

Virtual Heritage

Audio Design for Immersive Virtual Environments
Using Researched Spatializers.

Veli Laamanen

Advisors: Lily Diaz-Kommonen & Marko Tandefelt
Supervisor: Antti Ikonen

Master's thesis
Master's Programme in New Media: Sound in New Media
MediaLab, Department of Media
Aalto University
2018

Author Veli Laamanen

Title of thesis Virtual Heritage - Audio Design for Immersive Virtual Environments Using Researched Spatializers.

Department Media Lab, Department of Media

Degree programme Master's Degree Programme in Sound in New Media

Year 2018

Number of pages 91

Language English

This thesis work is based on a Virtual Heritage project being developed by the Systems of Representation research group. The objective of the project is to create a showcase demonstration on how the virtual reality (VR) could be used as an application for tourism in the heritage sector. In this context, my task was to develop a concept and prototype of how 'spatialized' sound could be used in a VR application. The initial location chosen for the concept was the ancient heritage burial site of Sammallahdenmäki, one of the Finnish heritage sites listed in the UNESCO register of World Heritage Sites.

The thesis, that is written from an audio designer's perspective, focuses on three aspects of this project. First is the sound design for the Virtual Heritage project and the second is the quality of currently available 'spatializer' plug-ins used for headphone listening. In order to evaluate the process of designing 3D audio for virtual environments, the methods and principles within binaural rendering, sound design and immersion must be understood. Therefore, functions and theories within audio spatialization and 3D audio design are reviewed.

Audio designers working on virtual reality content need the best possible solutions for creating believable 3D audio experiences. However, while working on the Virtual Heritage project, we did not find any comparative studies made about commercially available spatializer plug-ins for Unity. Thus, it was unknown what plug-in would have been the best possible solution for 3D audio spatialization. Consequently, two tests were conducted during this thesis work. First was an online test measuring which spatializer would be the most highly rated, in terms of perceived directional precision when utilizing head-related transfer functions without reverb or room simulations. The second was a comparative test studying if a spatialized audio rendering would increase immersion compared to non-spatialized audio rendering, when tested with the Virtual Heritage demonstration.

The central aim in the showcase demonstration was to create an immersive virtual environment where users would feel as if they were travelling from the present, back to the Bronze Age, in order to understand and learn about the location's unique history via auditory storytelling. The project was implemented utilising the Unity game engine. The research on music and other sound content used in the project's sonic environment is explained. Finally, results of the project work are discussed.

Keywords Sound Design, Virtual Reality, HRTF, Spatializer, Cultural Heritage

Contents

Acknowledgements	vii
Abbreviations.....	viii
1. Introduction.....	1
1.1 Motivation.....	1
1.2 Structure.....	3
1.3 Assignment	4
1.4 Audio Model	5
2. Background.....	7
2.1 Immersion	7
2.2 Role of Audio in VR.....	10
2.2.1 Informative Audio.....	10
2.2.2 Emotional Audio.....	11
2.2.3 Immersive Audio	12
2.3 Binaural Sound.....	13
2.3.1 Mono and Stereo	13
2.3.2 Natural Spatial Hearing.....	14
2.3.3 Evolution of 3D Audio.....	16
2.4 3D Audio	17
2.4.1 Basics	17
2.4.2 ITD, IID and SD.....	18
2.4.3 HRTF	21
2.4.4 Distance and Reverberation.....	22
2.4.5 Motion.....	23
2.4.6 Issues in 3D Audio.....	24
3. Evaluation of the Spatializers	26
3.1 Spatializers for Unity.....	26
3.2 Spatializer Test.....	29
3.2.1 Designing the Online Test.....	29
3.2.2 Data Acquisition and Distribution	33
3.2.3 Spatializer Test Results	37
3.3 Immersion Test.....	39
3.3.1 Test Procedure	39
3.3.2 Immersion Test Results	41

4. Case study: Virtual Heritage	43
4.1 Prototype.....	43
4.1.1 Location	43
4.1.2 Concept	44
4.1.3 Unity	45
4.1.4 Visuals	45
4.1.5 Audio Zoom	46
4.2 Music and Audio Content.....	48
4.2.1 Music and Instrumentation	48
4.2.3 Music Production	53
4.2.4 Other Audio Content	56
4.3 Audio Implementation.....	57
4.4 Final Discussion	62
5. Conclusion	65
Appendix	67
A. Audio Production of the Online Test.....	67
B. Instructions for the Online test.....	73
C. Histograms for the Online Test	74
D. Audio Components	80
List of Figures	82
List of Tables.....	83
Bibliography.....	84

Acknowledgements

Firstly I would like to express my deepest thanks to my thesis advisor Professor Lily Diaz-Kommonen for trusting me to work on the Virtual Heritage project, for guiding me through unknown waters, and for being incredibly encouraging and supporting in the hardest times. I would also like to show my deepest appreciation to thesis advisor Marko Tandefelt for all the valuable insight, feedback and all that positive energy. I am grateful to supervisor Antti Ikonen for all the possibilities that I got to experience during my studies in Media Lab. It has been a life changing experience.

Furthermore, I would like to express my thanks to all the others that have helped me to get through this process:

Associate Professor Perttu Hämäläinen for advising me in the online research.

Thomas Holmes for helping me remedy my grammar and for being my opponent.

Markku Reunanen for the advice at the master's thesis seminar.

Professor Ville Pulkki and Micah Hakala for introducing the DirAC project at ELEC.

Jukka Eerikäinen for helping me with graphical optimisation of the Virtual Heritage prototype.

Juhani Ala-Hannula from Gramex for advising with the creative licenses.

Anni Guttorm ja Kataja Raila for the information about ancient Sami drums.

Petteri Mäkinieniemi, Sebastien Piquemal for their advice during the thesis group discussions.

Tuomas Ahva for advising with the immersion test.

Sampo Laine, Samuli Riik, Johanna Rautiainen, Martti Sirkkola, Can Uzer, Helena Haaparanta for being my perfect Viking choir.

And of course media lab staff and my dear colleagues Gabi, Kim, Juha, Ville, Sami, Ava, Eero, Mikko, Ilpo, Jukka, etc... with whom I had the most fantastic educational adventures.

Abbreviations

3D	Three Dimensional
B.C.	Before Christ
dB	Decibel
ELEC	School of Electrical Engineering
HMD	Head-mounted Display
hpTF	Headphone Related Transfer Function
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
IID	Interaural Intensity Difference
ILD	Interaural Level Difference
INC.	Incorporation
ITD	Interaural Time Difference
LUFS	Loudness Units Relative To Full Scale
LTD.	Limited Company
R/D	Reverberant-to-direct Sound Ratio
RMS	Root Mean Square
SD	Spectral Differences
VE	Virtual Environment
VR	Virtual Reality

Chapter 1

Introduction

This chapter explains the motivation, context and structure of this thesis. Furthermore, the audio model that emerged from the Virtual Heritage concept is introduced.

1.1 Motivation

This thesis work is based on the Virtual Heritage project being developed by the Systems of Representation research group. It is part of a research project funded by the Aalto Digi Platform. I started working as a research assistant at Aalto University for Professor of New Media Lily Diaz-Kommonen in October 2017. One of the objectives of the Virtual Heritage project is to create a showcase demonstration on how Virtual Heritage and virtual reality (VR) can be used as a tourism application for the heritage sector.

I have worked as a composer and audio designer for multiple media production companies and projects in Finland and I have a personal motivation to build a career in the fields of games and VR, which require 3D audio design. Thus, the project work and the thesis emerged from the desire to apply my knowledge of sound design to the field of virtual reality. Additionally, I hope that other designers can benefit from this paper in their search for the tools, means and attitude in creating meaningful audio experiences in VR.

In the process of creating this thesis I researched, discovered and learned about topics that were new to me, but which were essential to understand in order to complete the project. As such the work has made use of some scientific methods but it is written from a sound artist's perspective. To conclude, this thesis is first and foremost a documentation and representation of my personal approach to the process of design as a sound artist.

This thesis is an artistic research project which utilises the practice-led method, where research is realised through professional project work.¹ Therefore, the questions emerged from the project work. There were several questions I asked myself as I engaged in the work, one of the more important ones was: ‘How can user immersion in the virtual environment be improved utilising only audio?’ My initial assumption was that immersion would increase if the audio was ‘realistically’ spatialized in the virtual environment.

Therefore, the mission to achieve immersion through audio spatialization led me to a secondary question: ‘In terms of perceived directional precision, what is the most highly rated audio spatializer plug-in, and how much can its usage increase immersion when compared to non-spatialized audio?’ This demanded further study of immersion and the technicalities of 3D audio and initiated a series of experiments designed to answer this question.

The third essential question that emerged during the project work was: ‘How can an authentic sonic representation of Bronze Age era Sammallahdenmäki be created?’ Unfortunately, I had only a little knowledge of Sammallahdenmäki or the Bronze Age, and so in order to create an authentic and believable virtual environment, further research on this subject matter was required.

The content in this thesis is focussed only on the topics related to the central audio scene (also referred to as the audio experience) of the Virtual Heritage prototype. Therefore, the prototype is not discussed in its entirety and the project development is explained only when it affects artistic decisions related to the central audio scene.

I have tried to explain everything as clearly as possible in order to avoid an overly academic and non-practical approach. The discussions about 3D audio refer to binaural rendering, utilising head-related transfer functions in headphone listened conditions. All the studies in this paper apply to headphone playback and auditory effects of external speakers are not researched. Basic concepts of audio production and basic audio terminology are not explained in the thesis, as these are beyond the scope of the research.

¹ Lily Díaz-Kommonen, *Art, Fact and Artifact Production: Design Research and Multidisciplinary Collaboration*. (Jyväskylä: University of Art and Design, 2002), 24, accessed March 30, 2018, http://sysrep.aalto.fi/mulli2015/thesis_files/pdf/ch2_art_act_discourse_Diaz.pdf.

1.2 Structure

This thesis is structured to approximately follow the progression of my (learning) process in the project work of Virtual Heritage. Although some processes were, in practice, overlapping there was a logical progression from the start of the given assignment. The assignment and the audio model that emerged from the designed concept are explained in this **Introduction** chapter.

Chapter 2: **Background** discusses the theoretical studies behind the project-led work. From a sound designer's perspective it was essential to consider the capabilities of audio design in the context of VR and to define a central feature of virtual environments: immersion. The theories behind 3D audio were studied in order to gain a better understanding of the key components of audio spatialization.

At the beginning of development it was clear that the default 3D audio solution in Unity did not provide satisfactory results for the project work. Consequently, there was a need for alternative solutions which are introduced in the beginning of Chapter 3: **Evaluation of the Spatializers**. In order to distinguish the best 3D audio spatializer, an empirical online test was conducted. Furthermore, the effect of the 3D audio spatializer on immersion was tested. Methods, procedures and results of these experiments are reviewed in Chapter 3.

After the theoretical and empirical research on 3D audio the focus shifted to audio content research and production for the audio scene of the Virtual Heritage prototype. The procedure of discovering the likely elements of sound and music in Sammallahdenmäki and the Nordic Bronze Age is discussed, and the transformation from research to actual audio content is presented in Chapter 4: **Case Study: Virtual Heritage**. In addition, audio implementation of the final audio scene is explained. Finally, the findings of this learning process are discussed.

At the end, Chapter 5: The **Conclusion** summarises the findings of this project and thesis.

1.3 Assignment

The assignment of the project was to design a virtual reality concept with a strong focus on the potential of audio, and to develop a prototype according to that concept. The objective for the prototype was to demonstrate how virtual reality could be used as a tourism application for the heritage sector. In this context, my task was to design the concept and use the Unity game engine to develop the whole Unity project with scripts, audio and visual appearance under the direction of Professor Lily Diaz-Kommonen.

The location chosen by Professor Díaz-Kommonen for this concept was an ancient heritage burial site in Sammallahdenmäki, close to Rauma. It is one of the Finnish heritage sites listed in the UNESCO register of World Heritage Sites. The vision was to design a concept that would work as an audio model (introduced in next section) for other tourism locations as well. As the outcome was intended to be sensitive to the historical context of the site, a lot of time and attention was focused on research and creation of an audio experience that would feel authentic.

The time period for producing the concept and prototype was limited to six months. From my previous experience with game projects, I knew that the scoping was essential to successfully complete a project of this nature. Therefore, the focus was on creating only one working audio scene as a proof of concept.



Figure 1: Sammallahdenmäki burial cairn on an October morning in 2017.

1.4 Audio Model

The audio scene in the Virtual Heritage prototype follows an audio model (Fig. 2) that emerged from the project's concept creation. The audio model is divided into three parts: audio technology, audio content and audio interaction, presenting the central attributes of the Virtual Heritage concept. The content of this thesis will discuss all audio model attributes, but as previously mentioned in the motivation section 1.1, the main focus is on audio spatialization and designing the cultural heritage related audio content. The interaction segment will have only a minor role in this thesis as it was a very simple procedure and needed no further re-research in order to be implemented. The Virtual Heritage concept, prototype and prototype's audio model attributes are described in Chapter 4.

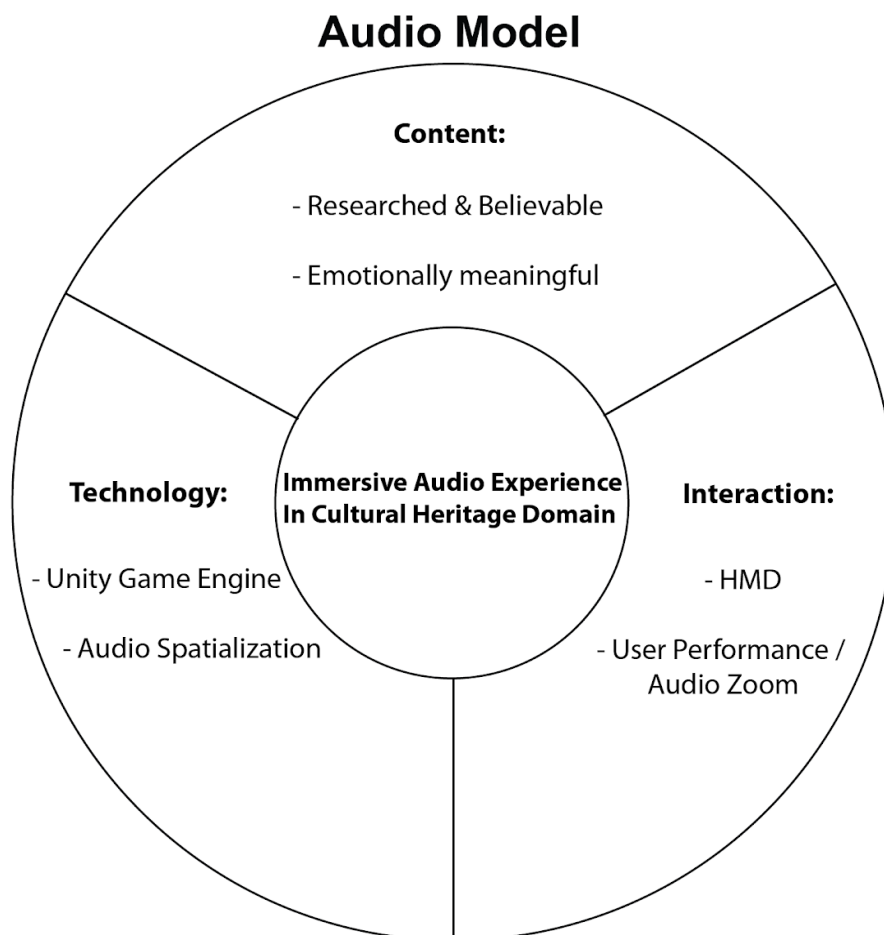


Figure 2: Audio model.

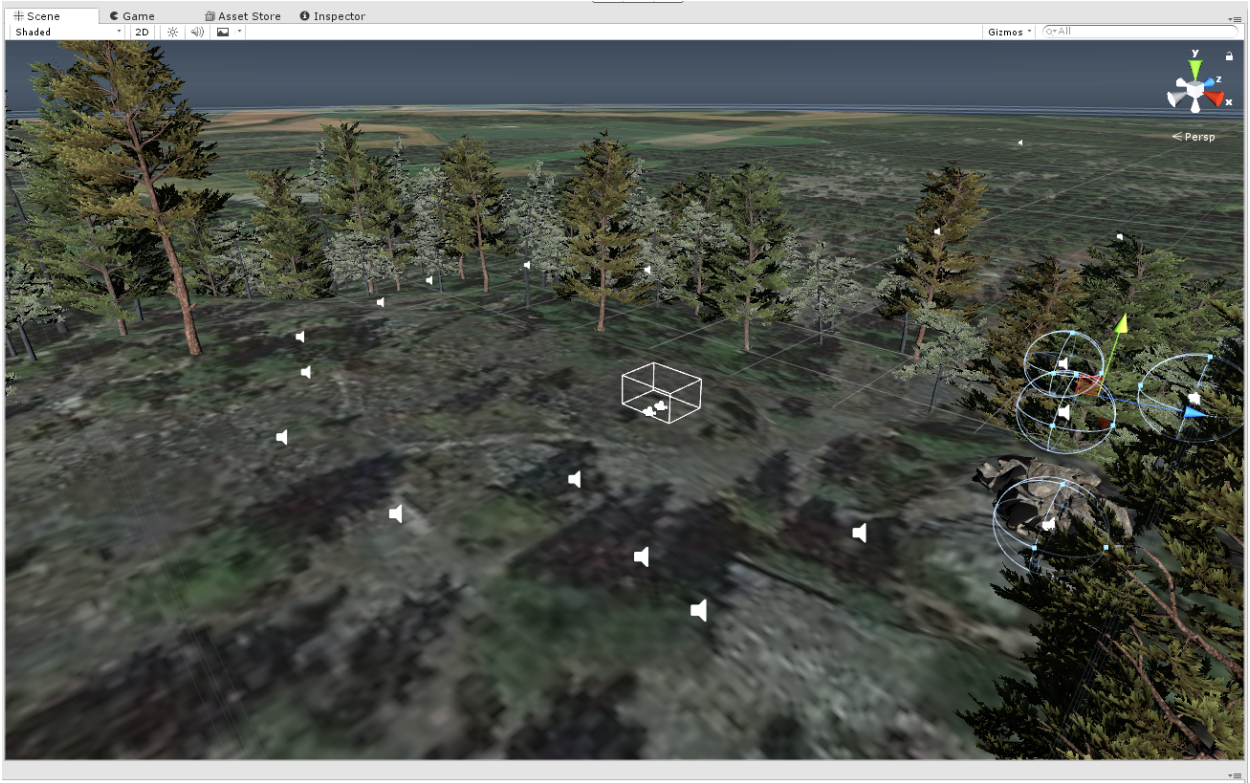


Figure 3: Prototype's visual appearance with audio sources locations.

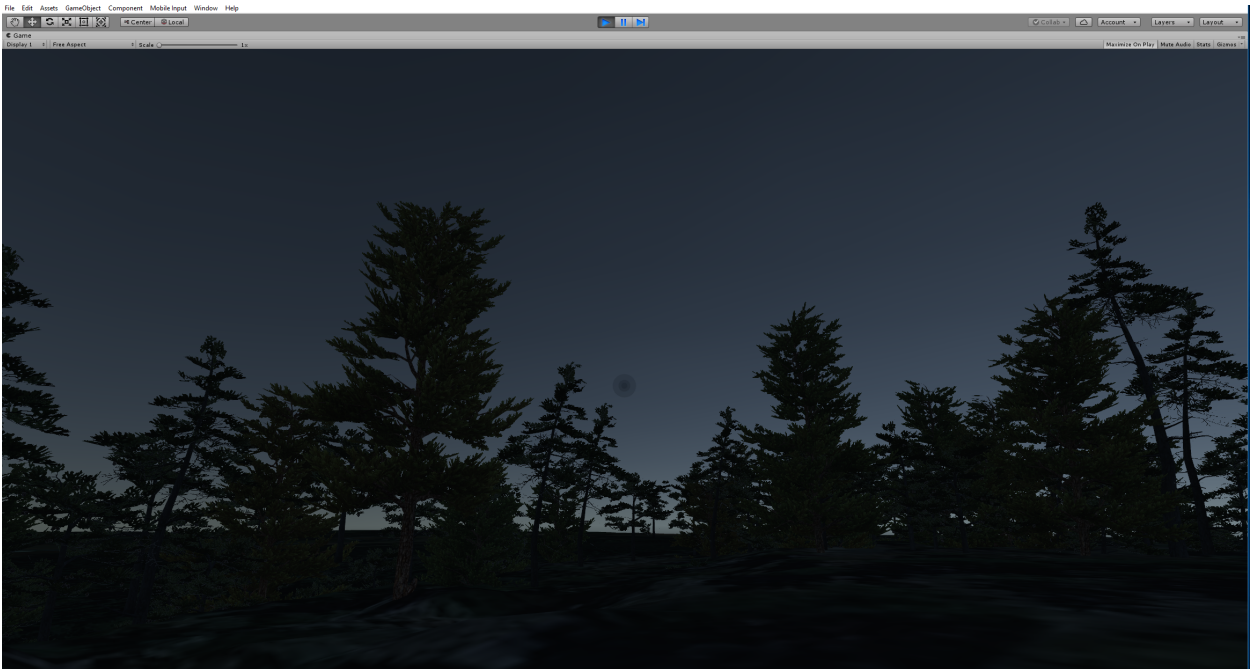


Figure 4: First person view of the final audio scene.

Chapter 2

Background

Chapter 2 discusses the theoretical studies behind the Virtual Heritage project work. It is necessary to first define a central feature of virtual environments: immersion. Secondly, it is essential to acknowledge the capabilities of audio in the context of VR. Furthermore, the evolution and theories behind 3D audio are studied in order to gain a better understanding of the spatial audio tools.

2.1 Immersion

In my own experience, the usage of the word 'immersion' is somewhat vague and without a clear definition, at least within the games industry. Therefore, it is necessary to clarify the meaning of immersion in this thesis as it is a central feature of the Virtual Heritage project and is used extensively in the text.

There are different interpretations of the term 'immersion', depending on the context. It is also often confused with the word 'presence'. So, let's begin by defining the word presence. According to the dictionary, the meaning of presence is a state of existing or being present.² However, presence doesn't necessarily mean that it needs to be associated with a physical body. For example, if away from a loved one, it is possible to feel closer to them when looking at a photo which is strongly connected to the absent person. This means that it is possible to bring the non-physical presence to us.³

Presence in the context of new media technologies has acquired a new relevance, derived from the term 'telepresence'. Telepresence refers to being present by means of verbal, gestural or graphical language without a physical form.⁴ In practice, a subject can discard the

² "Presence | Definition of Presence in English by Oxford Dictionaries." Oxford Dictionaries, accessed January 21, 2018, <https://en.oxforddictionaries.com/definition/presence>.

³ Schlemmer, Eliane, and Luciana Backes, "Immersion, Telepresence, and Digital Virtual Presence in Metaverses." In *Learning in Metaverses: Co-Existing in Real Virtuality* (IGI Global, 2015), 102, accessed January 03, 2018, doi:10.4018/978-1-4666-6351-0.ch005.

⁴ Eliane, and Backes, "Immersion," 107.

physical body and transform it into a form of a technologicised body.⁵ The subject can feel immersed in an environment when visualised in the first-person. An avatar works as a representative of a subject in the other world, visualising the events and interactions from an internal point of view and forming a sense of immersion.⁶ Immersion results from the involvement that a subject experiences.⁷ Total immersion can be said to be presence, and that by means of immersion, we feel like we are being inserted or submerged into another universe.⁸ Perhaps the most accepted definition for the word immersion is Janet Murray's:

The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. We refer to this experience as immersion. Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air.⁹

According to Brown and Cairns, the first level of immersion is to engage with the game. Immersion increases as the user invests time, effort and attention towards the game. Thus, the level of immersion is connected to the subject's involvement within the simulated environment.¹⁰

The user's sense of presence in virtual environments results in their engagement. This perceptual illusion involves responses within the cognitive and emotional human sensory processing to simulated objects and environments.¹¹ Thus, it can be said that the sense of

⁵ Eliane, and Backes, "Immersion," 103.

⁶ Ibid., 109.

⁷ Paul Cairns, Anna Cox, and A. Imran Nordin, "Immersion in Digital Games: Review of Gaming Experience Research." (New Jersey: John Wiley & Sons, 2000), 3, accessed January 20, 2018, <https://pdfs.semanticscholar.org/a9fb/2393b1e34be4155080bd02c6efba9e2cb807.pdf>.

⁸ Emily Brown, and Paul Cairns, "A Grounded Investigation of Game Immersion." In *proceedings of Human Factors in Computing Systems - CHI '04*, 2004, in Vienna, Austria (New York, NY: ACM, 2004), 1298, <https://doi.org/10.1145/985921.986048>; Eliane, and Backes, "Immersion," 114.

⁹ Alison McMahan, "Immersion, Engagement, and Presence, A method for Analyzing 3-D Video Games," in *The Video Game Theory Reader*, ed. Mark J.P. Wolf and Bernard Perron (New York: Routledge, 2009), 68, accessed January 15, 2018, <http://www.alisonmcmahan.com/sites/default/files/pdfs/papers/ch3-McMahanrev.pdf>.

¹⁰ Cairns, Cox, and Nordin, "Immersion in Digital Games," 3.

¹¹ Priscilla Chueng, "Designing Sound Canvas: The Role of Expectation and Discrimination." In *proceedings of Human Factors in Computing Systems - CHI '02*, 2002, in Minneapolis, MN. (New York, NY: ACM, 2004), 848, <http://dx.doi.org/10.1145/506443.506629>.

presence is always individual, depending on user's personal preferences, senses and cognitive capabilities.

Mel Slater clears the confusion between presence and immersion by mentioning that immersion is something that can be objectively assessed, whereas people may experience different levels of presence depending on subjective preferences. Slater compares and explains the difference: 'If immersion is analogous to wavelength distribution in the description of colour then presence is analogous to the perception of colour. Presence is a human reaction to immersion'.¹² In other words, the feeling of 'being there' is referred to as presence, and immersion is a stimulative and measurable part of the presence that may include spatialized audio, stereo displays and so on.¹³ From an auditory perspective, immersion is a complex set of interactions between the sounds and various aspects of the listener's sentiment processing.¹⁴ There is also a study which shows a clear connection between spatial presence and interactivity through a virtual avatar.¹⁵ This implies that the possibility of executing realistic interactions within the simulation maximises the immersion.¹⁶

It can be said that in an effective setup, immersion is a result of a user's attention, engagement, excitement and willingness. A measurable sense of presence is the result of believable sensory signals working together with coherent modalities. Simply put, immersion is a positive feeling of "being there".

¹² Mel Slater, "A Note on Presence Terminology", *Presence Connect* (January 2003), 1, accessed January 6, 2018,

https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology.

¹³ François De Sorbier De Pognadoresse et al. "From Research on the virtual reality installation," *Human-Computer Systems Interaction Advances in Intelligent and Soft Computing*, 2009, 337, accessed January 9, 2018, <https://hal-upec-upem.archives-ouvertes.fr/hal-00681737/document>.

¹⁴ Agnieszka Roginska and Paul Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (New York: Routledge, 2018), chap. 1, doc. 5, https://www.safaribooksonline.com/library/view/Immersive+Sound/9781317480105/xhtml/Ch01.xhtml#_15031-1076_h1_ch001_005.

