11111	novice	5	5	5	5	5	5	5	3
2	11111		5	5	5	5	5	5	5	4

Test	Participant	Do you find the sound to be localised?	TV sounds like headphones or cinema?	Was your voice clear? (didn't exist)	Can you understand yourself?	Is it natural?	Image/real? (1 image 5 real)	Virtual/real? (1 virtual 5 real)
1	2749	3	4	5	5	5		
2	2749	3	4	5	5	4	5	3
1	3350	4	5	5	4	3		
2	3350	5	5	5	4	3	5	4
1	3538	3	4	5	5	5		
2	3538	4	4	5	5	3	3	2
1	4214	4	4,5	4	4	4		
2	4214	5	4,5	4	5	3	3	3
1	5119	4	5	4	2,5	3		
2	5119	2	5	4,5	4	3	5	4
1	5155	5	3	5	5	5		
2	5155	5	3	5	5	5	4	3,5
1	5519	5	4,5	5	5	5		
2	5519	5	4,5	5	5	5	5	4
1	10208	2	5	4	4	5		
2	10208	5	5	4	5	4	5	5
1	11111	4	5	5	5	3		
2	11111	5	5	5	5	4	5	4

APPENDIX 3.D.

IDE Graduation Assignment (version 2017.09.21)
incl. the student's study progress (Appendix 3)

To be completed by the student
Please save your assignment as (format): IDE Graduation Assignment_family name, name_student number_dd-mm-yyyy
Place the proper document name on each page of your assignment in the headline, number the pages

Name student	Vincent Poppelaars	
Student number	4139275	
Address	599 Korvezeestraat	
Zip- code, City	2628CW Delft	
Telephone	0610698676	
E-mail address	Vincent.poppelaars@gmail.com	
Start at IDE 2011 (year)		Start at TU Delft 2011 (year)

Bachelor [1]	**Master** [1]	**Specialisation** [1]
☒ TUD Bachelor IO	☐ IPD	☐ Medisign
☐ TU/e or UT Bachelor IO	☒ DfI	
☐ TU Delft non-IO BSc	☐ SPD	**Annotation** [1]
☐ Other Dutch University Bachelor	☐ ……. = 2nd non-IDE master	☐ Techn. in Sustainable Design
☐ HBO Bachelor	☐ Individual programme, date of approval [2] …….	☐ Entrepreneurship
☐ Foreign Bachelor	☐ Master Honours Programme	

Name Chair	René van Egmond

1. Check study progress

To be completed by the Shared Service Centre O&S after approval of the assignment by the chair. The study progress will be checked for a 2nd time just before the green light meeting.

Bachelor degree:	☐ Yes	☐ No	☐ N.A.

Missing 1st year Master courses	1. …… 2. …… 3. ……	4. …… 5. …… 6. ……

Master electives, no. of EC credits accumulated:	

Name:	Date: ….. / ….. / 20…..	Signature:

2. Formal approval Graduation Assignment by the Board of Examiners

To be completed by the Board of Examiners

Approval of the content of the Grad. Assignment:	☐ Approved	☐ Not Approved
Procedural approval:	☐ Approved	☐ Not Approved

Comments:	

[1] *Tick where appropriate.*
[2] *Date of approval of your individual programme by the Board of Examiners.*

Name:	Date: ….. / ….. / 20…..	Signature:

IDE Graduation Assignment

GENERAL INFORMATION

Title Graduation Project [i]	**Evaluate, analyse and improve the auditory immersion in Virtual Reality**		
Chair of Supervisory Team [ii]	René van Egmond		
Department / Section	HICD		
Mentor of Supervisory Team [4]	Aadjan van der Helm		
Department / Section	DCC		
Project commissioned by [iii]	☐ Faculty	☒ Company	☐ Other, e.g. entrepreneurial
Project type [5]	☒ Design	☐ Research [iv]	☐ Other, e.g. entrepreneurial
Company name, if applicable	TNO		
City & Country	Den Hague, The Netherlands		
Company Mentor	Hans Stokking		
Start date	01/02/2018		
End date	05/07/2018		

CONTENT

Introduction

Give a sketch of the context of your assignment. Historical developments, if applicable relevant published scientific research results, new trends, status quo; materials, technologies, usage, etc.

- *In case of a faculty project: describe how your assignment reflects the research portfolio of the IDE Faculty [6].*
- *In case of a company project: provide company information.*
- *If other, e.g. entrepreneurial: describe the future enterprise and how your assignment will be of value to the enterprise.*

Include an illustration or visual which depicts the context of your assignment.
In case one or more extra parties are involved in your project, indicate which role they play.