¹⁵ Holger Regenbrecht, and Thomas Schubert, "Real and Illusory Interactions Enhance Presence in Virtual Environments." In *Presence: Teleoperators and Virtual Environments* 11, no. 4 (08 2002), 433, accessed February 28, 2018, doi:10.1162/105474602760204318.

¹⁶ Jiwon Lee, Kim Mingyu, and Kim Jinmo, "A Study on Immersion and VR Sickness in Walking Interaction for Immersive Virtual Reality Applications." In *Symmetry* 9, no. 5: 78. (2017), 1, accessed March 26, 2018, <https://doi.org/10.3390/sym9050078>.

2.2 Role of Audio in VR

2.2.1 Informative Audio

According to Nazemi and Gromala: “in the design of virtual environments, priority is often given to visual cues; auditory cues often are secondary in terms of importance.”¹⁷ This may be because the audio is more difficult to comprehend than visuals. Jens Blauert explained in his book *Spatial Hearing* that: “Human beings are primarily visually oriented. Compared to our visual sense, our other senses (auditory, tactile, etc.) are much less highly developed. Correspondingly, concepts and descriptions are based primarily on visual objects.”¹⁸ However, senses are connected to each other by knowledge, and as a result, spatial audio cues may be influenced by other senses, especially vision.¹⁹ Therefore, what is seen creates expectations of what should be heard.

In entertainment media, such as games and films, sound serves a certain purpose in storytelling and is a part of a larger whole. In a VR cinema workshop during the summer of 2015, I discovered that in virtual reality audio has to function as an informational guide to lead a participant’s attention. This was an eye-opening experience for me as I noticed that within virtual reality storytelling the role of sound may be of even more critical important than it is in two dimensional storytelling. Sound has to work as a directorial cue to lead a participant to look towards the narratively meaningful visual cues. In my opinion, the approach to sound design for VR cannot only be artistic but it also has to be highly functional.

VR systems combined with head-mounted displays (hereinafter HMD) have features that flat screen TVs or monitors have not been capable of. With HMDs, it is possible to expand the presentation from flat screen monitors to small display optics for both eyes. The simple flat screens presenting two dimensional information have made it easy to forget how we use hearing in our natural environment. We are able to hear spatial audio cues all around us very accurately and these cues contain information from beyond the 220 degree horizontal eyesight

¹⁷ Mark Nazemi, and Diane Gromala, "Sound Design: a procedural communication model for VE." In *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound - AM '12*, 2012, in Corfu, Greece. (New York, NY: ACM, 2012), 16, <http://dx.doi.org/10.1145/2371456.2371459>.

¹⁸ Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (London: MIT Press, 1983), 2.

¹⁹ Francis Rumsey, *Spatial Audio* (Oxford: CRC Press, 2017), 35.

of human visual ability.²⁰ In other words, our auditory system can provide information that HMD's cannot. This information may not be necessary when we are using flat screens, but in virtual reality where the subject is immersed in an artificially created space, spatial audio is a valuable addition and a fundamental functional element supporting the three dimensional experience.

Alex Riviere wrote in Gamasutra: 'in VR, hearing is the only sense able to provide full spatial information going beyond our field of view, including elevation, 360 degrees and depth, allowing us to guide our decisions and behaviours as well as understanding our virtual surroundings.'²¹ So when the VR production team is asking the question 'in which direction should the user look?' the audio system could provide a solution to that challenge. Therefore, in virtual environments, audio has a crucial *informative* role.

2.2.2 Emotional Audio

Sounds work on an unconscious level and evokes emotions in a very powerful way. It is well known that particular sounds are capable of triggering emotional responses.²² For example, music has an effect on infants as young as four months old.²³ Toddlers hearing music start swaying to the rhythm of the music without having any previous dancing experience. It is a common predisposition but one which is still not fully understood.²⁴

Studies show that it is not just music that affects people in this way, but that sounds and voices also evoke our emotions in a similar manner. When thinking about a four month old infant hearing the voice of their mother or the sound of a menacing dog snarling, it is easy to imagine the different reactions. There is a subliminal connection to these kinds of sounds that has

²⁰ "Field of View." Wikipedia, last modified January 02, 2018, https://en.wikipedia.org/wiki/Field_of_view#Humans_and_animals.

²¹ Alex Riviere, "How to Build Audio for VR Games." Gamasutra Article, accessed January 6, 2018, https://www.gamasutra.com/blogs/AlexRiviere/20171226/312290/How_to_Build_Audio_For_VR_Games.php.

²² Konstantinos Drossos, Andreas Floros, and Nikolaos-Grigorios Kanellopoulos, "Affective acoustic ecology: towards emotionally enhanced sound events." In *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound (AM '12, 2012)* in Corfu, Greece. (New York, NY: ACM, 2012), 109, <http://dx.doi.org/10.1145/2371456.2371474>.

²³ Catherine Pelachaud, *Emotion-Oriented Systems* (London: ISTE, 2012), chap. 9, doc. 1, https://www.safaribooksonline.com/library/view/emotion-oriented-systems/9781118603321/Text/17_chapter9.xhtml

²⁴ Staff, Live Science. "Babies Are Born to Dance." LiveScience, last modified March 15, 2010, <https://www.livescience.com/6228-babies-born-dance.html>.

developed during human evolution.²⁵ Some studies even suggest that sound evokes more emotions than any other content in audiovisual entertainment.²⁶

Research shows that there is a connection between the characteristics of sound and emotional responses. For example, rapid attack and release of audio signals (gunshots, alarms) are related to perception of urgency, and parameters of tone colouration and loudness can be used to either amplify or attenuate this kind of emotions. Furthermore, psychoacoustic properties can be associated with negative emotional experiences.²⁷

Complex sound design and dialogue that takes into account psychoacoustic phenomena generates an emotional response and thus engages and increases immersion.²⁸ One of the challenges in the field of virtual reality is to create environments which resemble physical reality as closely as possible. The use of sound micro-environment closely linked to the visual context can provide a significant boost to this kind of immersiveness.²⁹

2.2.3 Immersive Audio

In the virtual reality experience, the sensation of 'being there' in the simulated environment is not supported only by vision, but by sound modality too. Because our ears can support the effect of being in a virtual space as much as our eyes, appropriate auditory cues are as important as visual cues.³⁰ The complete 'being there' experience comes from the overall sensation of all sensory stimulus combined together. 'Although vision tends to be regarded as the dominant sensory channel, it is reported that auditory cues are important to the establishment

²⁵ Konstantinos Drossos, Andreas Floros, and Nikolaos-Grigorios Kanellopoulos, "Affective acoustic ecology: towards emotionally enhanced sound events." In *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound (AM '12, 2012)* Corfu, Greece. (New York, NY: ACM, 2012), 109, <http://dx.doi.org/10.1145/2371456.2371474>.II

²⁶ Tom Garner, Mark Grimshaw, and Debbie Abdel Nabi, "A Preliminary Experiment to Assess the Fear Value of Preselected Sound Parameters in a Survival Horror Game." In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound (AM '10, 2010)* Piteå, Sweden. (New York, NY: ACM, 2010), 2, <http://dx.doi.org/10.1145/1859799.1859809>.

²⁷ Tom Garner, and Mark Grimshaw, "A Climate of Fear." In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound (AM '11, 2011)* Coimbra, Portugal. (New York, NY: ACM, 2011), 6, <http://dx.doi.org/10.1145/1859799.1859809>.

²⁸ Natasa Paterson et al., "Spatial Audio and Reverberation in an Augmented Reality Game Design." In *Proceedings of the AES 40th International Conference (2010)* Tokyo, Japan. 6, <http://www.aes.org/e-lib/browse.cfm?elib=15532>.

²⁹ David Triantafyllou, Angeliki Antoniou, and George Lepouras, "Sound and Kinesthesia in Virtual Environments." In *Proceedings of the 20th Pan-Hellenic Conference on Informatics - PCI '16, 2016*, Patras, Greece. (New York, NY: ACM, 2016), 1, <https://doi.org/10.1145/3003733.3003737>.

³⁰ Thomas D Rossing, F.Richard Moore, and Paul Wheeler, *The Science of Sound* (San Francisco: Pearson, 2002), 686.

of a full sense of presence in virtual environments.³¹ Kevin Cheng and Paul A. Cairns claim that a particular barrier blocking the immersion exists when the different aspects do not cohere across different modalities.³² So, if a sound designer aims to create immersive audio, such design needs to go hand in hand with other modalities of the work.

Although the spatialized sound increases the feeling of immersion when properly combined with other modalities, it does not enhance the feeling of realism.³³ In my opinion, the audio content needs to be believable but not necessarily realistic. I deem that if the audio content is believable, it will be also engaging to users and lead to the increase of immersion. Therefore, careful research on audio content is required in order to create authentic and engaging experiences.

2.3 Binaural Sound

2.3.1 Mono and Stereo

From mono sound recordings, the listener can hear a single point sound location represented from one speaker (Fig. 5). There is one downside in this scenario; the mono sound can not convey the breadth and depth of sound.³⁴ Consequently, because mono sound limits the presentable spatial information, the concept of a stereophonic system was developed.³⁵

Headphones are a great example of stereo system with its left and right channel speakers. The brilliance of a stereo system is the capability to represent a wider sound image than a mono system, meaning that listeners are able to experience the horizontal location of sound (Fig. 5). In a sound mixing process, a recording engineer is able to relocate the sound by crossfading the amplitude of a signal from the left speaker to the right speaker, or anywhere in the middle between those two points. The area for this line is referred to as stereo image

³¹ Chueng, "Designing Sound Canvas," 848.

³² Kevin Cheng, and Paul A. Cairns, "Behaviour, Realism and Immersion in Games." In *proceedings of Human Factors in Computing Systems - CHI '05*, 2005, in Portland, OR. (New York, NY: ACM, 2005), 1272, <http://dx.doi.org/10.1145/1056808.1056894>.

³³ Claudia Hendrix, and W. Barfield, "Presence in Virtual Environments as a Function of Visual and Auditory Cues." In *proceedings of Virtual Reality Annual International Symposium '95* (1995) in Research Triangle Park, NC. 80, <https://doi.org/10.1109/VRAIS.1995.512482>.

³⁴ Paul Théberge, Devine Kyle, and Tom Everett, *Living Stereo: Histories and Cultures of Multichannel Sound* (London: Bloomsbury Academic, 2015), 129.

³⁵ Jens Ahrens, *Analytic Methods of Sound Field Synthesis* (Berlin: Springer, 2012), 4.

and the relocation of a sound on the line between the speakers is called amplitude panning.³⁶ Although this creates a feeling of a wider sound ambience, a listener can only hear the stereophonic sound reproduction within the limitations of a two-speaker array.

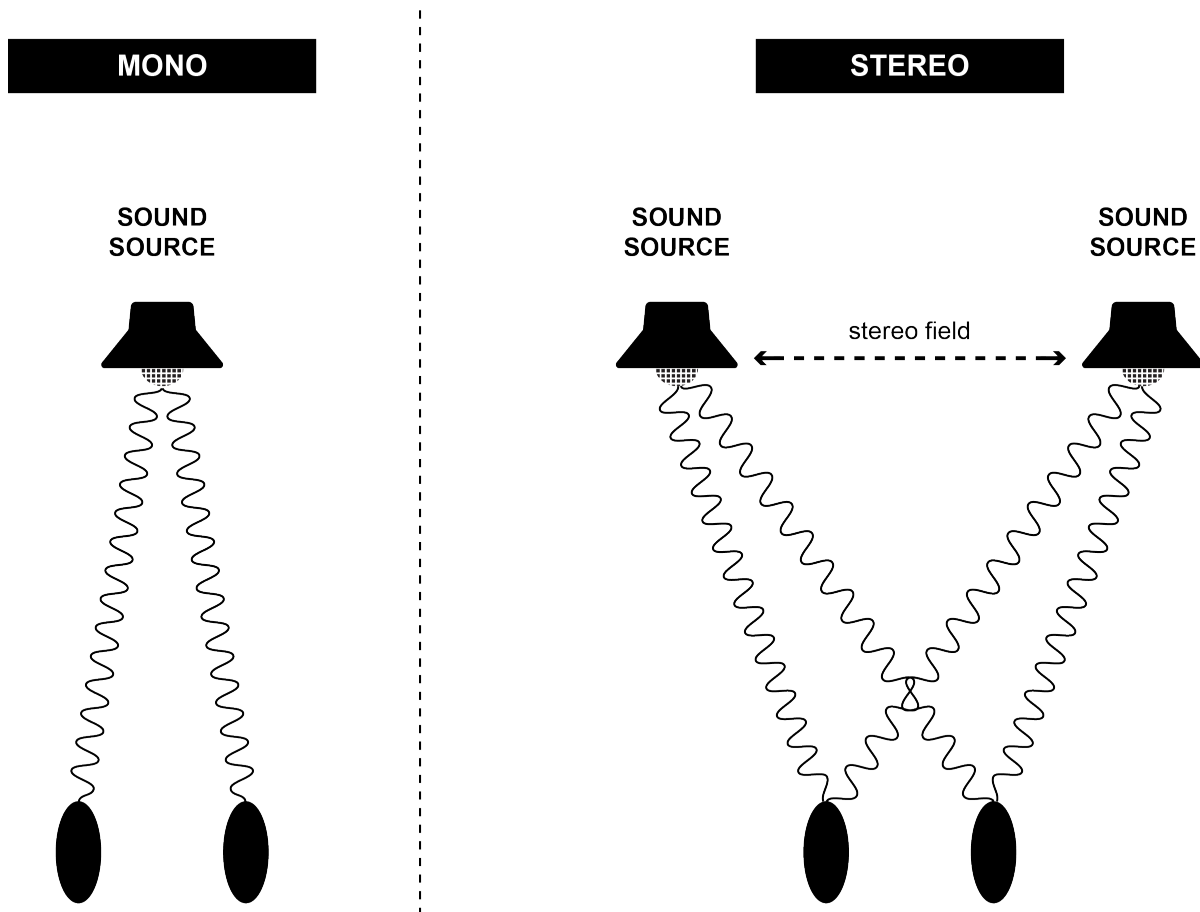


Figure 5: Example of mono and stereo sound sources.

2.3.2 Natural Spatial Hearing

According to Durand Begault's studies, the way we hear things naturally in our everyday environment is called 'natural spatial hearing'.³⁷ The human auditory system can easily recognize spatial audio cues from the surroundings, for example an aeroplane flying above or a dog barking in the distance. All of these can be easily located in the context of a 3D environment. A binaural audio system aims to create this effect of realistic spatial hearing when headphones are used.

³⁶ "Stereo Imaging." Wikipedia, February 18, 2018, https://en.wikipedia.org/wiki/Stereo_imaging.

³⁷ Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia* (San Diego: NTIS, 2000), 4.

The advantage of using headphones is that it enables the positional control of the speakers in relation to ears resulting a listening sweet spot; a position where the binaural sound functions the most effectively.³⁸

Binaural audio works in a similar manner to the stereo listening discussed earlier, except that binaural takes into consideration the functions of the human auditory system. There are reflections and obstacles that affect the sound waves travelling from the source to the ear which need to be translated into an auditory experience. Head, shoulders and our auditory system (ears) all affect how we hear surrounding sounds. The biggest difference between stereo and binaural hearing is that the latter provides spatial information in the form of time, intensity and coloration of sounds.³⁹

In a virtual environment, spatial audio cues provide valuable information about the surroundings. The sonic experience enveloping us from all around is an important factor in creating a believable virtual experience.

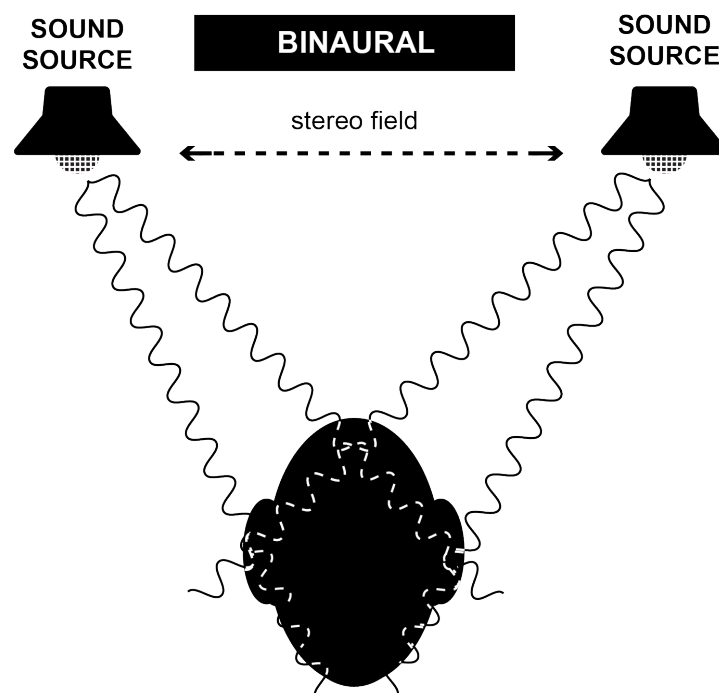


Figure 6: Example of a binaural listening situation.

³⁸ Agnieszka Roginska, and Paul Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (Routledge, 2018), chap. 4, doc. 1, https://www.safaribooksonline.com/library/view/Immersive+Sound/9781317480105/xhtml/Ch04.xhtml#_15031-1076_h1_ch004_001.

³⁹ Roginska, and Geluso, *Immersive Sound*, chap. 4, doc. 7.

2.3.3 Evolution of 3D Audio

The origins of binaural sound can be traced back to the 19th Century. At the Paris world expo of 1881 there was the first public representation of a “Théâtrophone”, a device with two separate speakers for both ears.⁴⁰ From there on the concept of 3D audio has developed from dummy head microphones to sophisticated algorithm-based software turning any sound sources into 3D audio.

In 1978, the BBC made history by producing a first binaural radio play entitled ‘Revenge’.⁴¹ Another narrative that used binaural recording is the ‘virtual barbershop’ made in 1993 by Qsound Labs.⁴² This is perhaps the most well-known binaural recording of all. These plays were both used dummy head microphones to create the binaural sound. Binaural sound recordings with dummy heads are static, meaning that it simulates an environment where listener’s head position does not change.

Head movements are part of our natural way of probing our surroundings and when discussing interactive applications such as virtual reality environments, it is necessary that the audio is not static but must follow the listener’s head movements. Thus, there is a need for a responsive audio representation method that follows the head-tracking and adjusts the played binaural sounds according to head movements. When head tracking is allowed in real-time applications, binaural audio is dynamic and updated continuously depending on the azimuth and elevation of a head tracked position, enabling accurate spatial localisation. It is essential to utilise a binaural rendering with head tracking when aiming for an immersive VR experience. One way to do binaural rendering in real-time is to utilise HRTF filtering discussed in more detail in next section.

Until recent times, spatialized audio has remained as an experimental audio effect instead of becoming a commonly established audio standard. Even though the possibilities of spatialized audio are unquestionable, the concept is not common in the field of audio design. Fortunately, after the first Oculus Rift headset was released in 2013, things have started to change regarding interest in 3D audio among audio content creators. It is only now that virtual reality systems

⁴⁰ "History (and Future) of 3D Audio." Hooke Audio, last modified August 21, 2017, <https://hookeaudio.com/blog/2016/10/11/history-and-future-3d-audio>.

⁴¹ Ibid.

⁴² "QSound Labs Inc. - Leaders in 3D Audio Enhancement & Virtual Surround Sound Technology." QSound Labs: Virtual Barber Shop, accessed on February 2, 2018, http://www.qsound.com/demos/virtualbarbershop_long.htm.

have reached consumers, presenting the possibilities and the need for 3D audio systems which virtual reality, in my opinion, deserves. Virtual reality media systems have huge potential in a wide range of applications in different fields of industry, from education to medical, design to entertainment and the estimated growth in virtual reality industries indicates an increasing need for high quality spatial audio systems.⁴³

2.4 3D Audio

2.4.1 Basics

According to Rossing, Moore and Wheeler, to spatialize a virtual sound there needs to be control over parameters of azimuth, elevation, distance and motion.⁴⁴ The location of sound sources are specified in terms of azimuth, elevation and distance. Sounds have properties such as pitch, loudness, tone, timbre.⁴⁵ Timbre is often known as a quality that distinguishes one sound from others with a similar loudness and pitch. It can be called a tone colour. In spatial hearing the timbre of sound differs by the attributes of placement in spatial locations.⁴⁶ Properties of sounds and their locations in space are mixed with different reverberations and echoes, reflecting from surrounding surfaces. The three-dimensional properties of these sounds come from their spatial location. These sounds are received by our auditory systems in two ways. The first is without any barriers through air conduction via the ears and the second is through the bone conduction of our head.⁴⁷

The human auditory system is a combination of three different sections: the outer ear, the middle ear and the inner ear. ⁴⁸ The outer ear consists of three different parts: pinna (bone and skin structure that is visible), auditory canal and the eardrum. The pinnae are like human fingerprints; every individual has a different structure of pinnae. These pinnae work as our probes for collecting and determining directions of high frequency sound waves.⁴⁹ The

⁴³ "Reality of VR/AR Growth." TechCrunch, last modified January 12, 2017, <https://techcrunch.com/2017/01/11/the-reality-of-vr-ar-growth/>; "3d Audio Market." New Report by TMR, accessed March 17, 2018, <https://www.transparencymarketresearch.com/3d-audio-market.html>.

⁴⁴ Rossing, Moore, and Wheeler, *The Science of Sound*, 687.

⁴⁵ Kirsten Cater et al., "An Investigation into the Use of Spatialised Sound in Locative Games." In *proceedings of Human Factors in Computing Systems (CHI '07, 2007)* in San Jose, CA. (New York, NY: ACM, 2007), 2316, <https://doi.org/10.1145/1240866.1241000>.

⁴⁶ Begault, *3-D Sound for Virtual Reality and Multimedia*, 27.

⁴⁷ Roy S. Kalawsky, *The Science of Virtual Reality and Virtual Environments: A Technical, Scientific and Engineering Reference on Virtual Environments* (Boston: Addison-Wesley, 1994), 69.

⁴⁸ Rossing, Moore, and Wheeler, *The Science of Sound*, 81.

⁴⁹ *Ibid.*, 81.

positional location of ears makes them great for positioning spatial sound cues in the horizontal axis. However, localizing sounds in the vertical axis is more complex for the same reason. Because ears are located on the opposite sides of the head, there are differences in the arrival time of sounds reaching each of our ears. This is called the interaural time delay.⁵⁰ In addition to time delay, there are masking effects caused by barriers that sound waves penetrate and reflect from, causing colouration of sounds. Therefore, before the sound reaches the inner ear, it gets affected by a complicated filtering process of acoustical interaction with the torso, head and pinnae.⁵¹ All of these elements contribute to the total spatial auditory perception.

2.4.2 ITD, IID and SD

The duplex theory is a spatial sound localisation theory proposed by Lord Rayleigh in 1907. It relies on two significant components: interaural time difference and interaural level difference (referred more commonly and in this master's thesis as interaural intensity difference).⁵²

Interaural time difference (hereinafter ITD) refers to the time difference of sound waves travelling to both ears. If a sound source is located straight in front of the head, ITD could, in theory, be non-existent (assuming that both of the listener's ears are absolutely identical). If sound waves were arriving from the right (or 90 degrees azimuth), the sound waves would arrive to the right ear first and then, after a short time delay, also to the left ear (Fig. 7).⁵³

The time difference between the ears depends on the body properties of the listener. Approximation of the maximal time difference between two ears can be said to be 660 microseconds (distance between two ears approximately 22-23cm).⁵⁴ Humans are able to perceive time differences as small as 10 microseconds.⁵⁵ This is an accurate way for the human auditory system to locate spatial sounds in azimuth in angles as small as 1° or 2° (time difference of about 11 microseconds), depending on the listener's position.⁵⁶

⁵⁰ Kalawsky, *The Science of Virtual Reality and Virtual Environments*, 69.

⁵¹ Cater et al., "An Investigation into the Use of Spatialised Sound in Locative Games," 2317.

⁵² Jeroen Breebaart, and Christof Faller, *Spatial Audio Processing: MPEG Surround and Other Applications* (Chichester: John Wiley & Sons, 2007), Chap. 3, Doc. 3, https://www.safaribooksonline.com/library/view/spatial-audio-processing/9780470723487/sec_3.3.html.

⁵³ Begault, *3-D Sound for Virtual Reality and Multimedia*, 32.

⁵⁴ "Interaural Time Difference." Wikipedia, modified December 18, 2017, https://en.wikipedia.org/wiki/Interaural_time_difference.

⁵⁵ Stephen Adams, and Frank Boland, "On the Distortion of Binaural Localization Cues Using Headphones." In *Proceedings of IET Irish Signals and Systems Conference (ISSC 2010)* in Cork, Ireland. 2, <https://doi.org/10.1049/cp.2010.0492>.

⁵⁶ David Murphy, and Flaithrí Neff, "Spatial Sound for Computer Games and Virtual Reality" In *Game Sound Technology and Player Interaction: Concepts and Developments*, ed. Mark Grimshaw (IGI Global, 2011), 290, <http://dx.doi.org/10.4018/978-1-61692-828-5.ch014>.

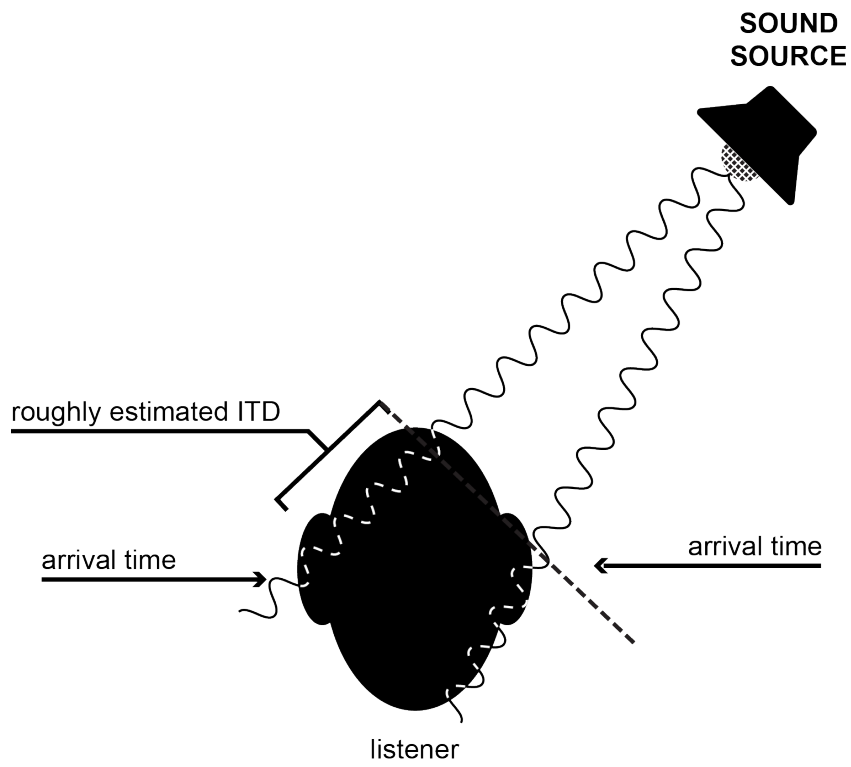


Figure 7: Time difference of sound waves travelling to both ears.

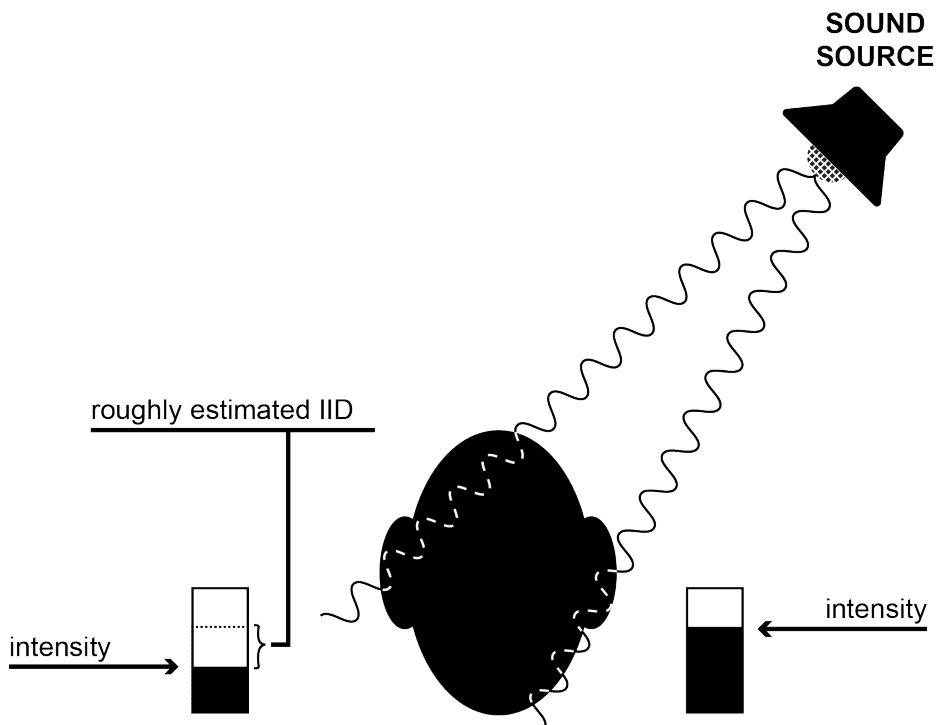


Figure 8: IID causing "acoustic shadow" absorbing energy from the left sound waves.

Interaural intensity difference (hereinafter IID) refers to the energy (sound pressure) difference of sound waves reaching our ears.⁵⁷ The effect of IID is caused by the interaction of sound waves and listener's head and shoulders, which creates an 'acoustic shadow', absorbing energy from sound waves. If sound waves were arriving from the right (or 90 degrees azimuth) the sound waves would arrive at the right ear first at a particular intensity but the barriers, head and shoulder, would drain energy from the sound wave before it arrived at the left ear. Thus, the intensity of the sound wave would be lower in the left ear (Fig. 8).⁵⁸

The IID is frequency dependent, meaning that some frequencies may arrive into both ears with similar energy whereas higher frequencies have huge differences in intensity. Human heads are only approximately 23cm wide, so very low frequencies, such as a 30 Hz sound wave (wavelength of 11.5 meters), travel through the head without losing any audible energy.⁵⁹ On the contrary, sounds that have a wavelength lower than 23cm will lose a significant amount of intensity when travelling through the head. Thus, the IID between two ears increases with frequency even though the spatial sound source was locationally constant.⁶⁰ At 1 kHz the sound is 8 decibels louder at the ear closer to the sound source. If the frequency is 10 times higher, the sound could be as much as 30 decibels louder at the closer ear.⁶¹

Similar to the frequency dependency of IID, the pinnae are also known to cause modifications to the frequency spectrum of sound. Due to the complex structure of pinnae, the spectrum modification is highly direction-dependent causing **spectral differences** (hereinafter SD) or spectral filtering where intensity depends on the direction of the sound source.⁶² The human ear is capable picking up even very subtle changes in frequency spectrum and translate those into spatial cues.⁶³

The unusual shape of pinna gives rise to reflections and resonances and filters frequencies. This changes the frequency spectrum of sound signals arriving to the eardrum. Pinnae are not the only body parts causing spectral differences, reflections of shoulders and other body parts

⁵⁷ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc.2.

⁵⁸ Murphy, and Neff, "Spatial Sound for Computer Games and Virtual Reality," 291.

⁵⁹ Michael Talbot-Smith, *Sound Engineering Explained* (Burlington: Focal Press, 2013), Chap. 1, Doc. 1, https://www.safaribooksonline.com/library/view/sound-engineering-explained/9780240516677/08_chapter01.xhtml.

⁶⁰ Begault, *3-D Sound for Virtual Reality and Multimedia*, 32.

⁶¹ Rossing, Moore, and Wheeler, *The Science of Sound*, 89.

⁶² Begault, *3-D Sound for Virtual Reality and Multimedia*, 41.

⁶³ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc. 2.

alters the spectrum.⁶⁴ However, it has been argued that pinnae cause the most significant locationally dependent effects on the spectrum of a sound source.⁶⁵

2.4.3 HRTF

HRTF stands for Head-Related Transfer Function. The basic assumption in using HRTFs is that it is possible to create a virtual acoustic space if the sound waves reaching the user's eardrums in free field conditions (in natural spatial hearing) are similar under the headphones. Therefore, the auditory experience should also be identical. In headphone listening conditions, HRTFs can filter non-binaural sound sources to simulate the natural hearing, enabling the perception of sound source location.⁶⁶

The general idea in creating HRTFs is to capture and study data regarding how sound waves behave in individual test subjects' ears. One method of studying this is to play short sound impulses (a single very short sound pulse or click) or other broadband sound stimulus from different angles around the test subject with a set of speakers with fixed intervals in vertical and horizontal axis. The played sound impulses are recorded with small probe microphones from the test subject's auditory canal. The smaller the fixed intervals are for the speakers, the better the directional resolution will be for the HRTF. When measuring these captured sound samples, it is possible to analyse the acoustic shaping made by spectral filtering, interaural time difference and interaural intensity difference for both ears from all the recorded angles. This collection of head related-impulse responses (hereinafter HRIR) can then be used to modify any sound source with a convolution that combines the original audio with HRTF and finally renders a binaural sound.⁶⁷ HRTF aims to create spatial directional cues without room acoustics. Therefore, HRTFs are created in silent and anechoic environments.⁶⁸

HRTF measurements are made for individual testers, which means that it works very well with a specific test subject's human properties but most probably not for others. Measuring a personalised HRTF for everyone would not be very practical and therefore one could argue that a generalised HRTF is needed. This can be generated averaging multiple HRTFs into a single

⁶⁴ Rumsey, *Spatial Audio*, 23.

⁶⁵ Begault, *3-D Sound for Virtual Reality and Multimedia*, 41.

⁶⁶ Alok Meshram et al., "P-HRTF: Efficient Personalized HRTF Computation for High-fidelity Spatial Sound." In proceedings of *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR, 2014)* Munich, German. 53, <https://doi.org/10.1109/ISMAR.2014.6948409>.

⁶⁷ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc. 3.

⁶⁸ *Ibid.*, chap. 4, doc. 3.

function.⁶⁹ However, this kind of generic HRTF in binaural reproduction may not sound realistic to everybody due to structural variations of pinnae. Generic HRTF is a compromise that can lead to unconvincing spatial impression but is the most practical solution for an average user and content producers.⁷⁰

2.4.4 Distance and Reverberation

Perception of acoustic distance is derived from two basic attributes: sound level and reverberation. First, sound waves arrive from a sound source to the auditory system in an unobstructed direct manner. This is called a direct sound. Based on research, it has been concluded that the amplitude of direct sound levels in anechoic room drop in accordance with the inverse square law, or about 6dB as the distance is doubled. However, this theory only applies to familiar sound sources such as a speech.

Sound level differences, in relation to distance, are also sound source related.⁷¹ Because of this, the creation of believable distance cues is more complex than just using 6dB attenuation with each doubling of the distance. The sound level of unfamiliar sounds (such as a synthesized tone) drops roughly 9 to 10 dB if the distance is doubled. Furthermore, with widely extended sound sources, such as motorways, the amplitude of the sound source drops as much as 3dB per each doubling of the distance.⁷²

Contrary to direct sounds, there are also indirect sounds, in other words, reflections. Sound waves bounce from surrounding surfaces causing echoes and reverberations. The human ear is very sensitive to environmental acoustics that affect the timbre of the sound, amplitude envelope (attack and decay) and cognitive familiarity of the sound.⁷³ In acoustic spatialization these are reverb cues that consists of early reflections and dense reverberation. Early reflections are sound waves that arrive to the auditory system from between 30 ms to 80 ms after direct sound. Reflections after 80ms tend to accumulate and spread rapidly, which leads to a dense audible reverberation. The character of the reverberation has a distinct effect on how the shape and size of an acoustic space is perceived.⁷⁴

⁶⁹ Rossing, Moore, and Wheeler, *The Science of Sound*, 690.

⁷⁰ Roginska, and Geluso, *Immersive Sound*, chap. 4, doc. 6.

⁷¹ *Ibid.*, chap. 1, doc. 5.

⁷² Rossing, Moore, and Wheeler, *The Science of Sound*, 688.

⁷³ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc. 5.

⁷⁴ Rossing, Moore, and Wheeler, *The Science of Sound*, 689.

The reverberant-to-direct sound ratio (hereinafter R/D) refers to the sound level balance between direct and indirect sounds. If a direct sound is dominant and a reverberant sound is quiet, the sound source is perceived to be closer to the listener compared to a reverse of the situation. R/D sound ratio changes the locational perception even if the sound level of a direct sound was constant, depending on the blend of the reverberation mix.⁷⁵

2.4.5 Motion

Head motion cues are an important source of information for sound localisation. When trying to localise sounds, we move or turn our heads to point towards the sound source in order to minimize interaural differences.⁷⁶ It has been proven that if a listener is allowed to move his/her head, the ability to localise sounds is improved.⁷⁷

The everyday environments that surround us are lively and constantly changing with sensory events. Just like moving our head causes dynamic variations, moving sound sources around us cause variations in our hearing.⁷⁸ We are consciously or unconsciously aware of this effect, taking advantage of the interactivity of sound sources to locate them.⁷⁹

Motion cues also include a Doppler effect which can create the impression of source distance.⁸⁰ It is a distance-dependent effect where sound waves increase and decrease in frequency depending on the direction of the movement in relation to the listener. If sound waves are emitted in the direction of the motion, the pitch of the sound is increased and vice versa.⁸¹ Pitch shifting caused by the Doppler effect may not always be desirable. In some contexts, a simulated Doppler effect can sound very unnatural and may cause musical pitches to go out of tune.⁸²

⁷⁵ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc. 5.

⁷⁶ Begault, *3-D Sound for Virtual Reality and Multimedia*, 39.

⁷⁷ *Ibid.*, 39.

⁷⁸ *Ibid.*, 40.

⁷⁹ Roginska, and Geluso, *Immersive Sound*, chap.4, doc.4.

⁸⁰ Rossing, Moore, and Wheeler, *The Science of Sound*, 689.

⁸¹ Jens Ahrens, and Sascha Spors, "Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect" In Audio Engineering Society Convention Paper 7363, Presented at the 124th Convention (2008) Amsterdam, Netherlands, 2, accessed January 5, 2018, <http://www.aes.org/e-lib/browse.cfm?elib=14493>.

⁸² Rossing, Moore, and Wheeler, *The Science of Sound*, 689.

2.4.6 Issues in 3D Audio

The 'cone of confusion' is a commonly acknowledged problem associated with audio spatialization.⁸³ If 3D audio relies only on ITD and IID, there are a series of positions where sound sources produce identical interaural values leading to front-back confusions. If there are no spectral differences between two sounds in cone of confusion, the binaural effect sounds identical and causes obscurity.⁸⁴ However, these ambiguities of signal directionality can be resolved by utilising spectral filtering or by means of head motion, as discussed in previous subsection.⁸⁵

There are also other problems complicating an accurate virtualisation of spatial audio. First of all, the headphone models that consumers use may reproduce negative results. Headphones differ in their frequency response, dynamic range and method of mounting, which may lead to distortion in HRTFs. It has been found in studies that headphone reproduction of delicate spatial localization cues have the potential to introduce artefacts to IID and ITD in some frequencies.⁸⁶ To compensate the issue with headphones, a proper calibration (equalisation) for the listening device should be executed. Headphone transfer function (hpTF) is a solution for this problem.⁸⁷ However, in practice it may not always be an executable method for entertainment consumers. A second possible problem may be phase and frequency response errors caused by the audio signal chain, for example a sound card.⁸⁸

Furthermore, there is the problem with generic HRTFs as discussed in subsection 2.4.3. Although, there are some common features in people's HRTFs, functional generalised filters are difficult to make for commercial systems because HRTFs are always specific to the individual.⁸⁹ Considerable research has taken place to identify HRTF characteristics that seem to occur in most people. This makes it possible to create a generic HRTF that functions for the majority of listeners.⁹⁰ However, it is theorised that to achieve truly immersive spatial sound, the individualisation of HRTF is crucial.⁹¹ Until a commercially viable solution for individual

⁸³ Roginska, and Geluso, *Immersive Sound*, chap.1, doc.2.

⁸⁴ Begault, *3-D Sound for Virtual Reality and Multimedia*, 41.

⁸⁵ Blauert, *Spatial Hearing*, 180; Roginska, and Geluso, *Immersive Sound*, chap.1, doc.2.

⁸⁶ Adams, and Boland, "On the Distortion of Binaural Localization Cues Using Headphones." 4.

⁸⁷ *Ibid.*, 2.

⁸⁸ Rumsey, *Spatial Audio*, 67.

⁸⁹ *Ibid.*

⁹⁰ *Ibid.*

⁹¹ Murphy, and Neff, "Spatial Sound for Computer Games and Virtual Reality," 288.

HRTFs is found, the front-back confusion and sound localisation of elevation remains a challenge.⁹²

In this chapter the meaning of immersion and the role of audio in VR context was presented. It became evident that immersion is a central feature in a virtual environment, and that audio has a lot of theoretical potential in supporting and improving the level of immersion in virtual environments. 3D audio is an essential element in creating those immersive VR experiences. Thus, it was necessary to give an overview of the basic theories and principles that spatializer plug-ins utilise, in order to create realistic binaural representations.

⁹² Murphy, and Neff, "Spatial Sound for Computer Games and Virtual Reality," 297.

Chapter 3

Evaluation of the Spatializers

Finding an audio spatializer plug-in that would work perfectly for all users is impossible, as the results depend on each individual's physical properties. However, sound designers working on VR content need the best possible solutions for creating 3D audio spatialization, regardless of the user's physical properties. Thus, an empirical test was necessary to indicate which spatializer plug-in would perform best, in terms of perceived directional precision, for the majority of people. In addition to this directional quality test, and to prove my point, I wanted to measure how much immersion increased when the most highly rated plug-in was used, compared to non-spatialized audio. In this chapter, 3D audio spatializer research will be discussed. Furthermore, the preparation, execution and results of the conducted experiments will be presented.

3.1 Spatializers for Unity

When developing the Virtual Heritage prototype, the results of Unity's default plug-in, Oculus, were insufficient. Therefore, a search for alternative plug-ins was needed. Eventually, five other feasible plug-ins were found: Resonance, Realspace 3D, Steam Audio, 3D Sound Labs and dearVR. Some of these spatializers also work on other platforms (as listed in Tab. 1, p. 28), although only Unity2017.3 was tested during this project. The reason that these specific spatializer plug-ins were included in this thesis was because during the time of the Virtual Heritage project work they were easily available with proper customer support, upcoming updates and support for both OSX and Windows operating systems. There are other spatializer plug-ins in addition to the ones researched, but those are excluded from this thesis.

In addition to the spatialization, there are other features (Tab. 1) of these plug-ins, that can contribute to the creation of audio experiences. Features such as tone control, room simulation, acoustic materials, and reverb are all tools that sound designers value in their work. It was assumed that all of the spatializer plug-ins utilise generic HRTFs discussed in section

2.4.3. However, apart from Unity and Google, the developers of the spatializers do not provide information about utilised HRTF databases.

Oculus plug-in is the default spatializer in Unity. It uses HRTF filtering based on KEMAR database. This data set is made by Bill Gardner at MIT Media Lab in 1994 and it is publicly available.⁹³

Resonance is a spatializer plug-in provided by Google. It has evolved from their previous 3D audio plug-in Google Audio.⁹⁴ Resonance uses HRTF filtering based on SADIE database.⁹⁵

dearVR is a spatializer plug-in made by a company called *Dear Reality*. The company was founded in 2014 in Düsseldorf.⁹⁶

Steam audio is a spatializer plug-in based on Phonon3D by *Impulsonic* that was acquired by video game developer and digital distributor Valve Corporation.⁹⁷

Realspace 3D audio plug-in is a spatializer created by *Visisonics*, a company founded in 2010 by the University of Maryland Computer Science faculty and researchers.⁹⁸

3D Sound Labs was founded in 2014 and is a French company based in Paris, focusing on 3D audio products.⁹⁹

⁹³ "HRTF Measurements of a KEMAR Dummy-Head Microphone." MIT Media Lab, accessed December 18, 2018, <http://sound.media.mit.edu/resources/KEMAR.html>.

⁹⁴ Jon Fingas, "Google Built a Spatial Audio Kit for Games and VR." Engadget, modified November 13, 2017, <https://www.engadget.com/2017/11/06/google-resonance-audio/>.

⁹⁵ "Google/spatial-media." GitHub, accessed April 4, 2018, <https://github.com/google/spatial-media/tree/master/spatial-audio>.

⁹⁶ "JOBS – We Are Hiring!" DearVR 3D Audio Reality Engine, accessed January 26, 2018, <http://dearvr.com/jobs/>.

⁹⁷ Michael Klier, "Home." Designing Sound, modified January 15, 2017, <http://designingsound.org/2017/01/impulsonic-acquired-by-valve/>; "Immersive Audio Solutions." *Steam Audio*, accessed January 26, 2018, <https://valvesoftware.github.io/steam-audio/#learn-more>.

⁹⁸ "Company." VisiSonics, accessed February 22, 2018, <http://visisonics.com/company/#background>.

⁹⁹ "3D Sound Labs." Crunchbase, accessed February 22, 2018, <https://www.crunchbase.com/organization/3d-sound-labs#section-overview>.

Features:	Oculus (Unity)	dearVR	Resonance	Steam Audio	Real Space 3D	3D Sound Labs
HRTF	x	x	x	x	x	x
Custom/Personalized HRTF					x	x
HRTF database	KEMAR	?	SADIE	?	?	?
HRTF Interpolation (Bilinear)				x	x	
Occlusion		x	x	x	x	
Air Absorption				x		
Reflections	x	x	x	x	x	x
Reverb		x	x	x		x
Reverb probe			x	x		
Reverb Baking			x	x		
Room/space simulation	x	x	x	x	x	x
Acoustic Material Options			x	x	x	
Quality adjustment			x	x		
Directivity adjustment			x			
EQ adjustment		x				
Supported:						
Unity	x	x	x	x	x	x
Unreal			x	x		
Fmod			x	x		
Wwise			x		x	
iOS		x	x			x
Android		x	x	x		x

Table 1: List of features for all the spatializers.

3.2 Spatializer Test

The main objective of the empirical test was to find out if there are subjective preference differences in the directional quality between the tested 3D audio spatializer plug-ins. My hypothesis was that if a large sample size (60 participants) of test subjects listened to the audio plug-ins with their headphones, and evaluated the directional quality perceived, differences would be detected indicating that one plug-in would be preferred over the others.

3.2.1 Designing the Online Test

The online test was chosen as the test method because of the possibility to gather many participants and large amount of data in a short time.¹⁰⁰ The test was planned following 'within-subjects design', where each of the participants would use exactly the same material.¹⁰¹ I was advised to use this research method by Assistant Professor Perttu Hämäläinen.

The objective was to gather a minimum of 60 test subjects from the game and audio industries. The online test was divided into two experiments. In the first part, users had to rate the overall and directional precision of the plug-ins. This test is called the 'quality test'. In the second test, the participants were asked to choose from six directions (front, back, left, right, below and above), corresponding to where they heard a test sound in a 3D space. This test is referred to as the 'direction test'.

At this point is necessary to clarify the term 'quality' used in this thesis. Quality is a rather broad concept, and it can be measured objectively or subjectively with a number of different factors and characteristics of a sound.¹⁰² However, in this work the word quality is used to express preferred sound characteristics by subjective listeners. As discussed in the book *Immersive Sound* by Roginska: "It is a listener's perceived judgement of the quality of a spatial auditory image that is more relevant to the actual listening experience."¹⁰³ The word has been used to drive test participants to answer the questions in a way that they have perceived the

¹⁰⁰ Gjoko Muratovski, *Research for Designers: A Guide to Methods and Practice* (London: Sage Publications, 2016), 117.

¹⁰¹ Kendra Cherry, "Learn What It Means When an Experiment Has a Within-Subject Design?" *Verywell Mind*, accessed January 19, 2018, www.verywellmind.com/what-is-a-within-subjects-design-2796014.

¹⁰² "Sound Quality." Wikipedia, last modify March 05, 2018, https://en.wikipedia.org/wiki/Sound_quality.

¹⁰³ Roginska, and Geluso, *Immersive Sound*, chap. 4, doc. 7.

spatial audio image in a positive context. It is also important to notice that the naming of the first experiment: ‘quality test’, is only to direct participants to focus and rate qualitative aspects of the audio samples. The second experiment, ‘direction test’, also measures the sound quality at a subjective level. However, the purpose of the test is to enable the audience to choose the correct direction of the sound source, thus naming it ‘direction test’.

It was assumed among my colleagues and thesis advisor (Petteri Mäkineniemi, Sebastien Piquemal, Marko Tandefelt) in a meeting on 2.12.2017, that the test package should be small in size in order to get enough participants to download it. The desirable small test package size forced me to forget uploading application builds (six builds would have been approximately three gigabytes) with interactive elements, such as a head-tracking. Additionally, the visual elements and appearance of the moving object had to be left out. These inadequacies have to be taken into consideration when evaluating the results, because both the head motions and visuals have a connection to the perception of spatial audio cues as already discussed in subsections 2.2.1 and 2.4.5.¹⁰⁴

I planned to create a package including uncompressed (wav.) test audio files. This way, the final compressed package size was 48,8 megabytes. To keep the test as simple and quick as possible, I included only the necessary information in the instructions, and kept the audio files less than 30 seconds long each.

A different procedure was used for the two testing methods. In the first test, called ‘quality test’, an audio object moved around the listener and the participant was asked to rate the directional quality for sounds emitting from the left/right, front/back, above/below, as well as the overall sonic quality. In the second test, called ‘direction test’, sounds were played randomly from six static main directions and after each play, participants needed to indicate the perceived direction of the test sample. The main spatial directions referred to in the tests in this chapter are left, right, front, back, below and above.

It was necessary to design and build a Unity project that included all the six spatializer plugins in order to generate the test audio content. The process of creating the test setup and recording the audio files had multiple phases, described in detail in Appendix section A. In the process 13 audio files were recorded: six for the ‘direction test’ and seven for the ‘quality test’.

¹⁰⁴ Francis Rumsey, *Spatial Audio* (CRC Press, 2017), 34; *Ibid*, 67.

One audio file in the 'quality test' was a sound example. The sound example file revealed the audio object's movements as to prepare participants for the actual test audio.

At this point it is important to underline that these spatializer plug-ins differ from each other. Testing all six 3D audio plug-ins in an equal manner is a rather difficult task to accomplish. This is because of the varying nature and differing factors of all the evaluated 3D audio plug-ins. Each of the plug-ins has a different interface, features, processing algorithms and original default settings. The central aim was to use settings which would present the directional capabilities of each plug-in clearly. The settings of each plug-in will not be discussed in detail because of the large number of different parameters in each of the tools. However, the main principle was to use the default settings of each plug-in, while disabling the room simulation, reverb, Doppler effect and tone colorations (equalisation).

I conducted subjective listening tests to ensure that there were no sound 'sweeteners' turned on in any of the spatializers. Unfortunately, it is impossible to be completely certain if some of these spatializers have unknown effect processing happening beyond what user sees. When evaluating the test results, it has to be understood that something like this may happen. To conclude, as an audio designer, I would use the tools that most of the people preferred to listen to, whether or not some mysterious effect processing was taking place unbeknownst to anyone.

An informative text file, entitled 'read me first', was created to provide instructions about the test procedure (Appendix B). The text file includes a link to an online form created with Google Forms. This was the form for gathering the responses. The final test package was constructed, including 13 different audio files, and an instructional note.

It was suggested by Perttu Hämäläinen that the audio file order would be counterbalanced in order to avoid priming of certain audio files in the listening sequence. Therefore, the Latin Squares with six-condition program was used to arrange the files into six different test packages.¹⁰⁵ The order of the audio files can be reviewed in Table 2.

¹⁰⁵ Geoffrey Keppel, and Thomas D. Wickens. *Design and Analysis: A Researcher's Handbook* (Academic Internet Publishers Incorporated, 2007), 384.

Quality test						
Counter-balanced audio file nro:	1	2	3	4	5	6
Test package 1	3DSL	dearVR	Oculus	RS3D	Resonance	Steam
Test package 2	Steam	3DSL	dearVR	Oculus	RS3D	Resonance
Test package 3	Resonance	Steam	3DSL	dearVR	Oculus	RS3D
Test package 4	RS3D	Resonance	Steam	3DSL	dearVR	Oculus
Test package 5	Oculus	RS3D	Resonance	Steam	3DSL	dearVR
Test package 6	dearVR	Oculus	RS3D	Resonance	Steam	3DSL
Directional test						
Counter-balanced audio file nro:	1	2	3	4	5	6
Test package 1	RS3D	Resonance	Steam	3DSL	dearVR	Oculus
Test package 2	Oculus	RS3D	Resonance	Steam	3DSL	dearVR
Test package 3	dearVR	Oculus	RS3D	Resonance	Steam	3DSL
Test package 4	3DSL	dearVR	Oculus	RS3D	Resonance	Steam
Test package 5	Steam	3DSL	dearVR	Oculus	RS3D	Resonance
Test package 6	Resonance	Steam	3DSL	dearVR	Oculus	RS3D

Table 2: Latin Squares counterbalanced coding system for test packages.







Name	Date Modified	Size	Kind
 Test package version 1.zip	7 Feb 2018 12:08	48,8 MB	ZIP archive
 Test package version 2.zip	7 Feb 2018 12:27	48,8 MB	ZIP archive
 Test package version 3.zip	22 Feb 2018 18:15	48,8 MB	ZIP archive
 Test package version 4.zip	7 Feb 2018 11:34	48,8 MB	ZIP archive
 Test package version 5.zip	7 Feb 2018 11:39	48,8 MB	ZIP archive
 Test package version 6.zip	7 Feb 2018 12:01	48,8 MB	ZIP archive

Figure 9: Counterbalanced test package versions.

Name	Date Modified	Size
READ ME FIRST	7 Feb 2018 11:33	4 KB
▼ Test1 Quality	25 Jan 2018 00:11	--
0_Quality_Example.wav	23 Jan 2018 22:29	5,5 MB
1_Quality.wav	23 Jan 2018 17:26	5,1 MB
2_Quality.wav	23 Jan 2018 17:26	5,1 MB
3_Quality.wav	23 Jan 2018 17:26	5,1 MB
4_Quality.wav	23 Jan 2018 17:26	5,1 MB
5_Quality.wav	23 Jan 2018 17:26	5,1 MB
6_Quality.wav	23 Jan 2018 17:26	5,1 MB
▼ Test2 Direction	23 Jan 2018 21:58	--
1_Direction.wav	23 Jan 2018 17:26	5,9 MB
2_Direction.wav	23 Jan 2018 17:26	5,9 MB
3_Direction.wav	23 Jan 2018 17:26	5,9 MB
4_Direction.wav	23 Jan 2018 17:26	5,9 MB
5_Direction.wav	23 Jan 2018 17:26	5,9 MB

Figure 10: Files in the test package.