Virtual Reality (VR) has quickly grown in recent years and reached consumers. It is currently used for professional applications such as training, simulations or healthcare while also finding its way to consumers through gaming and other audio-visual experiences. Through this significant progression, companies and content creators have mostly focused on the visual side of VR. The first thing experienced with VR is sight, as a large screen is placed in front of your eyes. Most solutions focus on that aspect and an external headphone is necessary. Even the more expensive headsets such as the HTC Vive only provide a basic headphone at additional costs. Most demonstrations also only focus on the visual aspect and often times the demonstrations are given in environments that do not benefit the sound experience. By limiting the sound experience, the general immersion takes a hit. Research shows that sound can positively influence other senses such as sight, it can also provide cues about the virtual environment such as size, space and localization. Evaluating and improving the auditory experience is therefore necessary to further increase the immersion of users in Virtual Reality.

Company information

TNO is an independent research organization that focuses on creating common experiences of social and economic value. It was founded in 1932 as a way to provide innovations to improve the competition between companies and to improve the welfare of society. TNO focuses on nine social themes: infrastructure & maritime, circular economy and the Environment, Defence, Safety and Security, energy, healthy living, industry, information & communication technology, strategic analysis & policy, traffic and transport.
Within TNO in The Hague, in the Networking department is a place called MediaLab where they are working on Virtual Reality experiences, with emphasis on what they call Social VR. This Social VR is a way of communicating with other people through VR by integrating them in the same environment and interacting together. The user is placed in a room represented by 360 image with additional interactive elements such as a screen. While TNO is currently working on improving the technical side and the visual aspect, they have yet to focus on the auditive side.

Problem definition

Indicate clearly, what should/could be improved compared to the present situation. When executing a research project: indicate the knowledge gap. What opportunities exist, what contradicting demands should be addressed, etc.

Due to VR hitting the mainstream, companies are improving the overall experience with each iteration. However, current demonstrations have a rather basic auditive experience and few spatial cues about the location of the sounds, or even in what type of environment they are. Users talking to each other is also kept basic, it does not sound like they are in the same environment. The challenge is to improve the user experience of Social VR through sound for a particular context (for example a stand up meeting or watching TV with friends).

Assignment

Briefly and to the point, describe what you are going to design, create or generate to solve (part of) the problem.
In case of a Specialisation and/or Annotation, address specifically how this is/these are included in the assignment.

The aim of this graduation project is to **evaluate, analyse and improve the auditory experience in Social Virtual Reality**. The focus will be on prototyping and testing solutions to make an immersive audio environment by providing spatial cues and also to integrate users in the environment. The experience being tailored to its context and user needs. The result would be an optimal demonstration of how sound should work in Social VR.

The main hypothesis for this project is that sound immersion improves the user experience in Social VR and gives the sense of being together in the same environment.

To support the decisions, iterations and guide the user tests, the leading research question throughout this project will be:

RQ: **What sound characteristics improve the user experience in Social VR?**

With following sub-questions:
S-RQ1: What setup would be appropriate for Social VR settings?
S-RQ2: What aspects from real life are important for sound immersion?
S-RQ3: Would a realistic immersion be too distracting?

Approach

What will be the approach to deal with the complexity of the assignment? What has to be done to meet the challenges? Indicate the main methodologies to be used. Indicate the same project phases as you distinguish in your planning. If one or more extra parties are involved in your project, indicate which role they play.
In case of a Specialisation and/or Annotation, address specifically how this is/these are dealt with.

Initially, a **literature research** will be done to get familiar with the subject of human hearing, 3D audio and the advancements in immersive audio in Virtual Reality. It will also be coupled with a field search to compare current hardware solutions and potentially interviews with experts. Several **scenarios** will be created based on the different applications possible in Social VR. With a **questionnaire**, the user requirements and interaction qualities can be determined for the chosen scenario/context. After which an **interaction vision** can be established for the type of experience that should be created.
Then, based on the theory and vision, **a user test will be designed** to compare different settings/concepts and their influence on the interaction qualities and sense of immersion. Comparing different setups is important to choose the hardware for the final concept. It would also indicate what is more appropriate for a different environment.
The final step would be to **design a demonstration** using available software in which users can interact with each other and feel like they are in the same environment. It will be user tested to grade the user experience and find final points of improvement to the experience.

In between these demonstrations, small tests can be done with colleagues to make minor iterations and find out what works and what does not.

Since the final step relies on the use of complex software, TNO can help on that aspect.

Graduation Project results

1. Describe the expected results or outcome of your Graduation Project. For instance, a product, a product-service combination, a strategy illustrated through product or product-service combination ideas.
2. Indicate the expected scientific and/or societal and/or commercial significance of the outcome of your project.
3. In case of a Specialisation and/or Annotation, address specifically the relevant results to be expected.