3.2.2 Data Acquisition and Distribution

In the online form the following demographic data was requested from the participants:

- Gender
- Age
- The state of hearing ability
- Educational background
- Headphones or earbuds

Studies reveal that gender and age affect how we perceived sounds. For example, females hear quieter sounds and understand language better, but males have a better capability for localizing sounds.¹⁰⁶ Also, the age effects the range of hearing.¹⁰⁷ It was assumed that level of education may have an effect on how people listen to audio sources. The state of hearing ability was requested. The users were also asked to mark if they used headphones or earbuds. This question was asked because the HRTFs work differently if the audio is produced outside of the ear canal (headphones), compared to inside (earbuds.)¹⁰⁸ All of this demographic data was gathered in case of prospective future research, but will not be studied further in this thesis.

¹⁰⁶ Leonard Sax, "Sex Differences in Hearing: Implications for best practice in the classroom." In *Advances in Gender and education*, 2, (Montgomery Center for Research in Child Adolescent Development, 2010), 18, accessed February 26, 2018, <http://www.leonardsax.com/sex-differences-in-hearing/>.

¹⁰⁷ "Age-Related Hearing Loss." National Institute of Deafness and Other Communication Disorders, modified January 16, 2018, <https://www.nidcd.nih.gov/health/age-related-hearing-loss>.

¹⁰⁸ Roginska, and Geluso, *Immersive Sound*, chap. 4, doc. 8.

The online study collected mainly ordinal data. In the “quality test”, Likert scale with seven points was used to gather the data as suggested by Perttu Hämäläinen. The participants were requested to rate the following qualities for each of the “quality” files (Fig. 11):

- Quality for Left/Right directionality
- Quality for Front/Back directionality
- Quality for Below/Above directionality
- Overall precision of the 3D audio

In the ‘direction test’ users were requested to choose the main direction of the audio cues (Fig. 11). The 3D audio locations were played in random order (Appendix A, Tab. 7, p. 70) from six different directions in each of the audio files. The participants then had to indicate the direction that they felt corresponded with the location of the sound source. A multiple choice grid was used to gather the answers for each of the tests.

Muratovski suggests in his book *Research for Designers*: “a few open-ended questions might give a new perspective or may provide with some valuable additional insight”.¹⁰⁹ Therefore, an open-ended question was added to get a better perspective on how the participants experienced the test and the spatialized audio content.

Dropbox links were used in the distribution of test packages. The objective was to gather a total of sixty test subjects; ten per each of the package versions. Test participants were found from my personal Facebook contacts and from three different Facebook groups: FIVR — Finnish Virtual Reality, Ambisonics VR 360 Audio, and Spatial Audio in VR/AR/MR (Fig. 12, p. 36). I promised to share the results for all the participants as a reward for their attendance. I contacted all people willing to participate individually. This way I could balance and direct the number of respondents towards testing each of the package versions. The postings were done on 22.02.2018 and the testing period lasted for one week.

The response was appreciable. People were very interested in participating in the test and some asked for permission to share the test links to their own connections as well. There were 98 answers in total, but they were not equally balanced between the packages, as seen in Table 3.

¹⁰⁹ Muratovski, *Research for Designers*, 111.

Quality test nro:1. Rate the quality for directions.

Rating: 1 = Very Poor, 2 = Poor, 3 = Fair, 4 = Neutral, 5 = Good, 6 = Great, 7 = Excellent

	1	2	3	4	5	6	7
Quality for Left/Right	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality for Front/Back	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality for Below/Above	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall precision for the 3D audio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Direction test nro:1

Choose the direction for the locations.

	Front	Right	Back	Left	Below	Above
Location A1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Location B2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Location C3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Location D4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Location E5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Location F6	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 11: Questionnaire for quality and direction tests.

Version:1 = 15	Version:2 = 14	Version:3 = 14	Version:4 = 14	Version:5 = 17	Version:6 = 23
----------------	----------------	----------------	----------------	----------------	----------------

Table 3: Responses rate to each of package versions.

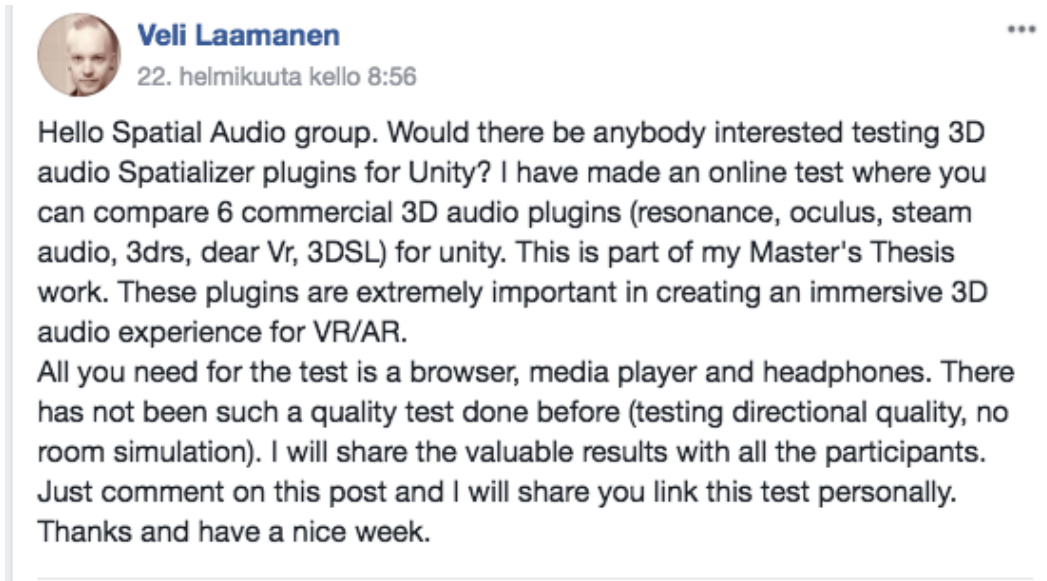


Figure 12: A following post was sent to each of the Facebook groups to gather participants.

Each of the package versions got at least 14 answers. However, there were three responses with several unanswered questions. Each of these responses were in different package versions. To avoid unbalanced answers within the packages, one response was removed from all the version packages including these three incomplete responses. Consequently, 13 responses per all of the six package versions were included in the data evaluation, resulting a total of 78 responses. For the data evaluation an average of subject Likert ratio was calculated for each question. The Google Sheets functional calculator was used for these calculations (Fig. 13).

	I	J	K	L	M	N
69	3	3	6	3	2	4
70	3	4	6	5	5	5
71	4	5	4	3	2	4
72	1	1	4	2	1	2
73	3	4	5	5	3	4
74	3	4	2	2	2	2
75	2	2	5	4	2	2
76	3	4	5	4	3	4
77	2	3	6	5	4	4
78	2	4	7	5	2	4
79	6	6	6	5		
80	3.179487179	4.038461538	5.833333333	4.5	3.538461538	4.545454545
81		4.105769231	DearVR			4.604312354

Figure 13: Calculating the average subject ratio per question.

3.2.3 Spatializer Test Results

A total of 78 responses were analysed for the final results. A seven-level format of the Likert scale was provided for the answers in the 'quality test'. The responses were organised in 24 histograms showing detailed subject ratio. The histograms can be reviewed in the Appendix C (Fig. 36-39, p. 74-77).

An average Likert ratio was calculated from all the 78 responses for each plug-in per question (Fig. 13). All of these mean values were rounded and organised into the histograms (Fig. 14). Furthermore, for each of the plugins, an average Likert ratio was calculated from the mean values derived from the four questions. The results indicate the total average of subject ratio in the 'quality test' (Fig. 15).

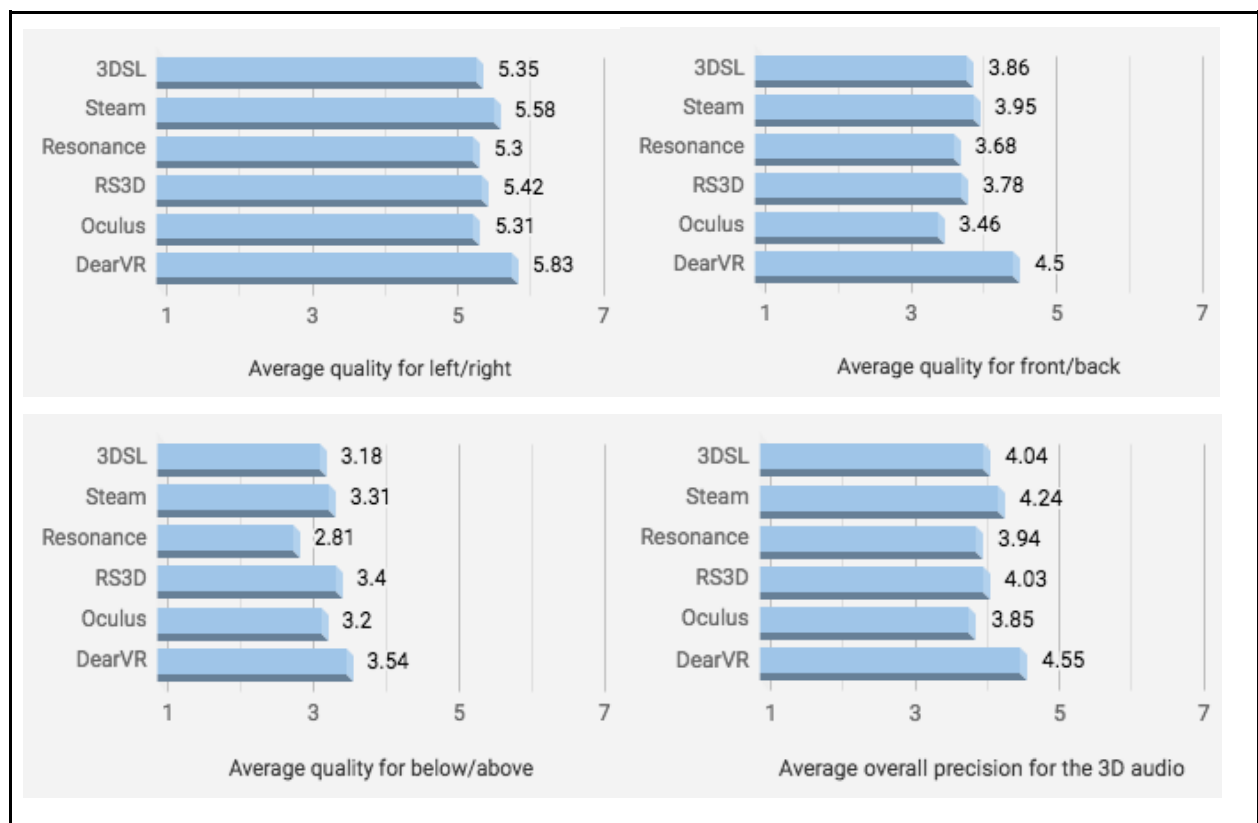


Figure 14: The average of subject Likert ratio for each plug-ins per question.

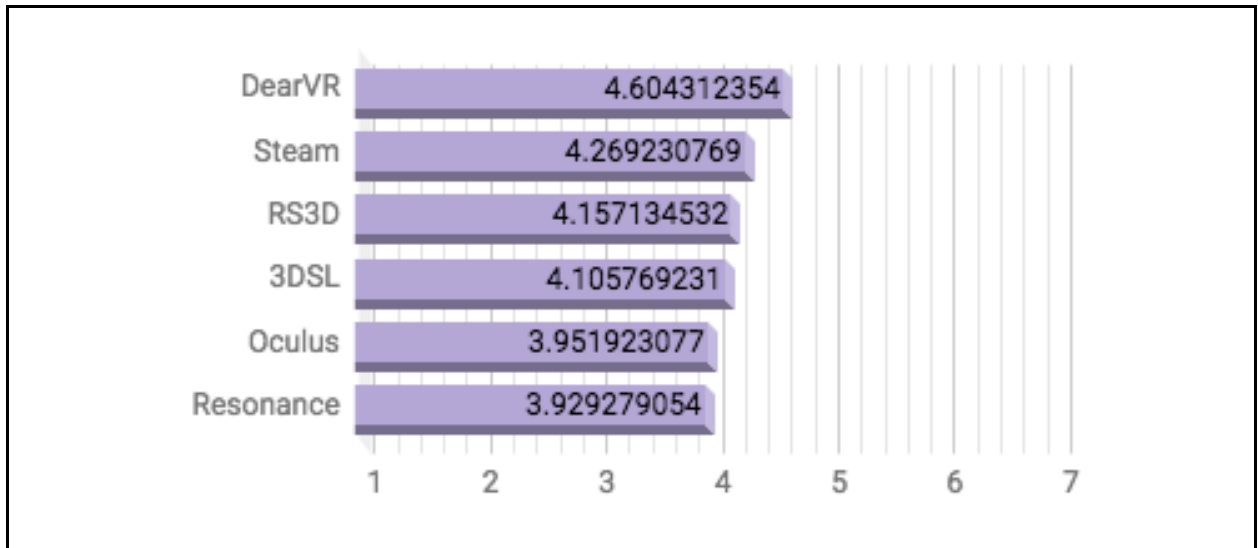


Figure 15: The total average of subject ratio in seven point Likert scale in quality test.

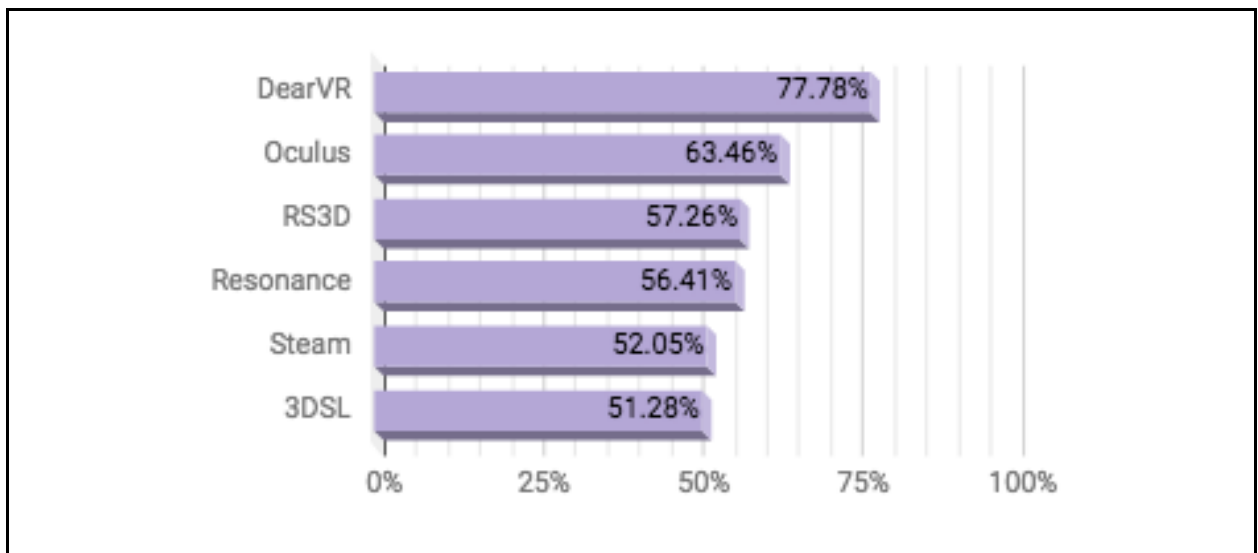


Figure 16: Average percentage of correct responses in direction test.

In the 'direction test' participants were asked to choose from six directions (front, back, left, right, below and above), corresponding to where they heard a test sound in a 3D space. A percentage ratio of correct answers per direction was calculated for each tested plug-in. These percentage ratios can be reviewed in Appendix C (Fig. 39 and 40, p. 77). The average percentage ratio of correct responses was calculated to indicate the average correction level for each plug-in (Fig. 16).

As reviewed in Figures 15 none of the spatializers were rated good, great or excellent (5, 6, 7 in Likert scale) in average. However, according to both experiments (Fig. 15 and 16), it is distinctly indicated that dearVR was the most highly rated spatializer of the six studied plug-ins. DearVR scored the best average ratio in all tests, except when questions about correct directions for left and right were requested in the 'direction test'. However, the average mean in these specific questions for all of the plug-ins was 97% and dearVR resulted 96.15% for both left and right. Therefore, it can be assumed that such a small difference between dearVR and other plug-ins is inconsequential. To conclude, dearVR was the most highly rated spatializer plug-in and was therefore tested in the immersion test, discussed in the next section.

83% of the participants responded to the open-ended questions. More than half of the comments reported front-back and above-below confusions. As discussed in section 2.4.6 this is a known issue and a commonly acknowledged problem associated with generic HRTFs, therefore these kinds of responses were expected. As reviewed in Figure 15, average ratings ranged from 3.93 to 4.6 (fair to neutral) on a 7-point scale. None of the spatializers got average ratings of 5, 6, 7 (good, great or excellent) possibly because of the reported front-back and above-below confusions. All of the evaluated data can be reviewed from the following URL (this link was shared to the participants):

<https://drive.google.com/drive/folders/14tNX-WwwHhZadPC1WJsZ65Z2vdTvH5wT?usp=sharing>.

3.3 Immersion Test

3.3.1 Test Procedure

The immersion test was conducted in order to find out how much the use of a spatializer plug-in in a VR project would have an effect on immersion. My hypothesis was that if two versions of virtual environments were compared in similar conditions, with the only differential factor being that one used spatial audio rendering and the other one did not, then immersion would be more highly rated in the spatialized version. In other words, the aim was to identify how much the use of spatial audio rendering increased immersion compared to non-spatialized audio.

Measuring the immersion of audio has challenges. The unclear definition of the word immersion and the emotional state of participant has an effect on results. Test participants were instructed with the text: "Your task is to think of how much you feel like "being there" in the

virtual environment while concentrating on audio.” Discussions were held with all participants in order to encourage the feeling of involvement and to ensure that all understood the meaning of immersion as concluded in the section 2.1.

The Virtual Heritage prototype was used as a test platform. The prototype will be introduced and explained in detail in Chapter 4. Three modifications were done to the prototype for the immersion test. Firstly, all reverb and room simulations were turned off. Secondly, interactive audio volume changes (Audio Zoom) were disabled, resulting in all audio sources playing with static sound levels. Thirdly, the display was turned off in order to have all participants focusing only on audio. Two different versions of the Virtual Heritage prototype were tested. The first version utilised the dearVR spatializer, and the other version used Unity’s proprietary audio localisation properties utilising amplitude panning methods.¹¹⁰

The testing for immersive audio was conducted in a closed classroom environment where test subjects were observed at all times. The testing environment was kept silent during the test, as surrounding sounds might distract the virtual experience.¹¹¹ All test participants used HTC Vive HMD with integrated Deluxe Audio Strap headphones.

In the test procedure the subjects tried the two VR versions for three minutes each. They were then requested to answer a questionnaire. The question about the immersion ratio was influenced by Hendrix and Barfield’s work on *Presence in Virtual Environments as a Function of Visual and Auditory Cues*. The main question was: ‘If your level of presence in the real world is “100”, and your level of presence is “1” if you have no presence, how would you rate your level of presence in tested virtual environments?’¹¹² A simplified supporting question ‘which version was more immersive?’ was also requested.

A group of seven male and one female participants, consisting of current and former students of Aalto University’s Media Lab, volunteered to participate in this experiment. Six of them were in the age group of 25-34 and two in the 35-44 group. All test subjects had a normal or excellent state of hearing ability. Half of the test group tested spatialized and the other half tested non-spatialized versions of the virtual environment first.

¹¹⁰ Technologies, Unity. "Audio Source Overview." Unity - Manual: Audio Source, accessed March 22, 2018, <https://docs.unity3d.com/Manual/class-AudioSource.html>.

¹¹¹ Triantafyllou, Antoniou, and Lepouras, "Sound and Kinesthesia in Virtual Environments," 2.

¹¹² Hendrix, and Barfield, "Presence in Virtual Environments," 75.



Figure 17: Test participant immersed.

3.3.2 Immersion Test Results

Seven out of eight of the subjects responded that virtual environment with spatialized audio was more immersive. Only one test participant felt that both of the versions were equally immersive. These thoughts were confirmed in open discussions that were held after the test sessions. Participants rated the level of presence according the Table 4. The version with spatial audio rendering was rated 15-60 percentage points higher in “level of presence” except for the one participant who felt no difference in presence. On average, subjects felt 36.25 percentage points higher “level of presence” when testing the Virtual Heritage prototype with dearVR. Four of the subjects consider the spatialized version to have twice or more higher “level of presence” compared to non-spatialized version.

No spatializer	dearVR	Percentage point difference in advantage of dearVR
40%	85%	45%
65%	80%	15%
40%	80%	40%
40%	90%	50%
20%	70%	60%
40%	80%	40%
70%	70%	0%
50%	90%	40%
Average 45.63%	Average 80.63%	Average 36.25%

Table 4: Subject ratio for “level of presence” in two versions of VR.

Therefore, it can be confirmed that the usage of dearVR spatializer plug-in in a virtual environment has an increasing effect on immersion when compared to an identical virtual environment not utilizing a spatializer. However, it has been said that the sense of presence is always individual, depending on a user’s personal preferences, senses and cognitive capabilities.¹¹³ Therefore, it has to be remembered that these results reflect preferences of university students studying new media art.

In this chapter six commercially available spatializer plug-ins were introduced and tested in an empirical online test to indicate which one of the plug-ins perform best in terms of directional precision. 78 responses were evaluated and the results indicate that majority of the participants preferred the dearVR spatializer. In addition, the effect of dearVR on immersion was tested in a closed environment with eight test subjects who tested two modified versions of the Virtual Heritage prototype. The first version utilised dearVR and the second version used the amplitude panning methods of Unity’s proprietary audio system. The test revealed that dearVR increased “level of presence” by average of 36.25 percentage points indicating that dearVR 3D audio solution increases immersion significantly.

¹¹³ Chueng, "Designing Sound Canvas," 848.

Chapter 4

Case study: Virtual Heritage

Presented in this chapter is the concept and prototype of the Virtual Heritage project.

The audio scene of the prototype was designed to be as authentic and historically valid as possible. Therefore, the research on discovering the cultural heritage of Sammallahdenmäki was an essential part of creating the musical and auditory content for the sonic environment. The transformation process from research to actual audio content is presented in this chapter. Furthermore, the audio implementation of the audio scene is discussed, and finally, the findings of this practice-led artistic research are discussed.

4.1 Prototype

4.1.1 Location

Sammallahdenmäki is the largest burial site from the Scandinavian Bronze Age, 1500-500 B.C. with 33 ancient burial cairns within an area of 36 hectares. Nowadays, all of these cairns are located between 19 and 43 meters above the present sea level.¹¹⁴ However, in the Bronze Age, the sea level was almost 30 meters higher due to the upthrust of ground, and the sea shore was located next to Sammallahdenmäki.¹¹⁵ The cairns are exceptional proof of religious structures and social life from an era before written documentation. There have been discoveries of burned human bones from the cairns and it has been stated that cremation was a common custom in Sammallahdenmäki during the Bronze Age.¹¹⁶ Some societies during the Bronze Age in Scandinavia practiced long distance trading, which enabled the craft of ship building to exist, and the trading of items to be possible.¹¹⁷ The fact that Sammallahdenmäki

¹¹⁴ Niina Uusi-Seppä, *Satakunnan Kulttuuriympäristöt - eilen, tänään, huomenna* (Pori: Satakunnan Museo, 2012), 30, https://issuu.com/satakunnan_museo/docs/eilen_tanaan_huomenna_web.

¹¹⁵ *Ibid.*, 34.

¹¹⁶ *Ibid.*, 32.

¹¹⁷ Johan Ling, and Per Cornell, "Violence, Warriors, and Rock Art in Bronze Age Scandinavia." In *Feast, Famine or Fighting? Studies in Human Ecology and Adaptation*, ed. Richard J. Chacon, and Ruben G. Mendoza (Springer, 2017), 19, accessed February 20, 2018, https://doi.org/10.1007/978-3-319-48402-0_2.

was once located next to the sea also provided fruitful opportunities for such trading customs. However, the artefacts found from the Sammallahdenmäki area are limited and evidence of the actual inhabitants is vague.¹¹⁸

4.1.2 Concept

I visited Sammallahdenmäki in October 2017 to get an overall impression of the location. After the visit, I designed a concept of audio experience using time travel as a metaphor. The core idea was to create virtual environments where users could feel as if they were travelling from the present back to the moments in the past in different tourism locations, and to understand and learn about their unique history via auditory storytelling. The idea is that the story is not told visually but relies completely on audio cues.

Cairns are the dominating elements in Sammallahdenmäki. Therefore, the audio scene in the developed prototype presents an ancient burial ceremony with Sammallahdenmäki inhabitants singing a funeral song.

Technology: <ul style="list-style-type: none">• Unity game engine as the main platform.• DearVR spatializer.
Audio Content: <ul style="list-style-type: none">• Carefully executed research on cultural heritage of Sammallahdenmäki in order to create authentic sonic environment.• Emphasis on mournful emotions when producing the musical content.
Interaction: <ul style="list-style-type: none">• HTC Vive as the main presentation hardware.• Interactive audio increasing/decreasing volume according to the user's field of vision, creating 'Audio Zoom' effect.

Table 5: Audio Model attributes for Virtual Heritage prototype.

¹¹⁸ "Asuminen." Sammallahdenmäki, accessed February 22, 2018, <https://www.sammallahdenmaki.fi/kouluille-pronssikausi-sammallahdenmaella/asuminen/>.

4.1.3 Unity

The Unity game engine 2017.3 with the HTC Vive HMD kit was used throughout the development of the prototype. The whole prototype was constructed in a single Unity scene. The project's functionality was planned and created in a modular structure so that changing and altering the scripts of components would be as flexible as possible. This prototype's functional structure is illustrated in Figure 18. However, I will not go into further details of the scripting and functional features of the prototype if they do not have connection to the audio design.