1. The expected result is a set of guidelines and theory about how to improve the user experience in Social VR through sound together with an experience demonstrating that theory.
2. The result would be of commercial significance by providing an all-round immersive VR experience. The social value is clear as the demonstrations will be designed for Social VR, so that people can interact with each other more naturally through VR.

Deliverables

List the extra graduation deliverables, if any (apart from the mandatory deliverables being the thesis report, annexes if any, the poster and the representative pictures). For instance, a working prototype or a paper.

The graduation project will result in deliverables best suited to communicate the findings and results of the project:

- Thesis, reporting on the research and design process during the project including key choices and guidelines on what elements are necessary for auditive immersion in VR.
- Executive summary for TNO
- Demonstration of immersive audio in Social VR
- Presentation
- Poster

Relation and relevance to the domain of Industrial Design Engineering, the chosen master direction and the IDE pillars

Explain the relation of your project with the domain of Industrial Design Engineering and your master direction IPD, DfI or SPD.
1. Relation of you project to the master IPD, DfI or SPD
Furthermore describe the interface of your project with each of the IDE pillars:
2. Business
3. Human Interaction
4. Technology

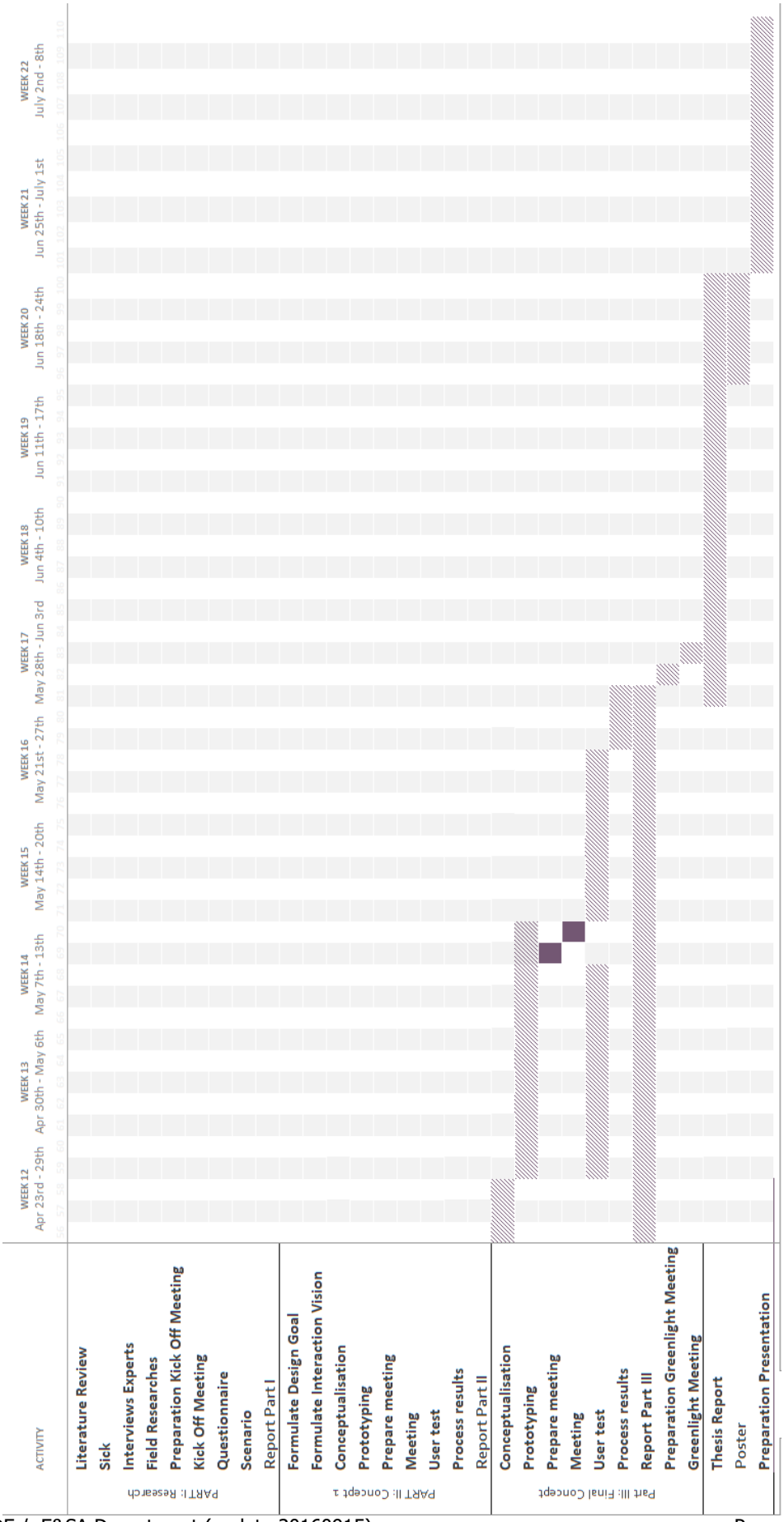
1. The nature of this project is research oriented yet the design, or audio design, will go through different iterations with the research and user tests to improve the user experience. The interaction would be between the user and the audio and between the two users in Social VR. The audio will react on the user's actions in the virtual world and represent them appropriately.
2. The **business** potential of this project is that the research and designed can be applied for B2B such as stand up meetings while also show the potential of audio to the VR industry.
3. In this project, the **users** are central to the solution. Through explorations and testing, the user requirements of the solution will be clearer. An interaction vision of how the interaction between the users should work in Social VR will also help create an experience that suits the users. Both the user requirements and interaction vision will shape the interface (audio) which facilitates the interaction between the users and their environment.
4. **Technology** is the core of the interface. Users interact with each other in Virtual Reality, exclusively through technology with an HMD (head-mounted display), controllers and sound.