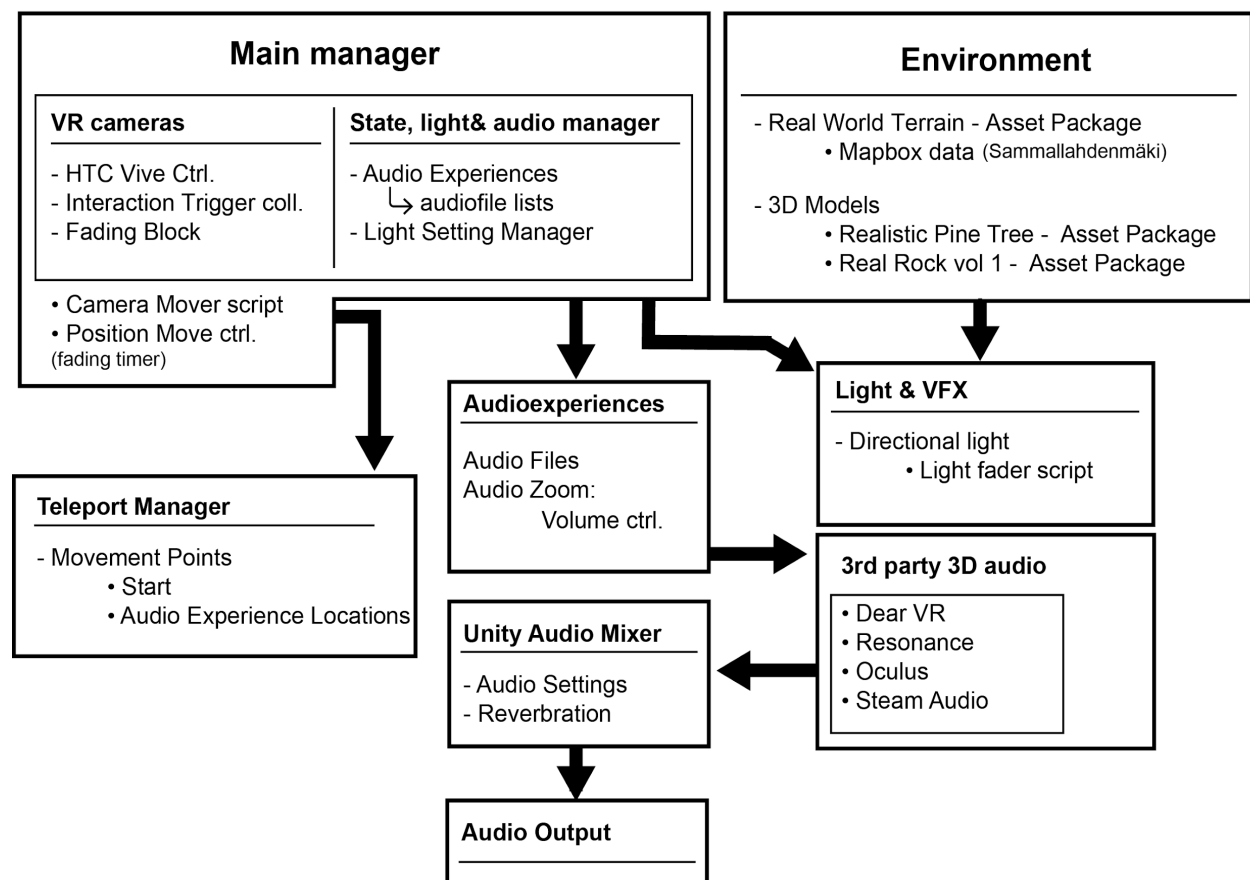


Figure 18: Simplified information architecture of the prototype's functionalities.

4.1.4 Visuals

The approach regarding the visuals of the Virtual Heritage project was that the graphics should be very minimalistic. Two aspects affected this approach: firstly, my technically limited skills in creating graphical content, and secondly, the thought that if there is nothing particular in the visuals to probe, then users would concentrate mostly on the sonic environment. Sounds were intended to work like an echoic memory from the moments of the past. I believed that the imagination would cover the lack of visual information if the auditory storytelling was

interesting and engaging enough. In practice, the display provides just enough information so that the head-tracking feels intuitive and natural. Therefore, the aesthetic of the visuals relies on static shadows and darkness, revealing only the horizon and profiles of natural objects. The inhabitants of Sammallahdenmäki are not represented visually. They can be only heard.

The overall appearance was realised using 3D modelled trees and rocks and with *Real World Terrain Builder* asset. All of these assets were bought from the Unity Asset Store. *Real World Terrain Builder* was used to take the Sammallahdenmäki area's geographical satellite photography, and utilising that as a terrain platform in the project in order to create accurate appearance for the location (Fig. 19).¹¹⁹ Mapbox service included in *Real World Terrain Builder* was used to create the ground elevation.¹²⁰ However, the elevation was somewhat incomplete and inaccurate, so it was customized with Unity's terrain elevation tool according to recollections from visiting Sammallahdenmäki.

The final visual setup can be reviewed from Figures 3 and 4 (p. 6). To conclude, it was the intention of this project to show the potential of audio. Therefore, aspects related to visual design are not discussed in any more detail in this thesis.

4.1.5 Audio Zoom

As discussed in section 2.1, it has been shown that there is a connection between interaction and immersion.¹²¹ As immersion is a desirable feature of a virtual environment, the sonic environment of the Virtual Heritage concept is designed to be reactive to the user's actions.

In the prototype, the audio interaction instance is referred to as 'Audio Zoom'. In practice, Audio Zoom means that sound levels increase or decrease according to the user's field of vision. This function enables users to control the sound levels of audio sources and therefore modify the overall soundscape to some extent.

¹¹⁹ "Real World Terrain." *Asset Store*, accessed November 10, 2017, <https://assetstore.unity.com/packages/tools/terrain/real-world-terrain-8752>.

¹²⁰ "About." *Mapbox*, accessed February 26, 2018, <http://www.mapbox.com/about/>.

¹²¹ Lee, Mingyu, and Jinmo, "A Study on Immersion," 1.

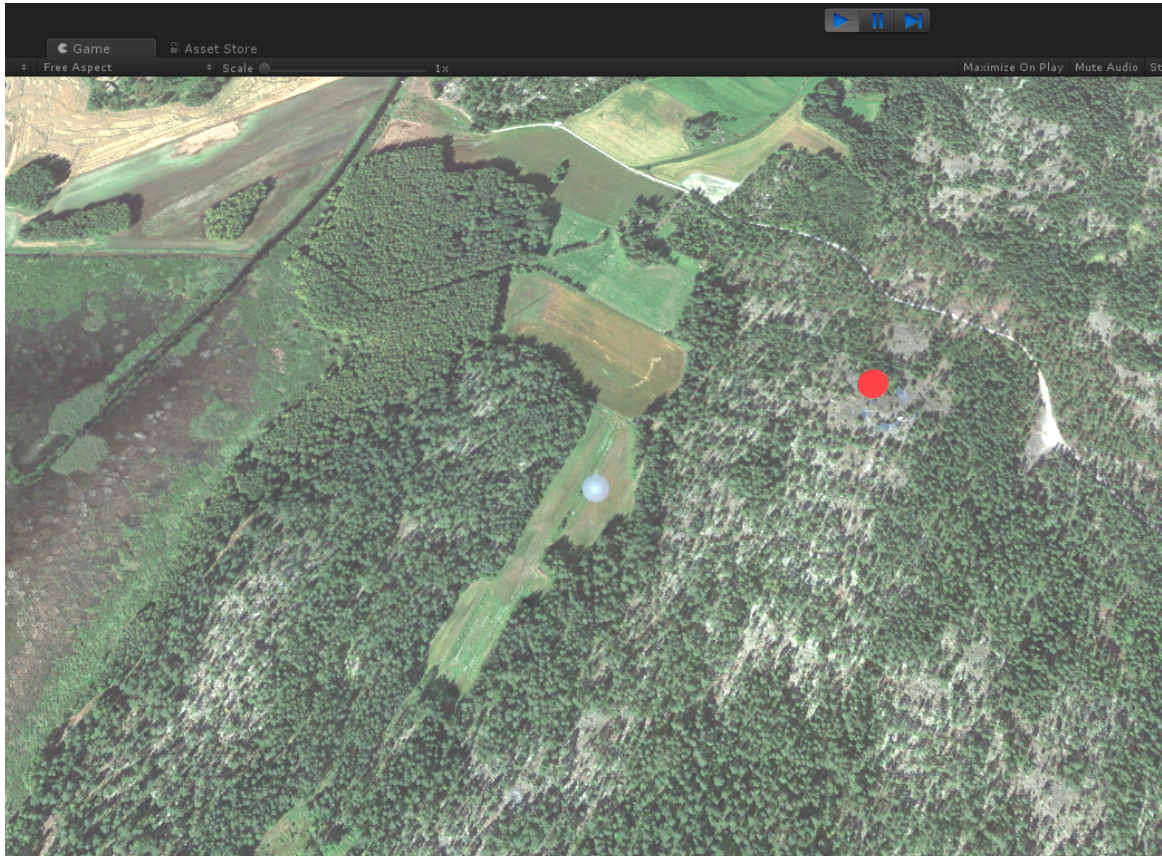


Figure 19: Bird's-eye view over Sammallahtenmäki in Unity editor.

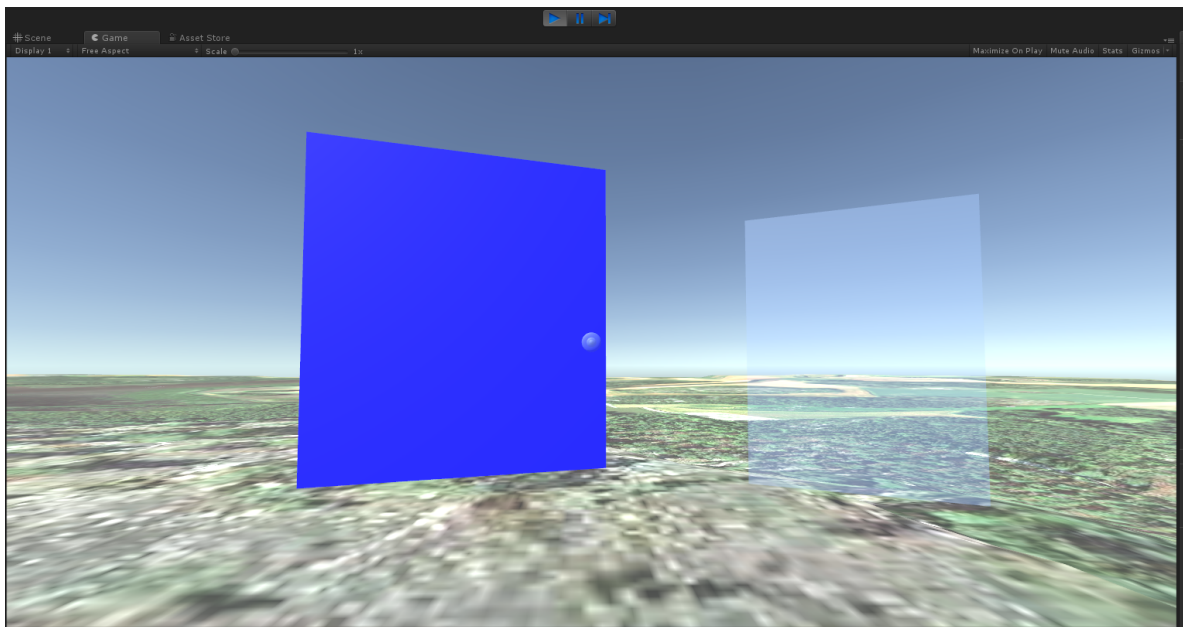


Figure 20: Early prototype with visible trigger colliders for testing Audio Zoom.

Functional realisation of the Audio Zoom effect was simple. Invisible colliders were implemented to all audio source objects (Fig. 20). The crosshair in the user's view triggers these colliders and makes the audio sources sound levels behave as predefined. When a collider is triggered the audio source automatically increases in volume to a pre-set threshold. The longer the user views a sound source, the louder that specific sound cue increases until it has reached the maximum predefined level. When the collider is not triggered the sound level decreases slowly until the original audio volume setting is reached (Tab. 6, p. 60). This audio effect allows users to focus and 'zoom in' on desired audio elements within the virtual environment.

4.2 Music and Audio Content

In this section the research behind the music, musical instrumentation, lyrics and other sound content for the Virtual Heritage prototype's audio scene is clarified. The sources of inspiration for music, and the reasons for choosing certain instruments are explained, as well as the process of producing the material.

4.2.1 Music and Instrumentation

As the theme for the prototype was an ancient funeral ceremony, I felt that the suitable way to narrate a burial rite was to design a musical performance. However, before composing the music, a study on the musical elements of the Nordic Bronze Age was required. The knowledge and assumptions of the conducted research are based mainly on rock paintings and artefacts that have been discovered from as early as 40.000 B.C. Musical content, lyrics and instrumentation that was included in the project are, in the end, based on my speculation on what could have been implemented during the Bronze Age in Sammallahdenmäki.

Instrumentation

The oldest known melodic instrument is a vulture-bone flute discovered in Europe.¹²² This artefact is 40,000-years old and by revealing Paleolithic music making it shows that melody

¹²² Bruce Bower, "Humans: Oldest Known Instruments Found: Ivory, Bone Flutes Date from 35,000 to 40,000 Years Ago." *Science News* 176, no. 2: 13, accessed February 18, 2018, <https://onlinelibrary.wiley.com/doi/abs/10.1002/scin.5591760214>.

and instruments have existed long before the Bronze Age.¹²³ Some studies suggest that: ‘musicality originated with the body alone, that instrumental play came after singing’.¹²⁴ Singing is said to be so basic to man that the origins of vocal music have been lost to history long before the era of written documentation. There is not a single culture without some kind of a singing tradition even if the culture is isolated and primitive.¹²⁵ Religion, social activities and entertainment have often been linked to vocal performances. Thus, we may assume that vocal music also existed in Sammallahdenmäki area in 1500-500 B.C. Therefore, vocal music is one element included in the auditory experience of the prototype.

In Finland, the oldest mention of a percussion instrument is in the anonymous *Historia Norvegiæ* from the late 12th century.¹²⁶ There are approximately 70 ancient Sami drum findings from the Scandinavian area but only three of them from the area of Finnish Lapland. Nevertheless, there are other mentions suggesting that frame drums were more commonly used by Sami people in Finland before the era of written documentation.¹²⁷ Although there is no actual proof of Sami people using drums before the 12th century, there are cave paintings found from Siberia (Karakol Valley) proving that similar frame drums were in use during the Bronze Age.¹²⁸ Furthermore, it is known that these similar percussion instruments with only one drum head have been spread over parts of Europe as well.¹²⁹ Thus, it is probable that frame drums could have also been used in Sammallahdenmäki during the Bronze Age.

Because there is no factual evidence of musical instruments in Sammallahdenmäki from the Bronze Age, I decided to use frame drums and vocals in the creation of the funeral song. Both of the instruments have existed in the Bronze Age, so in that sense they could have existed and also used in Sammallahdenmäki.

¹²³ James Owen, “Bone Flute Is Oldest Instrument, Study Says.” National Geographic News, accessed February 18, 2018, <https://news.nationalgeographic.com/news/2009/06/090624-bone-flute-oldest-instrument.html>.

¹²⁴ Jonathan De Souza, “Voice and Instrument at the Origins of Music”, *Current Musicology* Vol. 97 (2014), 22, accessed February 24, 2018, https://journals.cdrc.columbia.edu/wp-content/uploads/sites/13/2016/08/CM97_DeSouza.pdf.

¹²⁵ "Antiquity to 1590." A Brief History of Singing, accessed February 14, 2018, <http://www2.lawrence.edu/fast/KOOPMAJO/antiquity.html>.

¹²⁶ Juha Pentikäinen, *Saamelaiset: Pohjoisen Kansan Mytologia* (Helsinki: Suomalaisen Kirjallisuuden Seura, 1995), 164.

¹²⁷ *Ibid.*, 168.

¹²⁸ Andrzej Rozwadowski, "Historic and proto-historic shamanic rock art in Siberia: a view from the Altai," In *proto-historical and Historic Rock Art in the Iberian Peninsula International Conference* (2014) Castellón, Spain. 202, accessed February 14, 2018, http://repositori.uji.es/xmlui/bitstream/handle/10234/156585/Quaderns_33_12.pdf?sequence=1.

¹²⁹ Curt Sachs, *The History of Musical Instruments* (New York: Norton & Company, 1940), 33, https://archive.org/details/the_history_of_musical_instruments_curt_sachs.

Lyrics

After the decision to use vocal music in the project, an investigation into the ancient Finnish language was needed to make the writing of lyrics possible. In the light of archaeological findings there have been inhabitants in ancient Finland for approximately 9000 years but it is uncertain what kind of language those ancient native inhabitants spoke.¹³⁰ However, etymological studies have proved that some parts of ancient proto-Finnic language can be still found in modern Finnish.

According to *Nykysuomen sanavarat* written by Jouko Vesikansa and *Mistä Sanat Tulevat: suomalaista etymologiaa* written by Kaisa Häkkinen, there are modern Finnish words that have their origin from back in distant history. From these words, I picked those that could be related to burial ceremonies. Words like “me, pala, sala, tuli” (us, piece, secret, fire) were quite possibly developed originally over 9000 years ago.¹³¹ Other words that belong to the original proto-Finnic language according to Häkkinen are: kuu, maa¹³², kuolla, ei¹³³, jää, kivi¹³⁴, nähdä¹³⁵, hän and olla¹³⁶ (moon, ground, to die, no, ice, stone, to see, him and to be). These words provided a starting point to write the lyrical content.

Because of the uncertainty of Finnish language in the Bronze Age, it was decided to mix these linguistic discoveries (proto-Finnic words) with gibberish. There were two main sources of inspiration for this process. The first was a theatre play called *Del Destierro* that my thesis advisor Professor Lily Diaz-Kommonen introduced me to. *Del Destierro* is an improvised act directed by Juan Carlos De Petre, where the actors use gibberish language to express their emotions.¹³⁷ Second was an invented language called *Vonlenska* used by band Sigur Ros in their album “()”. *Vonlenska* is a form of gibberish language that doesn’t have an actual vocabulary or a grammar but it fits beautifully with the music.¹³⁸ After combining proto-Finnic words with gibberish, I wrote the following lyrics:

¹³⁰ Jouko Vesikansa, *Nykysuomen Sanavarat* (Helsinki: WSOY, 1989), 7.

¹³¹ *Ibid.*, 13.

¹³² Kaisa Häkkinen, *Mistä Sanat Tulevat: Suomalaista Etymologiaa* (Helsinki: Suomalaisen Kirjallisuuden Seura, 2006), 170.

¹³³ *Ibid.*, 172.

¹³⁴ *Ibid.*, 178.

¹³⁵ *Ibid.*, 183.

¹³⁶ *Ibid.*, 184.

¹³⁷ TeatroAltosf. "DEL DESTIERRO..." YouTube, last modified June 16, 2010, https://www.youtube.com/watch?v=_xYAtJXXyiY.

¹³⁸ "Sigur Rós - Frequently Asked Questions." *Return to the Front Page*, accessed November 22, 2017, <https://sigur-ros.co.uk/band/faq.php#08>.

“Hal-laaa, nie-toi-ne ka-laa
Mei ol-laa ne-jaa kuu,
ki-vil-lää-nää ki-vil-lää-nää-kuu

Hal-laaaa, näh-joon en-ei ook-kaan.
Mei ol-laa pa-la - maa,
kuo-lol-laa-hän oon jää-rän-nää maa”

Composition

The musical documentation from the Nordic Bronze Age is understandably non-existent. Therefore, the musical content of the prototype relied more on my musical training and intuition than researched material. However, documentations and notes about ancient Nordic music were explored and some findings set the framework for the artistic approach.

The funeral song is divided into two sections following a structure: chorus - verse - chorus (Fig. 21). The music is composed in the key of C minor and it follows the time signature of 3/4. The rhythmic grounding is based on simple frame drum hits on the quarter beats. The drum beat forms a connection between the ‘viking style’ choir music of chorus, and improvised shamanistic performance of verse, making the both sections sound coherent.

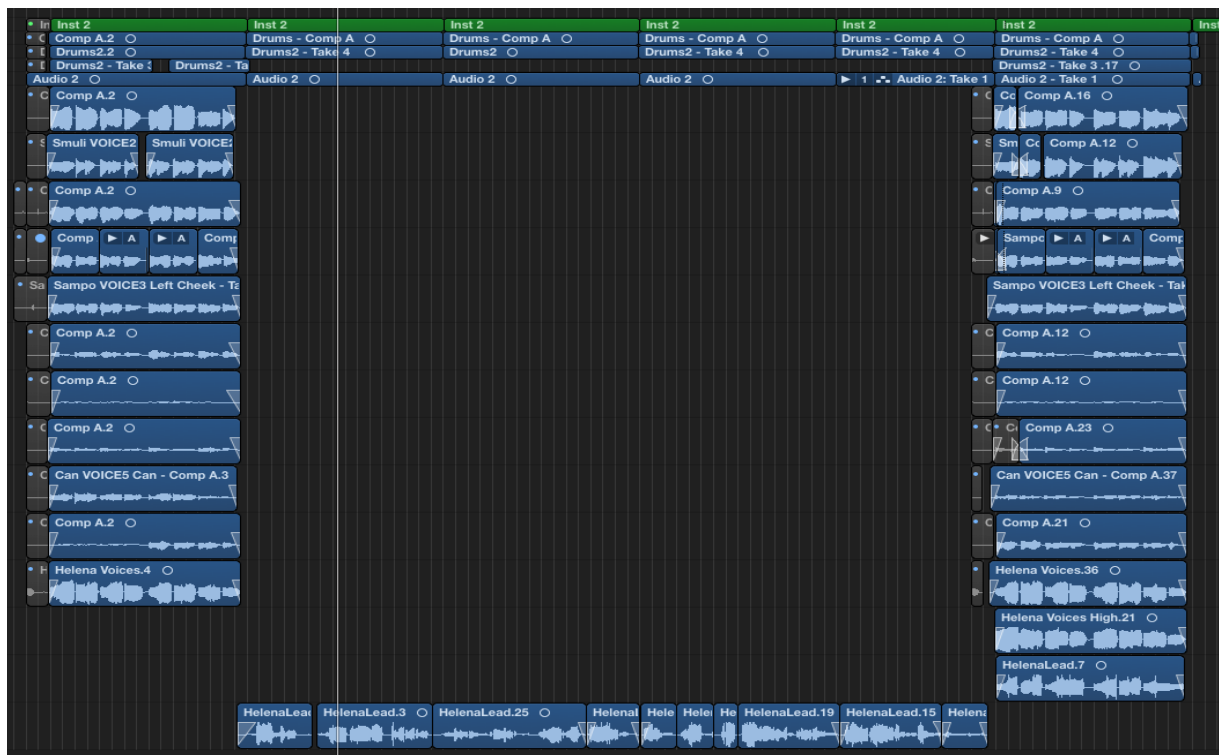


Figure 21: The chorus-verse-chorus arrangement of the vocal choir and drums.

There were two sources influencing the composition of the 'viking style' chorus section. The first is a song called *Helvegen* which is modern style Viking music performed by a collaboration of three music groups: Wardruna, Aurora and Oslo Fagottkor.¹³⁹ The second source is the oldest notation of traditional music from Scandinavia dating from 1300. It was found from the manuscript called *Codex Runicus*.¹⁴⁰ The music decoded from this manuscript has been performed by an unknown music group in a youtube video that served as a source of inspiration.¹⁴¹

The melody (Fig. 22) for the chorus section of the song was composed for a group of unprofessional singers, therefore the melody was kept fairly simple and easy to learn. There is no lyrics on this part as it would have been impractical and complicated to make each vocalist pronounce the unconventional gibberish sentences consistently. Thus, simple humming sound was used instead.



Figure 22: Melody for humming choir in the key of C minor.

In Sami culture there is no written language, therefore there are no references to the origins of traditional Sami singing style: joik.¹⁴² The thought behind joik is that it is often spiritual and there are shamanistic elements included.¹⁴³ It has been assumed that the singing style before the era of written documentation was less structured and more improvisatory and individualistic.¹⁴⁴ Inspired by these findings, the verse section of the funeral song was performed completely without a form as a totally improvisatory performance. Only previously introduced lyrics were guiding the vocalist.

¹³⁹ "Wardruna, Aurora and Oslo Fagottkor: 'HELVEGEN.'" YouTube, last modified 8 July 2017, www.youtube.com/watch?v=cg0TQyjdHJ0.

¹⁴⁰ "Codex Runicus." Wikipedia, last modified August 10, 2017, https://en.wikipedia.org/wiki/Codex_Runicus#Content_of_the_Codex.

¹⁴¹ "Listen to The World's Oldest Known Secular Norse Song." Realm of History, last modified October 13, 2017, <https://www.realmofhistory.com/2017/08/02/oldest-norse-song-codex-runicus/>.

¹⁴² "Joik." Wikipedia, last modified March 19, 2018, https://en.wikipedia.org/wiki/Joik#cite_note-2.

¹⁴³ "A Brief Introduction to Traditional Sami Song and the Modern Music." Same Etnam, accessed January 5, 2018, <http://boreale.konto.itv.se/smusic.htm>.

¹⁴⁴ "Antiquity To 1590." A Brief History of Singing, accessed February 14, 2018, <http://www2.lawrence.edu/fast/KOOPMAJO/antiquity.html>.

4.2.3 Music Production

As introduced in the Audio Model (Fig. 2, p. 5), emotional content is one of the key attributes in the Virtual Heritage concept. The objective of the funeral song was to create an atmosphere with a strong emotional mood of valediction in it. Therefore, the emotional performance was prioritised over the technical elements in the audio production. Six singers were recorded for the hummed chorus section of the song (Fig. 22). The vocalists for the chorus were chosen from people that I am familiar with. I did not want to use professional singers as such, but just somebody capable of singing in fairly good tune. In this way the hymn choir imitated ordinary inhabitants of ancient Sammallahdenmäki while still maintaining the musical coherence within a short recording session. I wanted the singers to get into the proper emotional mood instead of trying to execute a technically perfect musical performance. Therefore, the recording sessions with vocalists included profound conversations about burial ceremonies and Sammallahdenmäki itself.

The vocalist on the verse section was a professional singer Helena Haaparanta. Helena was instructed to forget traditional western music forms, and to enchant a spiritual story with a melody. Before the vocal recording, Sami Grand Prix 2011 performances were previewed from online.¹⁴⁵ This joik music served as an influence for Helena's performance that is completely improvised, has no structure, and has an emotional spirit that I was searching for. The performance was guided by the written lyrics discussed before. However, Helena was allowed to come up with her own gibberish phrases as long as it suited her improvised melody and followed the phonetic tone of the written lyrics. In the end, the original lyrics were only vaguely followed.

All the singers sang separately and each isolated performance was recorded on a single audio track with singers face straight towards the microphone capsule (Fig. 23). However, rather an unconventional method was used as an alternative. One choir singer (audio source nro: 15 in Fig. 28, p. 60) was recorded sideways; microphone pointing towards left cheek (Fig. 24). This technique produced vocals with a different timbre. It can be assumed that not all funeral singers around the listener are facing towards him/her but towards the grave instead. Thus, the timbre of vocals can also be coloured according to the orientation of performers.

¹⁴⁵ SÁMI GRAND PRIX 2011 | All the yoiks, YouTube, accessed February 16, 2018, www.youtube.com/watch?v=cg0TQyjdHJ0.



Figure 23: Direct vocal recording.



Figure 24: Vocal recording face sideways.



Figure 25: Frame drum recording.

A frame drum was recorded three times to be used as a one drum group. I played the frame drum myself using Schlagwer RTS61 (Fig. 25) which has natural skin and fairly similar construction to Sami drums. The recorded frame drum group sounded thin in lower frequencies, so it was duplicated with an additional track of drum samples to thicken and solidify the sound. EastWest Stormdrum library patch “Chinese Lion Drum Hits” was used for this enhancement.

Before exporting audio files from Logic Pro X the audio content was processed with light equalising to remove the lowest frequencies¹⁴⁶ and tame highest sibilance¹⁴⁷ to make the content better balanced in frequency response. A ‘brickwall’ limiting was used in order to control the dynamics and to make the audio files as loud as possible. Renaissance Equalizer, C4 Multiband Parametric Processor and L1 limiter plug-ins of Waves Audio Ltd. were used for these processes.

¹⁴⁶ Renaissance Equalizer: Low cut at 85 Hz and -4dB at 210 Hz with wide bell curve.

¹⁴⁷ C4 Multiband: high frequency compression above 9 kHz.

All the music recordings were done in a small vocal booth in a project studio in Herttoniemi, Helsinki during February 2018. For all of these recordings a Neumann 103 condenser microphone was used. Focusrite Sapphire Pro 40 worked as a sound card. For musical recordings a Logic Pro X digital audio workstation was used. All the performers agreed on Creative Commons Licence (BY-ND) to avoid problematic situations with copyright if the prototype is showcased publicly.¹⁴⁸

4.2.4 Other Audio Content

Although the focus of the audio experience is more in the musical content, that alone is not sufficient to create the auditory sensation of the location and space. Thus, other sounds were needed to represent the sonic environment of the Sammallahdenmäki location.

The climate was different in 1500-680 B.C. The average temperature was warmer and resembled that of modern France, meaning that the fauna and flora were different as well.¹⁴⁹ Because of this, the environment had to sound similar to modern France. Therefore, two different field recordings of French forests from Sound Ideas: General 6000 sound library were used to represent the sounds of nature: birds singing in the forest meadow.

As mentioned in 4.1.1, Sammallahdenmäki used to be located next to the sea.¹⁵⁰ Therefore, the sounds of sea waves were included in the sonic environment. A combination of four audio files from the General 6000 sound library were used to create the whole maritime atmosphere. Two of the audio files have a huge and deep sound of ocean roaring and the other two audio files have lighter sounds of waves hitting the shore. Furthermore, a subtle sound of waves rocking a wooden boat was added to underline the trading customs practiced in Bronze Age. This sound was from my own field recording from Jollas, Helsinki, captured in 2015.

The findings of burned human bones refer to cremation being a common custom in Sammallahdenmäki during the Bronze Age.¹⁵¹ Therefore, a sound of a funeral pyre was justifiable for the burial ceremony in Sammallahdenmäki. This sound is constructed from four different audio

¹⁴⁸ "About the Licenses." *Creative Commons*, accessed February 23, 2018, <https://creativecommons.org/licenses/>.

¹⁴⁹ "Nordic Bronze Age." *Wikipedia*, last modified January 22, 2018, https://en.wikipedia.org/wiki/Nordic_Bronze_Age#Climate.

¹⁵⁰ Uusi-Seppä, *Satakunnan Kulttuuriympäristöt*, 34.

¹⁵¹ *Ibid.*, 32.

files. I used a combination of two audio files of campfire sounds from my own field recordings (recorded in 2011 in Kulosaari, Helsinki), a forest fire sound from the General 6000 sound library and a bonfire sound bought from <https://stockmusic.com/>.

The reason for having several audio files for the funeral pyre, ocean and nature sounds was that it made it possible to spread the audio sources to a wider area, instead of having only a one spot emitting the sounds. This created an effect of being surrounded by a detailed and sensorially rich sonic environment. As discussed in section 2.2.2, a complex sound design has a connection to immersion.

4.3 Audio Implementation

The online test procedure (section 3.2) indicated that dearVR was the best spatializer to be implemented in the prototype. The spatial audio rendering of spatializer plug-ins works automatically when set properly in the Unity's audio manager settings (Fig. 26). A sample rate of 48 kHz was used in the prototype and in all related audio files.

During the first instance of implementing audio into the Unity project, it became apparent that when a spatializer plug-in was utilised the overall sound levels dropped drastically. Therefore, all the audio files used were treated with Waves L1 limiter plug-in and normalized to -0.1 dB (in Logic Pro) to maximize the overall loudness.

The objective of the audio design was to create a feeling of 'being there' in the centre of all the action, witnessing the auditory event of a burial ceremony. The listener is placed near to the burning funeral pyre while drummers and singers are located in a half-circle around her/him, facing towards the fire. The maritime sound sources are located further to the north side of the listener, and environmental forestry sounds are emitted from east and west approximately 50 Unity units (meters) away. In practice, all the sound objects are placed around the listener (camera) component in Unity editor. Audio source locations can be reviewed in Figures 27 and 28 (p. 59 and 60).

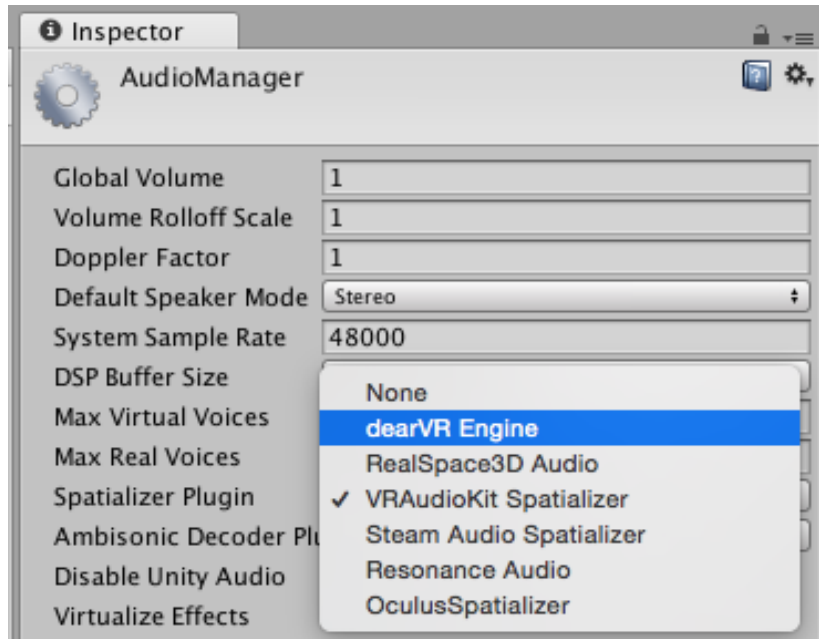


Figure 26: DearVR was implemented into the prototype.

The dearVR plug-in allows the utilisation of Unity's proprietary audio parameters by enabling the *Unity distance graph* property from dearVR source component (Appendix D, Fig. 43, p.81). This makes it possible to use Unity's audio source component's (Appendix D, Fig. 42, p.80) parameters to control distance, volume and reverb levels, for instance. These parameters can also be controlled from the dearVR source, but it was more convenient to rely on the controls of an audio source component that I was familiar with.

DearVR has some other features worth mentioning, such as occlusion and obstruction simulations, but those were disabled in the audio scene as the virtual environment takes place in spacious forest environment with no acoustic barriers close to the listener. Furthermore, a bass boost property of dearVR was utilised for audio sources in the drum group to enhance the bass frequencies of the musical beat. Unity's proprietary Doppler effect was disabled from all sound sources because it sounded unnatural.

The Table 6 (p. 60) reveals the audio parameter setups used for each audio object. Minimum and maximum distance parameters were controlled by the logarithmic sound level curve, also known as volume rolloff. Within the minimum distance the sound will play at the volume set for the audio source. When the distance is outside the minimum distance value the sound begins to attenuate according to the volume rolloff. Maximum distance indicates the range at which the sound level stops attenuating. This minimum sound level can be set from the sound

level curve and it is set to value 0.01 to all audio sources (Unity's default setting for audio component).¹⁵²

I approach the audio design of a complete soundscape as if it were a painting. Sound location placement in 3D environment could be compared to the creation of a large spherical picture being built from smaller graphical pieces. Even if the image is created from separate pieces, everything needs to feel like it fits together to form a complete and unified picture. As mentioned (in subsection 2.4.4) the basic attributes for setting the distance is sound level and reverberation. It is known that the proper use of these attributes can increase immersion, thus careful attention was paid to these properties.¹⁵³ The audio implementation requires understanding of cognitive associations for each of the sound sources when it comes to creating auditory distance cues. There are narrative and emotional objectives when setting distances for audio cues.¹⁵⁴ Thus, the final settings for distances and reverb amounts were based on my sound engineering skills and artistic intuition.

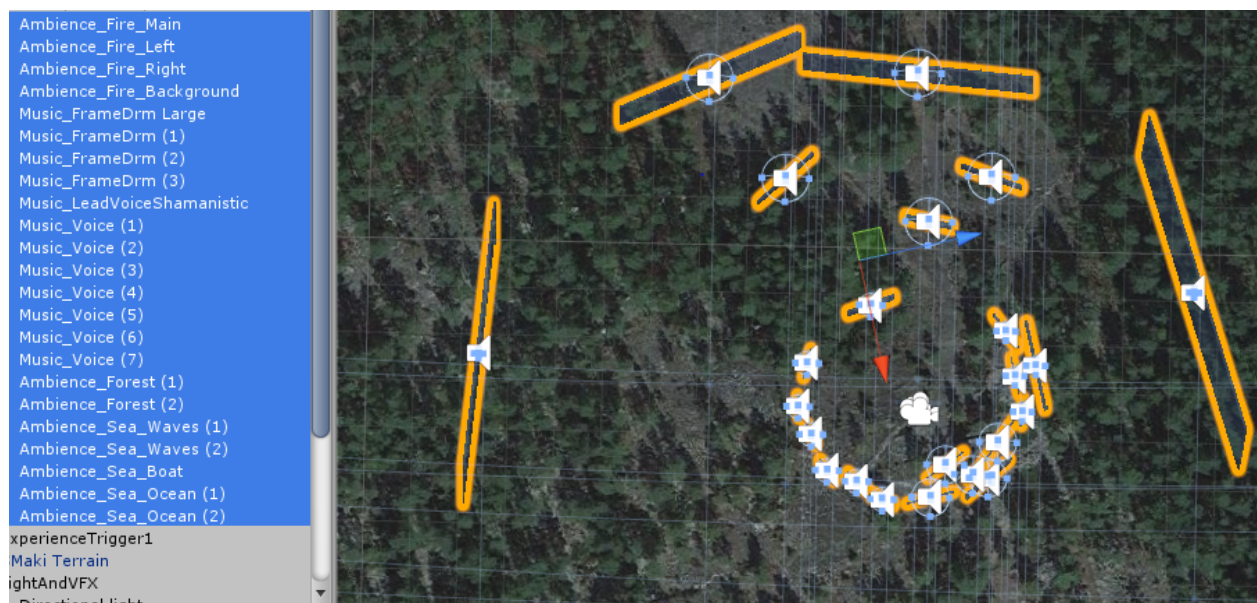


Figure 27: Audio Source locations in Unity Editor

¹⁵² Technologies, Unity. "Audio Source Overview." Unity - Manual: Audio Source, accessed March 22, 2018, <https://docs.unity3d.com/Manual/class-AudioSource.html>.

¹⁵³ Rossing, Moore, and Wheeler, *The Science of Sound*, 69.

¹⁵⁴ Roginska, and Geluso, *Immersive Sound*, chap. 1, doc. 5.

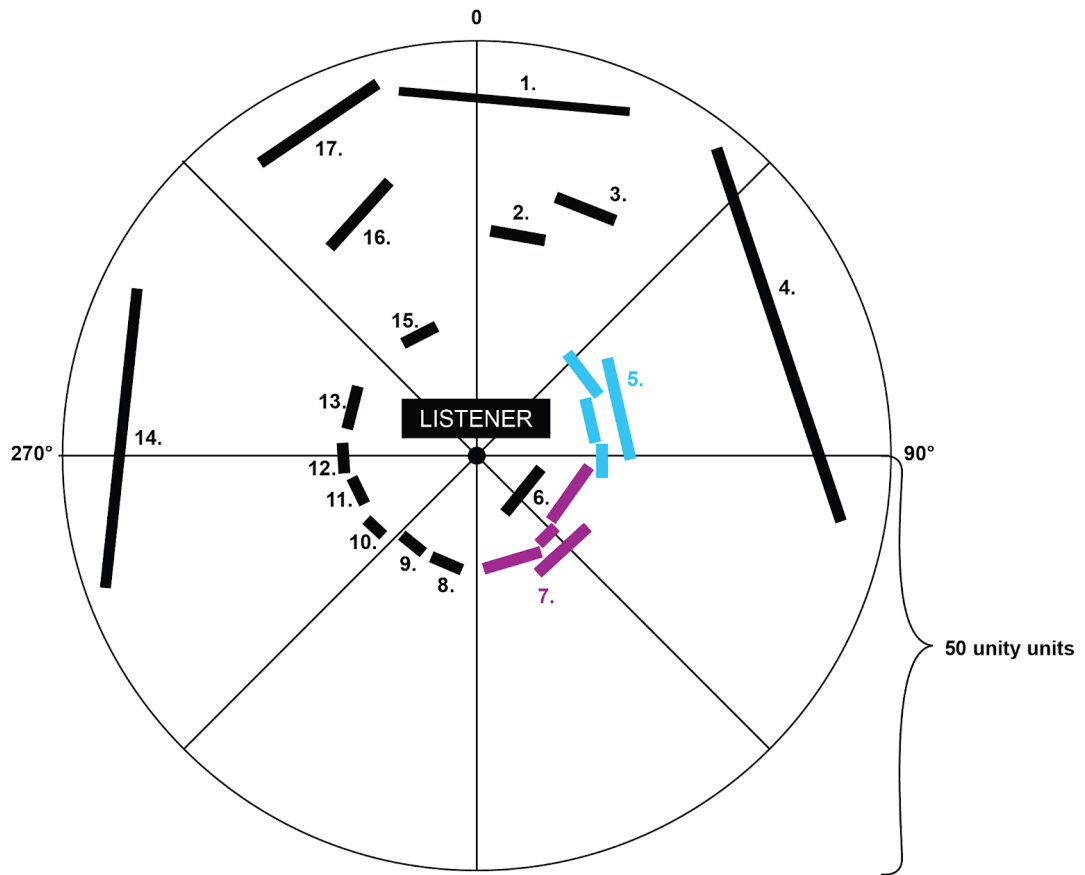


Figure 28: Info architecture of audio source locations.

Nro:	Name:	Volume:	Reverb Mix:	Min/Max Distance:	Loop:
1	Ambience_Sea_Ocean (1)	0.09	0.5	4/500	On
2	Ambience_Sea_Boat	0.09	0.6	5/500	On
3	Ambience_Sea_Waves (1)	0.09	0.6	4/500	On
4	Ambience_Forest (1)	0.1	0.1	1/500	On
5	Bonfire Group	0.25	0.2	1.5/500	On
6	Music_LeadVoiceShamanistic	0.25	1	2.5/500	Off
7	Frame Drums Group	0.35	0.3	2/500	Off
8-13	Music_Voice (1-4,6,7)	0.2	0.8	2.5/500	Off
14	Ambience_Forest (2)	0.1	0.1	1/500	On
15	Music_Voice (5)	0.2	0.75	2.5/500	Off
16	Ambience_Sea_Waves (2)	0.09	0.6	4/500	On
17	Ambience_Sea_Ocean (2)	0.09	0.5	4/500	On

Table 6: The settings for audio parameters in audio source components.

In practice, all audio objects were placed in the virtual environment in relation to the listener object by trial and error. Audio sources were moved in the virtual environment until a believable location for the sound was found. The challenge in this process was my close miked vocal and drum recordings. These audio files have a timbre of a sound being emitted from very near to the listener. Therefore, these audio sources seemed as if they were placed at a shorter distance within the virtual environment. In order to balance this, extra distance was added to audio objects and reverberant-to-direct sound ratio (as explained in 2.2.4) was set to enhance the auditory sensation of distance.

The acoustic space and distances were enhanced by using two different reverb presets of dearVR: Outdoor_Alley and String_Plate (Fig. 29). Neither of these reverbs sounded like a forest meadow in spatial sense but they were the best fitting presets available. The reverb parameters were set to enhance the sense of distance between the listener and audio sources, but also to 'sweeten' the musical content. In the end, the audio source reverberation for musical content sounded dreamlike, which fit well with the spiritual mindset of the audioscene. A low pass filter was used for both of the reverbs to 'roll off' the highest audio frequencies (Fig. 29). This resulted in the reverbs sounding more transparent and blending better with all sound sources, 'gluing' together a unified soundscape.

The Audio Zoom property (discussed in subsection 4.1.5) was set to increase the sound levels to the maximum in less than two seconds in all audio sources, fading back to the initial volume setting (Tab. 6) in approximately 0.2 value difference per second. This effect made it possible to focus on a specific singer from the choir or to listen sounds of ocean with a distant singing in the background, thus creating variation and vividness to the sonic environment.

When audio sources are spatialized in Unity, the loudness of audio sources only attenuate from the original volume. Therefore, to balance this decrease in loudness, all audio sources were amplified in the audio mixer by 20 dB (Fig. 30).

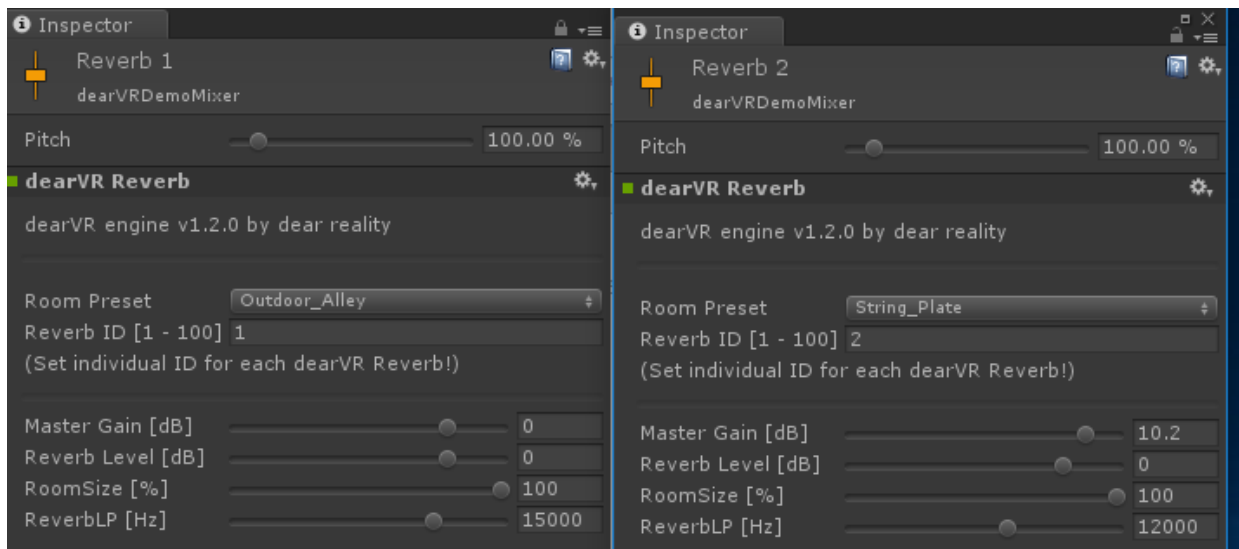


Figure 29: DearVR reverb component settings.

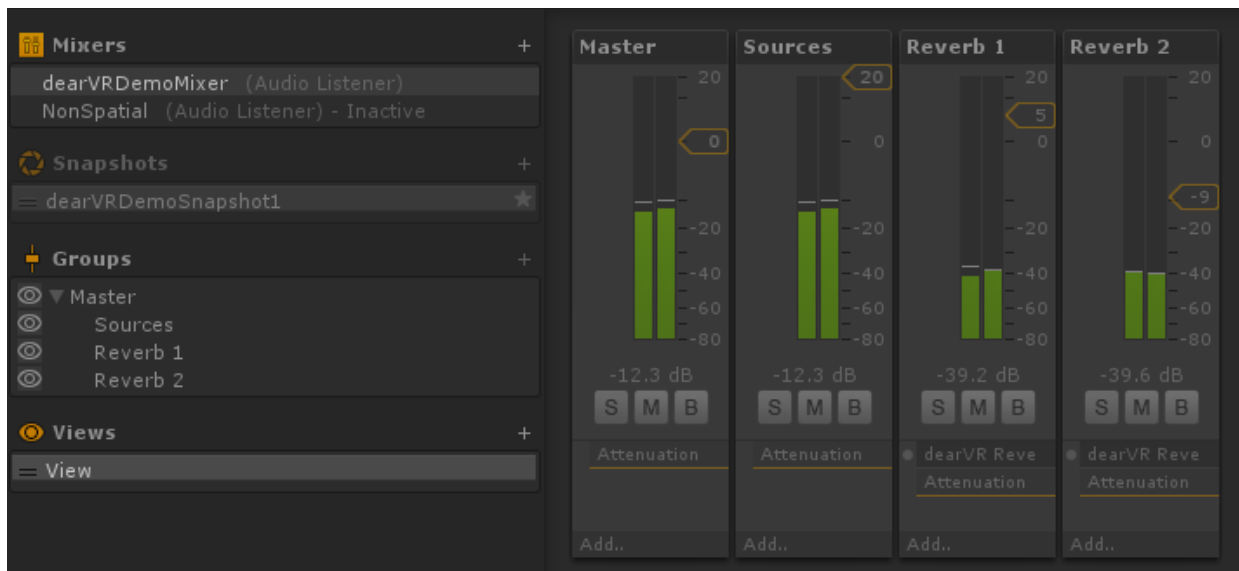


Figure 30: Unity mixer settings.

4.4 Final Discussion

I recall my colleague asking me, about year ago, which spatializer I prefer to use for VR projects in Unity. During that time, I was not sure how to answer as I was only familiar with Unity's default spatializer: Oculus. Now, after the Virtual Heritage project, I could give a long elaborate lecture to answer that question. It could be said that during the process of working on the Virtual Heritage project I have discovered and learned more than I was expecting to. In this subsection these findings are discussed.

In the field of sound design there is a common idiom: 'dialogue is king', meaning that whatever happens in the audio scene the dialogue has to always be distinct and easy to follow. In virtual

environments 3D audio representations with plentiful overlapping audio cues may potentially lead to chaotic and incoherent results, with difficulties in focussing one's auditory attention on the essential, such as dialogue. When developing the Audio Zoom feature it was thought to be a pleasant effect for the users to engage and perform with. However, when performing with the feature myself I realised that it could be used as a method to emphasize the relevant auditory information. In the case of incoherent sound environment 'zooming in' would make it convenient to focus into the most interesting and essential audio cues. I could see the Audio Zoom feature having the potential to be used as a practical method of controlling sound levels in any virtual environment with several simultaneously playing audio cues.

When localizing the musical audio sources into the virtual space of Unity, there were difficulties in setting the desired distance perspective between the source and the listener. This was because the sounds have a timbre of being emitted from very near to the listener; since the vocal and drum recordings were made only 10 cm away from the microphone. The resulting spatial audio effect (or the lack of it) made me realise that in order to create convincing spatial audio design and representations the timbre of sound sources should be genuine from the start. This means that optimal audio content should be recorded with authentic distance and perspective.

The improvement in the overall quality of 3D audio, when switching from Unity's default spatializer Oculus to dearVR, was significant. I personally considered dearVR to be the best spatializer in terms of precise audio spatialization, from the first time that I tried all the spatializers. However, as discussed, the perception of 3D audio is individual and thus it is impossible to make a subjective conclusion on which plug-in works for the majority of people. Now, after conducting the spatializer test online, it can be stated with confidence that dearVR is the best 3D audio solution out of those mentioned in this thesis. Nevertheless, it must be added that as long as generic HRTFs are used in spatializers they are still far from being realistic and this is likely to be the case for a long time, at least until a commercially viable and practical solution for individual HRTFs is found. Conducting the online test was a successful experience in other ways as well. As I contacted all of the test participants individually I made several new and useful contacts and received two job offers.

During this project work, a lot of time and energy went into discovering the best possible 3D audio solution in order to create as immersive virtual environment as possible. However, when

talking about the audio production, it is not only the spatial audio rendering that enhances the immersion but the content plays a critical role as well. If thinking about the experience of reading a fantastic book, it is the content that engages and draws the reader into the story, enabling them to lose their sense of time and place as the imagination takes a journey through the story. I believe that with sounds a similar sensation of immersion can be achieved. With an authentic audio design it can be possible to engage the user and make them feel immersed from the believability of the audio content.

After completing the prototype, I realised that there is a clear contradiction between auditory and visual representations. The design idea was to use minimalistic visuals which rely on shadows, darkness and user's imagination. In the prototype's sonic environment a funeral pyre can be heard but lights or flames cannot be seen. In my opinion, this forms a huge conflict between the modalities creating a barrier that blocks user immersion. This was not the initial aim but as often happens if creating content for new media, the result is not quite what you were expecting. Therefore, I will continue developing the prototype further in the future. There were initial discussions with Professor Lily Diaz-Kommonen, about improving the visual content using photogrammetry and lightning effects. However, that will be excluded from this thesis as such a visual upgrade might be executed only after completing this written work.

Chapter 5

Conclusion

While building the prototype of the Virtual Heritage project, I often found myself feeling like a tourist discovering the possibilities and limitations of 3D audio design in a VR context. Regarding 3D audio design, the practicalities do not exist in similar conditions as in conventional media, such as films and games. Therefore, my personal approach to design 3D audio is based on experimentation. The audio positioning in a virtual environment requires, in addition to the spatial audio rendering, careful audio engineering of the content production. Timbre of sound differs by the attributes of placement in its spatial location and therefore, it is recommended that audio content should be recorded with accurate distances and perspectives to make them 'fit' naturally into the sonic environment. The 3D audio design process requires understanding of how humans perceive audio spatialization and what are the key components in audio spatialization, in order to be able to compensate spatial incoherencies in post-production.

Spatial audio theories were explored in this thesis and it became evident that until a commercially viable solution for individual HRTFs is found, all the 3D audio rendering tools remain somewhat compromise solutions. However, during the Virtual Heritage project there was a need to know which of the commercially available audio spatializers, compromises or not, would be the best possible 3D audio solution for immersive virtual environments. Therefore, six different spatializer plug-ins (Oculus, Resonance, Realspace 3D, Steam Audio, 3D Sound Labs and dearVR) were compared in the empirical test procedure. 78 people participated to evaluate the perceived directional precision of the spatializers and the results indicate that dearVR is the most highly rated 3D audio solution.

A secondary test was organized in a closed environment to measure how much the usage of the dearVR spatializer plug-in increases immersion compared to non-spatialized audio. The prototype of Virtual Heritage without visual feedback was used as a test platform. Results indicate that using dearVR increased the perception of 'level of presence' by averagely 36.25% percentage points when tested with university students from the department of media.

The results show that spatialization of audio is extremely important in order to achieve the most immersive experience.

During the project work it became clear that enhancing the immersion is not completely dependent on a 3D audio solution but there are several other means to achieve this. When interactions, complex sound design and believable content work hand in hand with other modalities it generates an emotional response and thus engages users' attention and finally increases the user's immersion.

In my view, 3D audio design utilising spatializers can be thought as a combination of conventional audio design (design for 2D films and games) and spatial audio design for physical spaces. The spatializers reveal a new artistic territory where spatial audio design approach could be taken from physical spaces to virtually simulated worlds. Furthermore, they enable to create narrative audio scenes in virtual environments in a way that is not possible in 2D media content. To conclude, I believe that my work on Virtual Heritage project proves that 3D audio has a significant potential in creating immersive virtual reality experiences in cultural heritage domain.

Appendix

A. Audio Production of the Online Test

In the online test audio production process 13 audio files were recorded: six for the ‘direction test’ and seven for the ‘quality test’. There was a need to have a control over the audio objects’ movements in the ‘quality test’, therefore I decided to use Unity’s feature called Timeline.¹⁵⁵ With this asset, it was possible to create a linear motion for an object with position curves of Timeline editor. The Timeline controlled the object’s movements to emphasise clearly the directions: front, right, back, left, below and above, in this particular order. After being placed in each of those basic directions, the audio object was controlled to spin around the user two times before returning to its initial position in front of the listener. This Timeline animation is shown in Figure 31.