Planning

Present your planning in a Gantt Chart, which can easily be made in Excel, see example underneath. Make sure a print in black and white is still readable. Mention the main phases of the project as described at Approach + number of weeks. Indicate only main activities, milestones, meetings. Take notice: 33 EC = 22 full-time weeks! Indicate periods of part-time graduation project activity and/or periods of not spending time on your graduation project, if any, for instance because of holidays [3].

[3] *Only by approval of the Board of Examiners , a not yet passed course may be combined with the Graduation Project. In such case, show the approval to your Chair and indicate the period of not spending time on your Graduation Project for this reason.*

Graduatio
Today (day #): 26
Plan
Actual
% Complete
Actual (beyond plan)
% Complete (beyond plan)
ACTIVITY
WEEK 1 Feb 1st-11th
WEEK 2 Feb 12th - 18th
WEEK 3 Feb 19th - 25th
WEEK 4 Feb 26th - Mar 4th
WEEK 5 Mar 5th - 11th
WEEK 6 Mar 12th - 18th
WEEK 7 Mar 19th - 25th
WEEK 8 Mar 26th - Apr 1st
WEEK 9 Apr 2nd - 8th
WEEK 10 Apr 9th - 15th
WEEK 11 Apri 16th - 22nd
PART I: Research
Literature Review
Sick
Interviews Experts
Field Researches
Preparation Kick Off Meeting
Kick Off Meeting
Questionnaire
Scenario
Report Part I
PART II: Concept 1
Formulate Design Goal
Formulate Interaction Vision
Conceptualisation
Prototyping
Prepare meeting
Meeting
User test
Process results
Report Part II
Part III: Final Concept
Conceptualisation
Prototyping
Prepare meeting
Meeting
User test
Process results
Report Part III
Preparation Greenlight Meeting
Greenlight Meeting
Thesis Report
Poster
Preparation Presentation


ACTIVITY
WEEK 12 Apr 23rd - 29th
WEEK 13 Apr 30th - May 6th
WEEK 14 May 7th - 13th
WEEK 15 May 14th - 20th
WEEK 16 May 21st - 27th
WEEK 17 May 28th - Jun 3rd
WEEK 18 Jun 4th - 10th
WEEK 19 Jun 11th - 17th
WEEK 20 Jun 18th - 24th
WEEK 21 Jun 25th - July 1st
WEEK 22 July 2nd - 8th
PART I: Research
Literature Review
Sick
Interviews Experts
Field Researches
Preparation Kick Off Meeting
Kick Off Meeting
Questionnaire
Scenario
Report Part I
PART II: Concept 1
Formulate Design Goal
Formulate Interaction Vision
Conceptualisation
Prototyping
Prepare meeting
Meeting
User test
Process results
Report Part II
Part III: Final Concept
Conceptualisation
Prototyping
Prepare meeting
Meeting
User test
Process results
Report Part III
Preparation Greenlight Meeting
Greenlight Meeting
Thesis Report
Poster
Preparation Presentation

APPROVAL BY CHAIR

Date of approval	
Signature of Chair	

[i] *Keep the title compact and simple. Do not use abbreviations.*
[ii] *Avoid team members from the same section. In case a non-IDE mentor is preferred over an IDE-mentor, the Chair should request so for approval by the Board of Examiners (including a motivation letter and c.v. of the proposed non-IDE mentor).*
[iii] *Tick where appropriate. See the IDE Graduation Manual, paragraph 2.5. If necessary, explain at Introduction.*
[iv] *See webpage http://www.io.tudelft.nl/en/research/*