Unity’s distance measurement units referred to meters in the physical system.¹⁵⁶ As the head related impulse responses for HRTF are usually measured at a distance of at least one meter away from the listener, I used the distance of one Unity unit for sound sources in the ‘quality test’.¹⁵⁷ The audio listener component was attached to a main camera object (Fig. 32). This object was positioned to be 1.8 Unity units high in order to resemble the head of a person approximately 1.8 metres tall.

The “direction test” used the same audio listener in the main camera object as the “quality test”, but not the Timeline controlled movable audio source object. Instead, there were six static audio source objects (visible cubes in Fig. 33), two Unity units distance away from the audio listener. During the test audio recording, these sound sources were triggered to play one by one with “Play on Awake” function.

¹⁵⁵ Technologies, Unity. "Timeline Overview." Unity - Manual: Timeline Overview, accessed February 7, 2018, <https://docs.unity3d.com/Manual/TimelineOverview.html>.

¹⁵⁶ Alan Thorn, *Mastering Unity 2017 Game Development with C# - Second Edition* (Birmingham: Packt Publishing, 2017), chap. 9, doc. 6, <https://www.safaribooksonline.com/library/view/mastering-unity-2017/9781788479837/2fab2eb0-81a9-447b-8365-183fef00615.xhtml>.

¹⁵⁷ Roginska and Geluso, *Immersive Sound*, chap. 4, doc. 3.

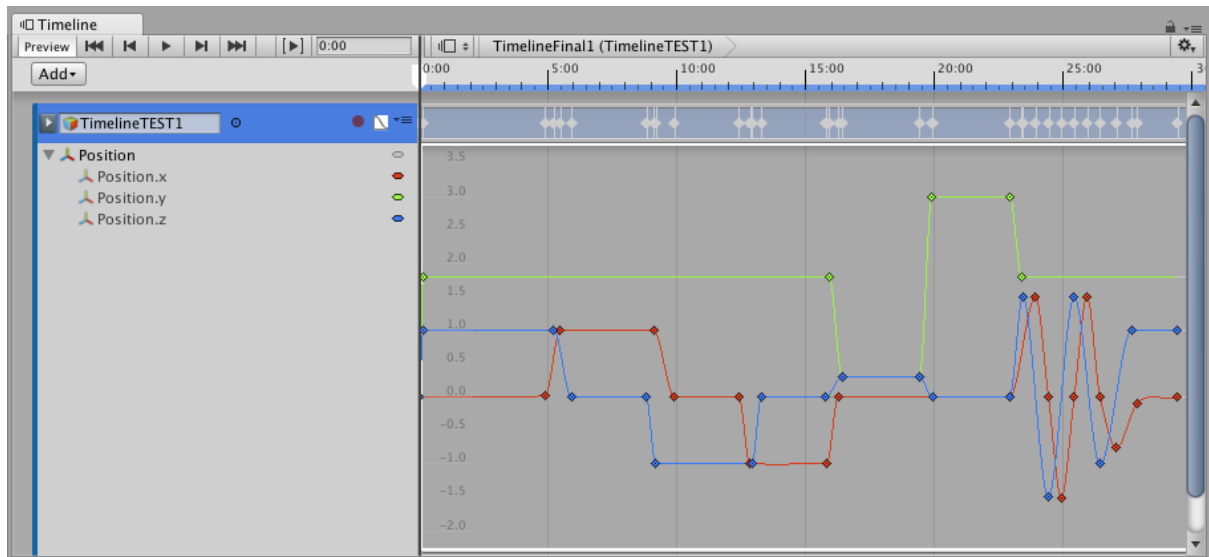


Figure 31: Timeline curves controlling the positional movements of the audio source object.

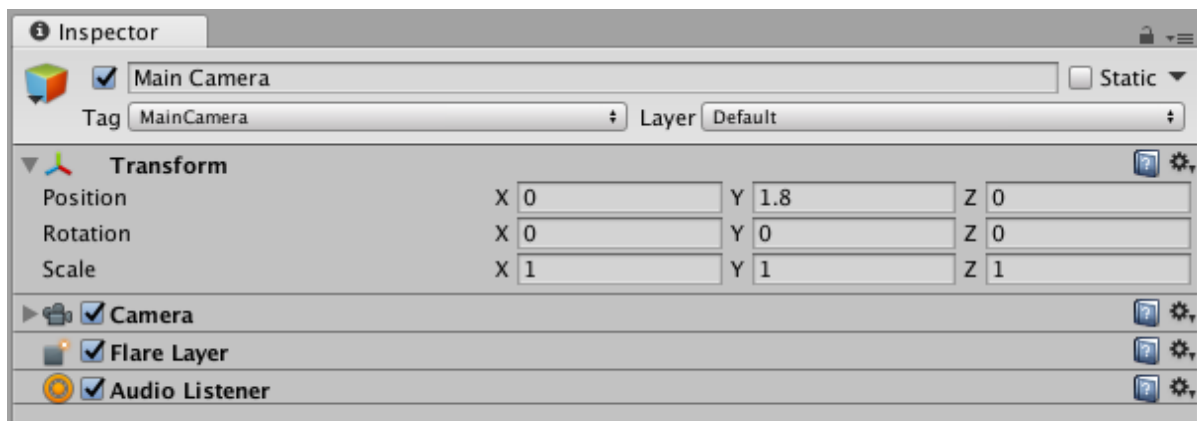


Figure 32: The Audio Listener component was located 1.8 Unity units high in the scene.

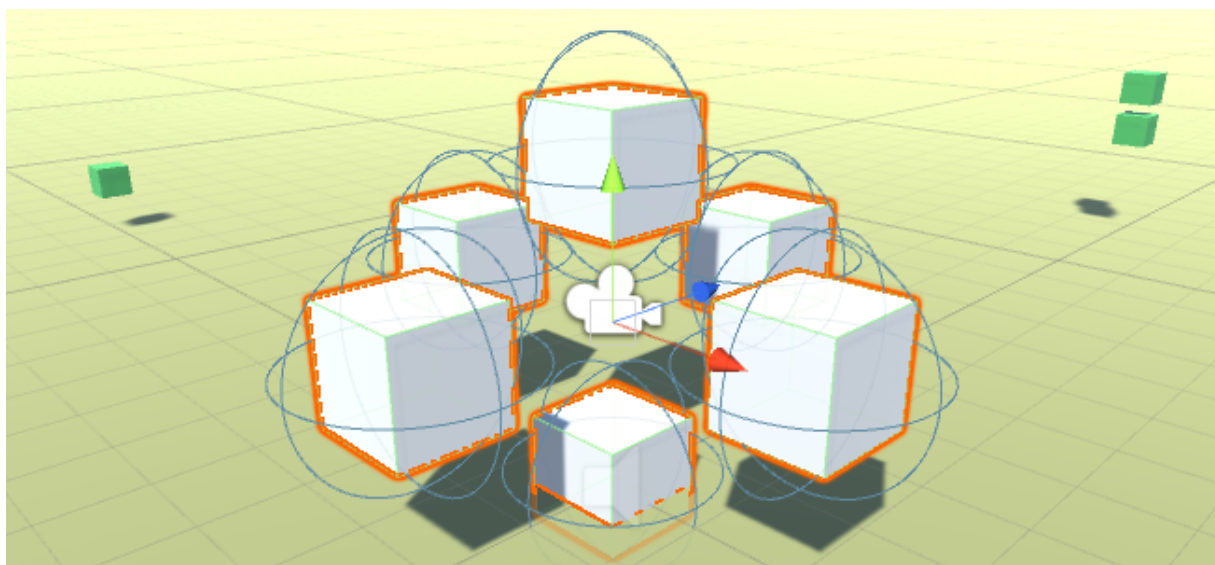


Figure 33: 'Direction test' with static audio source objects in six directions.

The audio content for both test methods needed to be clear and simple to hear.

Human ears are accustomed to hearing human voices and it is known that the hearing range is at its best between 1 kHz and 5 kHz frequencies, where the most significant intelligible information within the human voice exists.¹⁵⁸ Thus, a human voice is a main element for the test audio.

For the moving sound in the “quality test”, an announcement was recorded in which the current location movements of the audio object were explained. The recording was adapted to the audio source object’s movements in the Timeline editor. I announced the following text for the recording: *“This is a 3D audio test, the sound is now in front of you. The sound is now on your right side. The sound is now on behind you. The sound is now on your left side. The sound is now below at your feet. The sound is now above your head. The sound is now moving around. The sound is now back in front of you.”* The recorded announcement was edited with Pro Tools into a single audio file, timed perfectly to sync with the audio object movements.

A looped audio source was added to accompany the voice-over recording. This audio file was consistent and ceaseless (which the voice-over recording was not) so that the movement between the directional announcements could be distinct and easy to localize. It was made of three separate sounds. The first was a high-pitched clicking sound operating in the sensitive area of the human range at around 2 Khz.¹⁵⁹ The second sound was a constant rustle in the higher frequency range of 4 kHz — 14 kHz. The third sound was a synthesized humming operating in the frequency range of around 200 Hz. All the files were mixed and recorded to a mono file in Pro Tools editor and exported to a single loop file.

In the “directional test”, only my own voice was recorded in six audio files, for each audio source object, announcing different location codes: A1, B2, C3, D4, E5 and F6. These recordings were arranged in main directions for each of the plug-ins. The location randomisation was done so that it did not follow any detectable linear pattern. The sound locations were organised according to Table 7.

¹⁵⁸ Dave Swallow, *Live Audio: The Art of Mixing a Show* (Focal Press, 2012), chap.2, doc.1, https://www.safaribooksonline.com/library/view/live-audio-the/9780240816043/xhtml/11_chapter02.xhtml.

¹⁵⁹ Ibid.

Direction code	A1	B2	C3	D4	E5	F6	Main directions
Oculus	F	Ba	L	R	A	Be	F = front
Realspace3D	R	L	Be	Ba	F	A	Ba = back
Resonance	A	L	F	Be	R	Ba	L = left
dearVR	L	Be	A	R	F	Ba	R = right
Steam Audio	Be	R	Ba	L	F	A	A = above
3D SoundLabs	R	F	Be	Ba	L	A	Be = below

Table 7: Randomized sound directions for all the spatializer plug-ins.

The voice-over recordings were done using Neumann 103 condenser microphone and Focusrite Sapphire Pro 40 sound card. All the audio files were recorded and processed in Pro Tools with several audio processing tools in order to enhance the clarity and loudness of the voice-over recordings (Fig. 34). The recording of test audio file was done from the Unity editor. The recording setup was Mbox Pro 3 sound card as an output and Focusrite Sapphire Pro 40 sound card as an input.

Before the actual audio recordings, a thorough calibration for stereo channel balance was executed in order to maintain the correct stereo image. Two identical sine wave oscillations at 1 kHz were generated from the Unity workstation to the recording computer where the signals were analysed. A phase cancellation test was done to ensure that both of the input channels were receiving exactly identical signal balance in Pro Tools. The phase cancellation means that when the audio polarity of one of two identical audio files is inverted, the signals on both channels cancel each other out. The polarity inversion was done with the EQ III plug-in by Avid.

After successful calibration the test audio recordings begun. In the “quality test” there were six recording runs for each of the Timeline controlled setups. In the “directional test” different recordings were executed for each main directions for each of the plug-ins, consisting of 36 different variations of short recordings in total. All 36 “directional test” audio files were combined and arranged to longer test audio files for each spatializer plug-ins according to the Table 7.



Figure 34: Various audio processing tools used for the voice-over recordings.



Figure 35: Audio files and loudness calibration as recommended in the EBU R128 standard.

Each of the spatializer plug-ins outputted differential volume levels and the scale between maximum and minimum loudness (measured as loudness units in full scale) of audio files was as much as 12.1dB.¹⁶⁰ It is known that some people prefer louder sounds over quiet.¹⁶¹ Therefore, a loudness balancing was performed to all recorded audio files to avoid test participants preferring a louder audio. An equal gain adjustment of -23.0 LUFS was set for all the audio files (Fig. 35). This is a broadcast standard EBU R128 for loudness recommendation.¹⁶²

¹⁶⁰ Bob Katz, *iTunes Music* (Burlington, MA: Focal Press, 2013) chap. 5, https://www.safaribooksonline.com/library/view/itunes-music/9780415656856/Text/12_Chapter05.xhtml.

¹⁶¹ "Why Do We like Our Music Loud?" MusicRadar, modified July 13, 2009, <https://www.musicradar.com/tuition/tech/why-do-we-like-our-music-loud-212790>.

¹⁶² "Loudness (PLOUD)." EBU Technology & Innovation, accessed February 8, 2018, <https://tech.ebu.ch/loudness>.

B. Instructions for the Online test

3D audio test

This is a spatial audio test studying the quality and usability of generic HRTF (Head Related Transfer Function/Function that creates binaural audio effect) in commercial Unity plugins. The spatializer plugins included in the test are: Unity default(Oculus), Steam Audio, Resonance, Real Space 3D, Dear VR and 3D Sound Labs.

This study is divided in two tests: **quality** and **directional** test.

Quality test (testing directional and overall quality):

-Listen to the example file. *The sound source follows always similar path as heard in the example.*

(Front, Right, Back, Left, Below, Above, Circling around and Front).

- Listen through the files and rate the quality and precision for left-right, front-back and below-above.

- Rate also the overall quality of the sound sample.

Directional test:

- In all files you will hear six different sound locations.

- Choose the direction as best as you can for each of the locations (1A, 2B, 3C, 4D, 5E, 6F).

Note: This test works only with headphones. Do not use loudspeakers. Make sure that your operating system doesn't have spatial processing on.

(WINDOWS 10: From the **Sound** control panel, select a playback device and click **Properties**. In the page that opens, there is a **Spatial sound** tab.

If the device supports spatial sound, you can select one of the available formats from the dropdown. Choose None.)

When listening to these 3D audio examples it is necessary that you keep your head still because head movements are connected to hearing perception. Hearing 3D sounds as presented during this test is absolutely subjective experience, in other words these example sounds works differently for all participants. For someone they might work brilliantly and for someone very poorly. **So, remember that there are no right or wrong answers.**

Open the test link here and follow the instructions first.

After instructions you can start playing the audio files.

<https://goo.gl/forms/TI47i1bW3YYh6qZB3>

C. Histograms for the Online Test

Quality Test Histograms

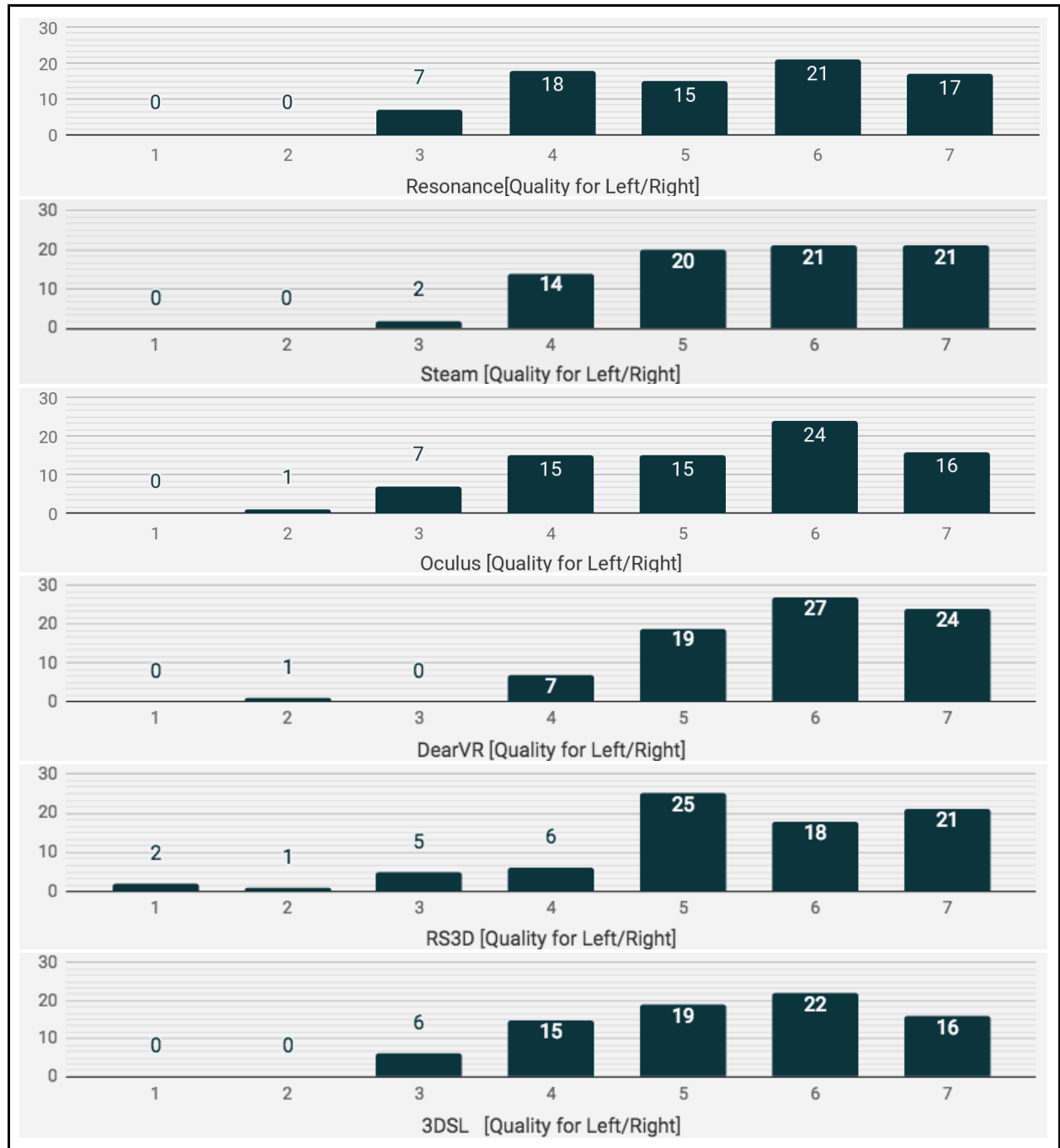


Figure 36: The Likert scale ratio for the quality for left/right as rated by the test subjects.

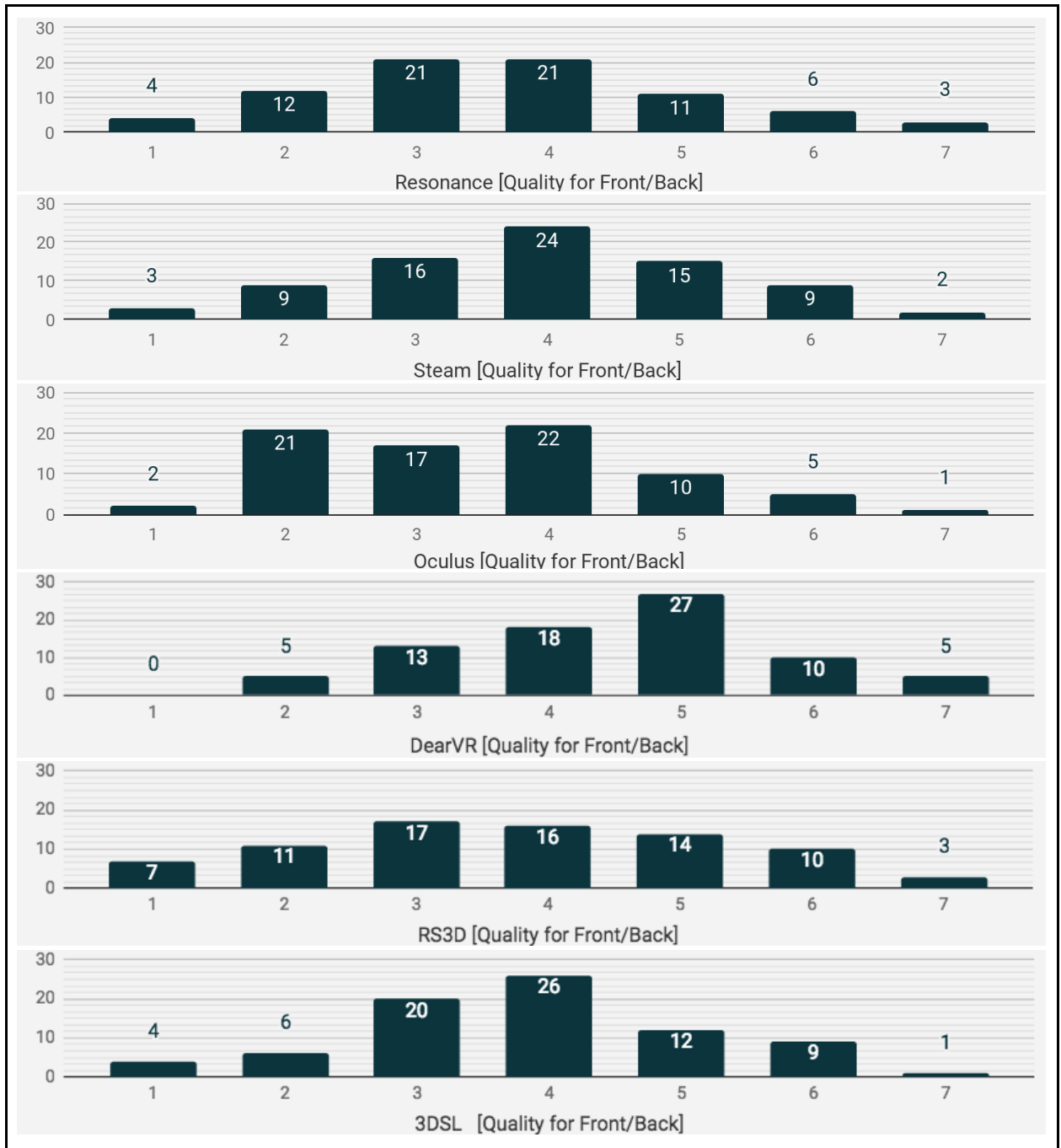


Figure 37: The Likert scale ratio for the quality for front/back as rated by the test subjects.

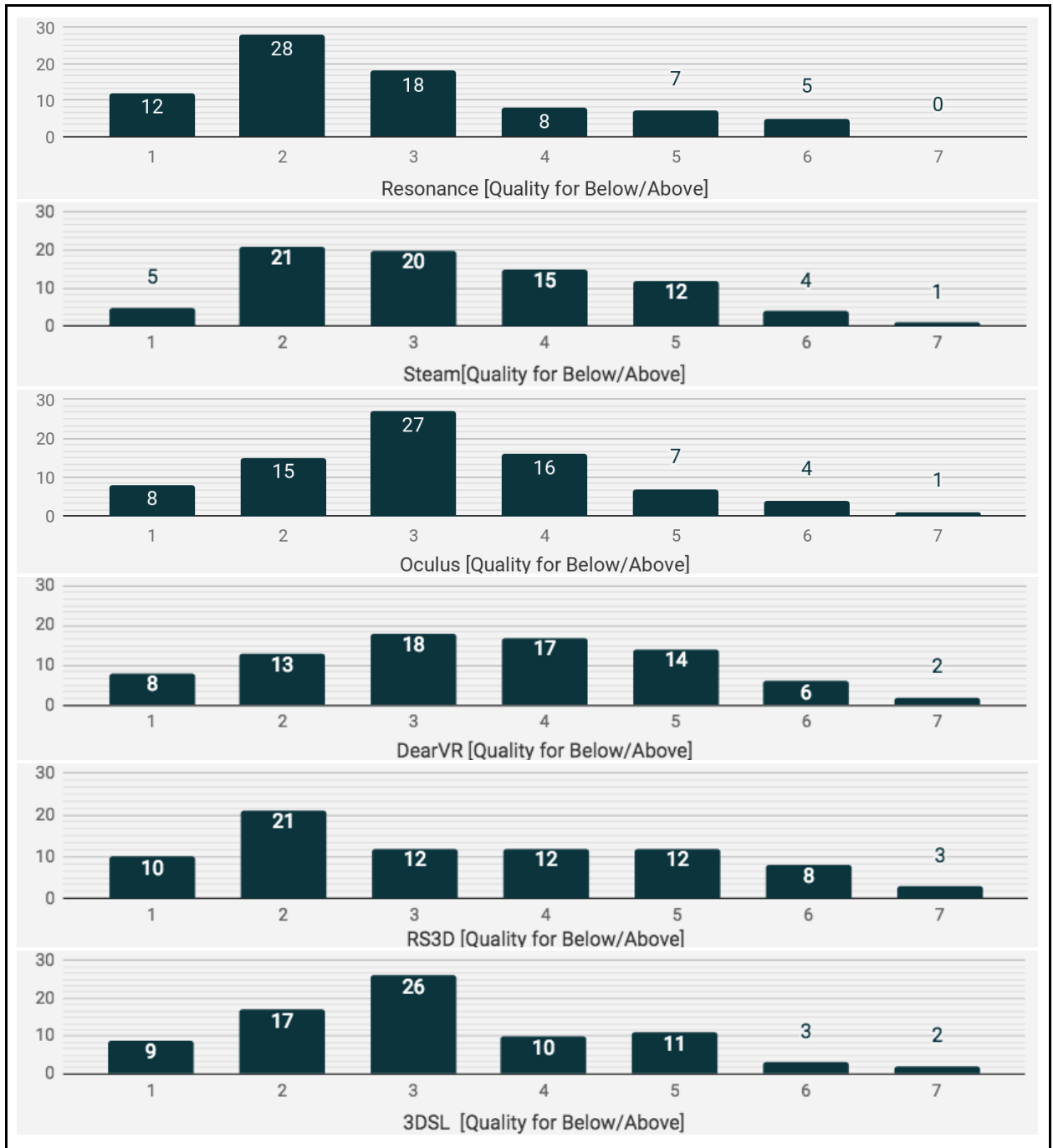


Figure 38: The Likert scale ratio for the quality for below/above as rated by the test subjects.

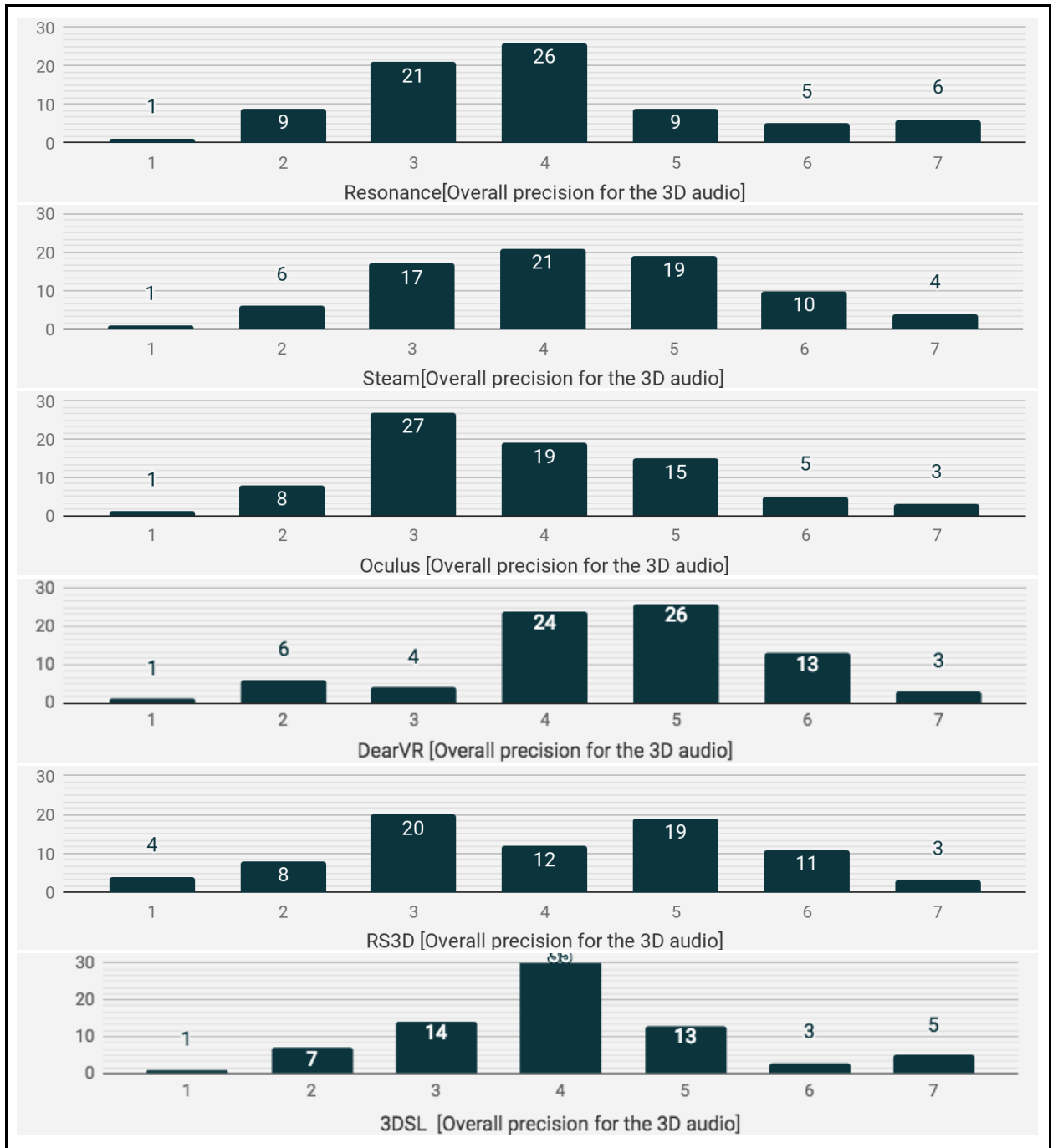


Figure 39: The Likert scale ratio for the overall precision as rated by the test subjects.

Direction Test Histograms

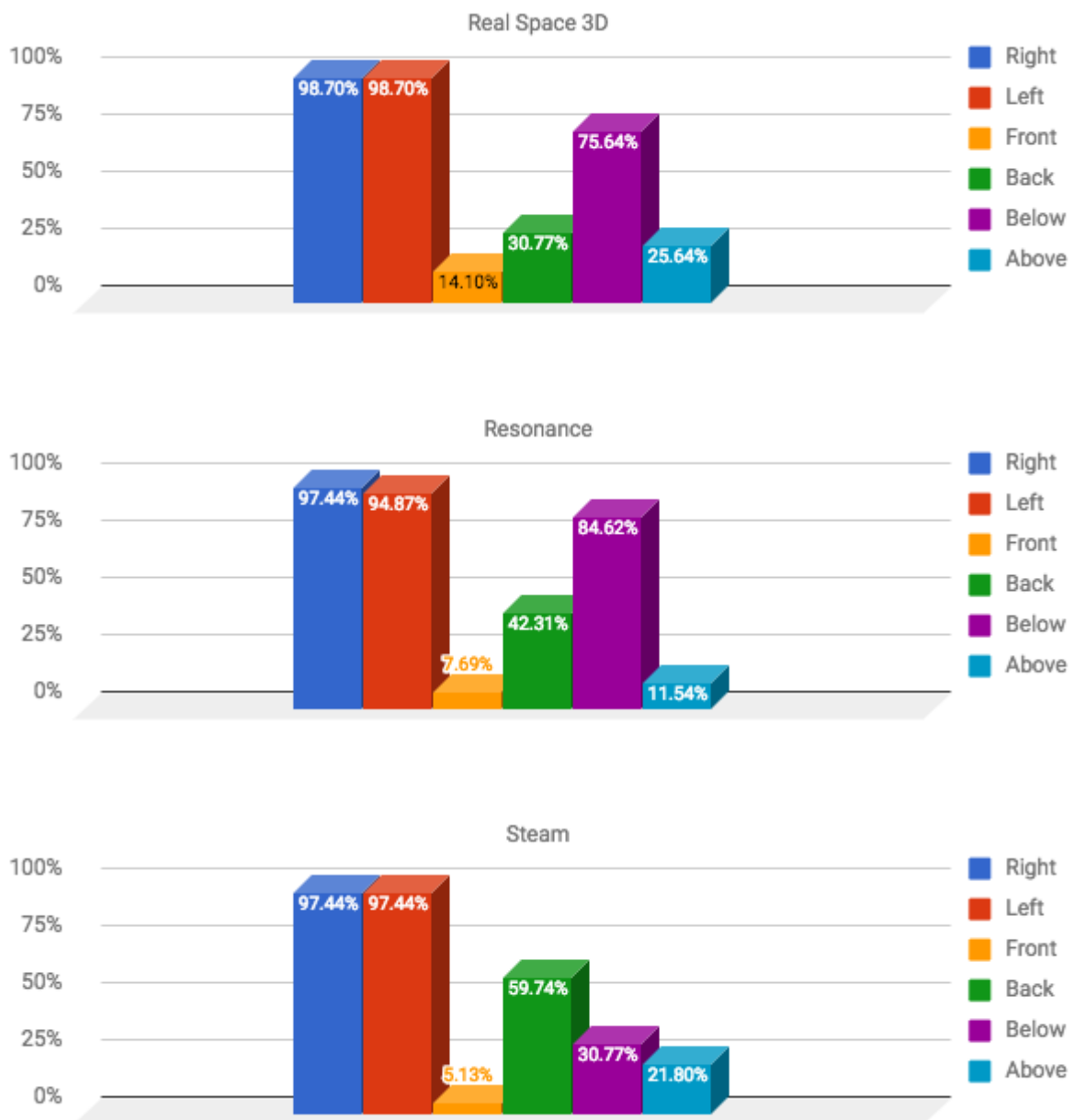


Figure 40: The average percentage ratio of correct answers from 78 participants.¹⁶³

¹⁶³ 77 answers in Steam about back. This taken into account when calculating the average results.

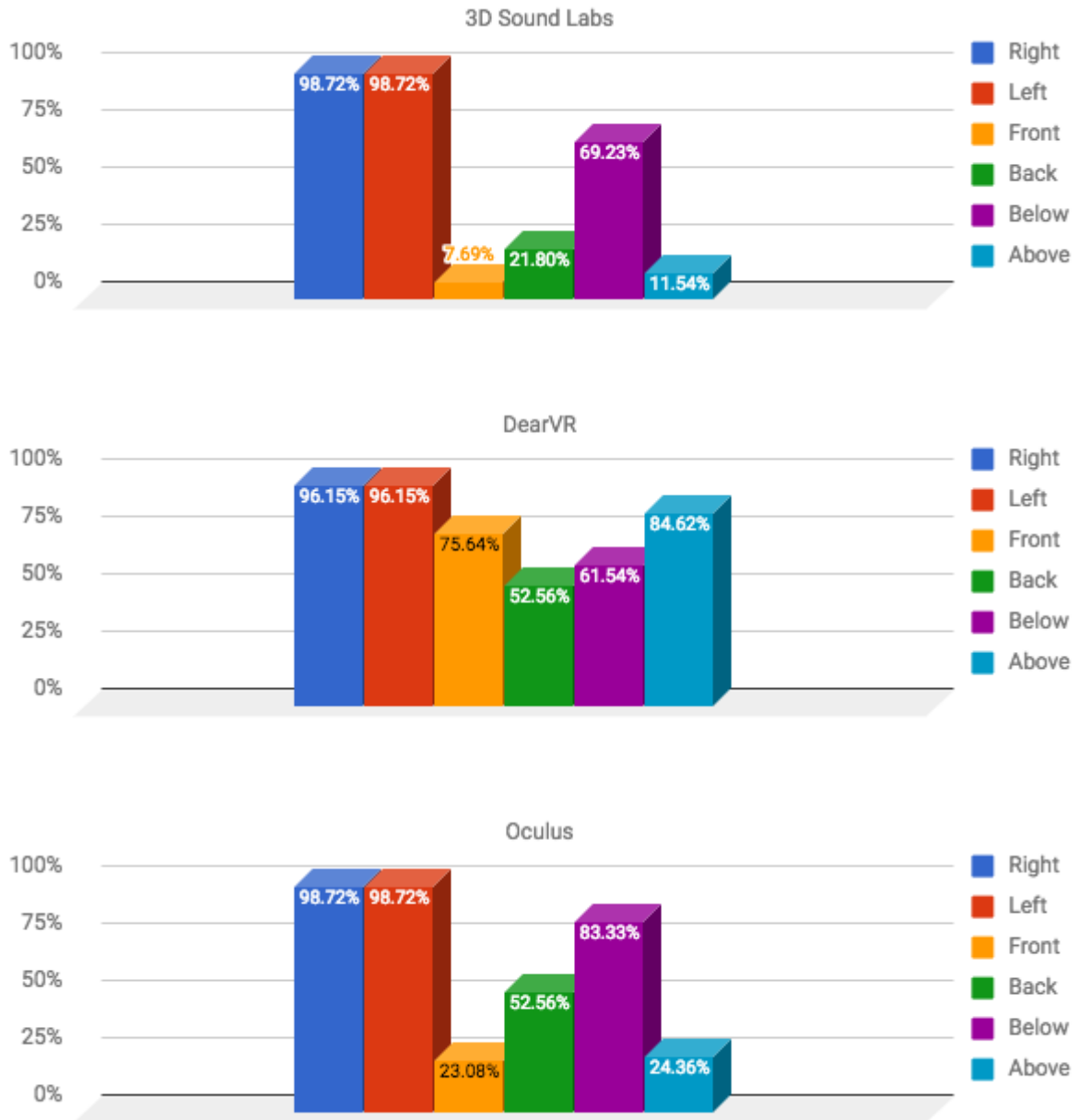


Figure 41: The average percentage ratio of correct answers from 78 participants.

D. Audio Components

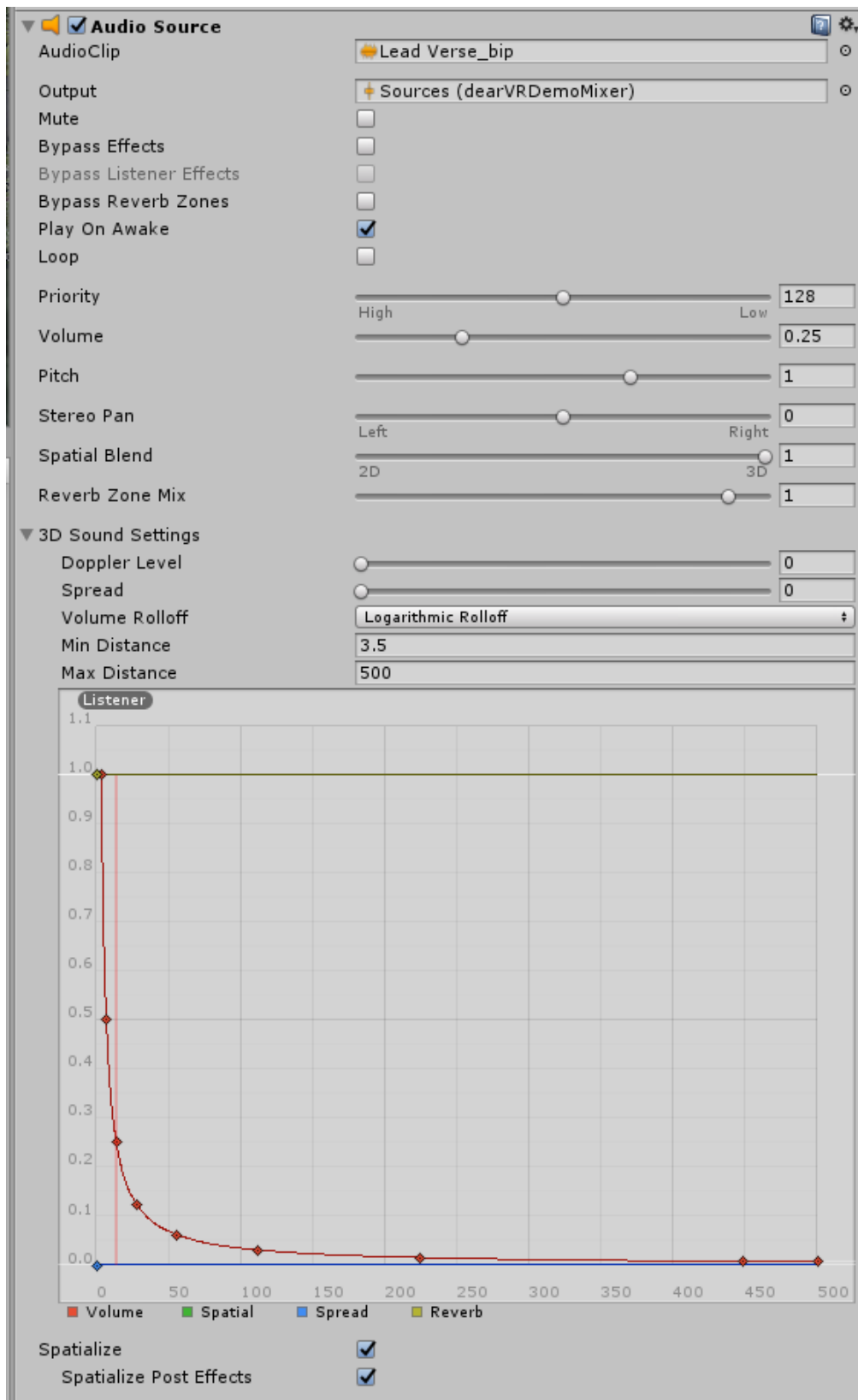


Figure 42: Unity's proprietary audio source.



Figure 43: DearVR source component.

List of Figures

Figure 1: Sammallahdenmäki burial cairn on an October morning in 2017.	4
Figure 2: Audio model.....	5
Figure 3: Prototype’s visual appearance with audio sources locations.....	6
Figure 4: First person view of the final audio scene.....	6
Figure 5: Example of mono and stereo sound sources.....	14
Figure 6: Example of a binaural listening situation.....	15
Figure 7: Time difference of sound waves travelling to both ears.....	19
Figure 8: IID causing “acoustic shadow” absorbing energy from the left sound waves.....	19
Figure 9: Counterbalanced test package versions.....	32
Figure 10: Files in the test package.....	33
Figure 11: Questionnaire for quality and direction tests.....	35
Figure 12: A following post was sent to each of the Facebook groups to gather participants.....	36
Figure 13: Calculating the average subject ratio per question.....	36
Figure 14: The average of subject Likert ratio for each plug-ins per question.....	37
Figure 15: The total average of subject ratio in seven point Likert scale in quality test.....	38
Figure 16: Average percentage of correct responses in direction test.....	38
Figure 17: Test participant immersed.....	41
Figure 18: Simplified information architecture of the prototype’s functionalities.....	45
Figure 19: Bird’s-eye view over Sammallahdenmäki in Unity editor.....	47
Figure 20: Early prototype with visible trigger colliders for testing Audio Zoom.....	47
Figure 21: The chorus-verse-chorus arrangement of the vocal choir and drums.....	51
Figure 22: Melody for humming choir in the key of C minor.....	52
Figure 23: Direct vocal recording.....	54
Figure 24: Vocal recording face sideways.....	54
Figure 25: Frame drum recording.....	55
Figure 26: DearVR was implemented into the prototype.....	58
Figure 27: Audio Source locations in Unity Editor.....	59
Figure 28: Info architecture of audio source locations.....	60
Figure 29: DearVR reverb component settings.....	62
Figure 30: Unity mixer settings.....	62
Figure 31: Timeline curves controlling the positional movements of the audio source object.....	68
Figure 32: The Audio Listener component was located 1.8 Unity units high in the scene.....	68
Figure 33: ‘Direction test’ with static audio source objects in six directions.....	68
Figure 34: Various audio processing tools used for the voice-over recordings.....	71
Figure 35: Audio files and loudness calibration as recommended in the EBU R128 standard.....	71
Figure 36: The Likert scale ratio for the quality for left/right as rated by the test subjects.....	74
Figure 37: The Likert scale ratio for the quality for front/back as rated by the test subjects.....	75
Figure 38: The Likert scale ratio for the quality for below/above as rated by the test subjects.....	76
Figure 39: The Likert scale ratio for the overall precision as rated by the test subjects.....	77
Figure 40: The average percentage ratio of correct answers from 78 participants.....	78
Figure 41: The average percentage ratio of correct answers from 78 participants.....	79
Figure 42: Unity’s proprietary audio source.....	80
Figure 43: DearVR source component.....	81

List of Tables

Table 1: List of features for all the spatializers.	28
Table 2: Latin Squares counterbalanced coding system for test packages.	32
Table 3: Responses rate to each of package versions.....	35
Table 4: Subject ratio for “level of presence” in two versions of VR.....	42
Table 5: Audio Model attributes for Virtual Heritage prototype.....	44
Table 6: The settings for audio parameters in audio source components.	60
Table 7: Randomized sound directions for all the spatializer plug-ins.	70

Bibliography

- "3D Audio Market." New Report by TMR. Accessed March 17, 2018.
<https://www.transparencymarketresearch.com/3d-audio-market.html>.
- "3D Sound Labs." Crunchbase. Accessed February 22, 2018.
<https://www.crunchbase.com/organization/3d-sound-labs#section-overview>.
- "About." Mapbox. Accessed February 26, 2018. <https://www.mapbox.com/about/>.
- "About the Licenses." Creative Commons. Accessed February 23, 2018.
<https://creativecommons.org/licenses/>.
- "A Brief Introduction to Traditional Sami Song and the Modern Music." Same Etnam. Accessed January 5, 2018. <http://boreale.konto.itv.se/smusic.htm>.
- Adams, Stephen, and Frank Boland. "On the Distortion of Binaural Localization Cues Using Headphones." In *Proceedings of IET Irish Signals and Systems Conference - ISSC 2010*, Cork, Ireland. Extended abstracts: 82-87. 2010. <https://doi.org/10.1049/cp.2010.0492>.
- "Age-Related Hearing Loss." National Institute of Deafness and Other Communication Disorders. Last modified January 16, 2018.
<https://www.nidcd.nih.gov/health/age-related-hearing-loss>.
- Ahrens, Jens. *Analytic Methods of Sound Field Synthesis*. Berlin: Springer, 2012.
- Ahrens, Jens, and Sascha Spors. "Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect." In Audio Engineering Society Convention Paper 7363, Presented at the 124th Convention 2008 Amsterdam, Netherlands.
Accessed January 5, 2018. <http://www.aes.org/e-lib/browse.cfm?elib=14493>.
- "Antiquity to 1590." A Brief History of Singing. Accessed February 14, 2018.
<http://www2.lawrence.edu/fast/KOOPMAJO/antiquity.html>.
- "Asuminen." Sammallahdenmäki. Accessed February 22, 2018.
www.sammallahdenmaki.fi/kouluille-pronssikausi-sammallahdenmaella/asuminen/.
- Begault, Durand R. *3-D Sound for Virtual Reality and Multimedia*. San Diego, CA: NTIS, 2000.
- Blauert, Jens. *Spatial Hearing: The Psychophysics of Human Sound Localization*. London: MIT Press, 1983.
- Bower, Bruce. "Humans: Oldest Known Instruments Found: Ivory, Bone Flutes Date from 35,000 to 40,000 Years Ago." *Science News* 176, no. 2 (10, 2009): 13. Accessed February 18, 2018. <https://onlinelibrary.wiley.com/doi/abs/10.1002/scin.5591760214>.

- Breebaart, Jeroen, and Christof Faller. *Spatial Audio Processing: MPEG Surround and Other Applications*. Chichester: John Wiley & Sons, 2007.
<https://www.safaribooksonline.com/library/view/spatial-audio-processing/9780470723487/>.
- Brown, Emily, and Paul Cairns. "A Grounded Investigation of Game Immersion." In *proceedings of the 2004 Conference on Human Factors and Computing Systems - CHI '04*, Vienna, Austria. Extended abstracts: 1297-1300. New York, NY: ACM, 2004.
<https://doi.org/10.1145/985921.986048>.
- Cairns, Paul, Anna Cox, and A. Imran Nordin, "Immersion in Digital Games: Review of Gaming Experience Research." New Jersey: John Wiley & Sons, 2000. Accessed January 20, 2018. <https://pdfs.semanticscholar.org/a9fb/2393b1e34be4155080bd02c6efba9e2cb807.pdf>.
- Cater, Kirsten, Richard Hull, Tom Melamed, and Robin Hutchings. "An Investigation into the Use of Spatialised Sound in Locative Games." In *proceedings of the 2007 Conference on Human Factors and Computing Systems - CHI '07*, 2007, San Jose, CA. Extended abstracts: 2315-2320. New York, NY: ACM, 2007.
<https://doi.org/10.1145/1240866.1241000>.
- Cheng, Kevin, and Paul A. Cairns. "Behaviour, Realism and Immersion in Games." In *proceedings of the 2005 Conference on Human Factors and Computing Systems – CHI '05*, 2005, Portland, OR. Extended abstracts: 1272-1275. New York, NY: ACM, 2005.
<http://dx.doi.org/10.1145/1056808.1056894>.
- Cherry, Kendra. "Learn What It Means When an Experiment Has A Within-Subject Design?" Verywell Mind. Accessed January 19, 2018.
<https://www.verywellmind.com/what-is-a-within-subjects-design-2796014>.
- Chuang, Priscilla. "Designing Sound Canvas: The Role of Expectation and Discrimination." In *proceedings of the 2002 Conference on Human Factors and Computing Systems – CHI '02*, Minneapolis, MN, Extended abstracts: 848-849. New York, NY: ACM, 2002.
<http://dx.doi.org/10.1145/506443.506629>.
- "Codex Runicus." Wikipedia. Last modified August 10, 2017.
https://en.wikipedia.org/wiki/Codex_Runicus#Content_of_the_Codex.
- "Company." VisiSonics. Accessed February 22, 2018.
<http://visisonics.com/company/#background>.
- De Souza, Jonathan. "Voice and Instrument at the Origins of Music." *Current Musicology* Vol. 97 (2014), 21 - 36. Accessed February 24, 2018.
https://journals.cdrs.columbia.edu/wp-content/uploads/sites/13/2016/08/CM97_DeSouza.pdf.
- Díaz-Kommonen, Lily. *Art, Fact and Artifact Production: Design Research and Multidisciplinary Collaboration*. Jyväskylä: University of Art and Design, 2002. Accessed March 30, 2018.
http://sysrep.aalto.fi/mulli2015/thesis_files/pdf/ch2_art_act_discourse_Diaz.pdf.

- Drossos, Konstantinos, Andreas Floros, and Nikolaos-Grigorios Kanellopoulos. "Affective Acoustic Ecology: towards emotionally enhanced sound events." In *proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound - AM '12*, 2012, Corfu, Greece. Extended abstracts: 109-116. New York, NY: ACM, 2012. <http://dx.doi.org/10.1145/2371456.2371474>.
- "Field of View." Wikipedia. Last modified January 02, 2018. https://en.wikipedia.org/wiki/Field_of_view#Humans_and_animals.
- Fingas, Jon. "Google Built a Spatial Audio Kit for Games and VR." Engadget. Last modified November 13, 2017. <https://www.engadget.com/2017/11/06/google-resonance-audio/>.
- Garner, Tom, and Mark Grimshaw. "A Climate of Fear." In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound - AM '11*, 2011, Coimbra, Portugal, Extended abstracts: 31-38. New York, NY: ACM, 2011. <http://dx.doi.org/10.1145/1859799.1859809>.
- Garner, Tom, Mark Grimshaw, and Debbie Abdel Nabi. "A Preliminary Experiment to Assess the Fear Value of Preselected Sound Parameters in a Survival Horror Game." In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound - AM '10*, 2010, Piteå, Sweden. Article No.10, 9 pages. New York, NY: ACM, 2010. <http://dx.doi.org/10.1145/1859799.1859809>.
- "Google/spatial-media." GitHub. Accessed April 4, 2018. <https://github.com/google/spatial-media/tree/master/spatial-audio>.
- Hendrix, C., and W. Barfield. "Presence in Virtual Environments as a Function of Visual and Auditory Cues." In *Proceedings Virtual Reality Annual International Symposium '95*, 1995, in Research Triangle Park, NC. Extended abstracts: 74-82. <https://doi.org/10.1109/VRAIS.1995.512482>.
- "History (and Future) of 3D Audio." Hooke Audio. Last modified August 21, 2017. <https://hookeaudio.com/blog/2016/10/11/history-and-future-3d-audio/>.
- "How to Build Audio for VR Games." Gamasutra Article. Accessed January 6, 2017. https://www.gamasutra.com/blogs/AlexRiviere/20171226/312290/How_to_Build_Audio_F_or_VR_Games.php.
- "HRTF Measurements of a KEMAR Dummy-Head Microphone." MIT Media Lab. Accessed December 18, 2018. <http://sound.media.mit.edu/resources/KEMAR.html>.
- Häkkinen, Kaisa. *Mistä Sanat Tulevat: Suomalaista Etymologiaa*. Helsinki: Suomalaisen Kirjallisuuden Seura, 2006.
- "Immersive Audio Solutions." Steam Audio. Accessed January 26, 2018. <https://valvesoftware.github.io/steam-audio/#learn-more>.
- "Interaural Time Difference." Wikipedia. Last modified December 18, 2017. https://en.wikipedia.org/wiki/Interaural_time_difference.

- "JOBS – We Are Hiring!" dearVR 3D Audio Reality Engine. Accessed January 26, 2018.
<http://dearvr.com/jobs/>.
- "Joik." Wikipedia. Last modified March 19, 2018. https://en.wikipedia.org/wiki/Joik#cite_note-2.
- Kalawsky, Roy S. *The Science of Virtual Reality and Virtual Environments: A Technical, Scientific and Engineering Reference on Virtual Environments*. Boston: Addison-Wesley, 1994.
- Katz, Bob. *iTunes Music*. Burlington: Focal Press, 2013.
<https://www.safaribooksonline.com/library/view/itunes-music/9780415656856/>.
- Keppel, Geoffrey, and Thomas D. Wickens. *Design and Analysis: A Researcher's Handbook*. New Jersey: Pearson, 2004.
- Klier, Michael. "Home." *Designing Sound*. Modified January 15, 2017.
<http://designingsound.org/2017/01/impulsonic-acquired-by-valve/>.
- Lee, Jiwon, Kim Mingyu, Kim Jinmo. "A Study on Immersion and VR Sickness in Walking Interaction for Immersive Virtual Reality Applications." In *Symmetry* 9, no. 5: 78. (2017). Accessed March 26, 2018. <https://doi.org/10.3390/sym9050078>.
- Ling, Johan, and Per Cornell. "Violence, Warriors, and Rock Art in Bronze Age Scandinavia." *Feast, Famine or Fighting? Studies in Human Ecology and Adaptation*, edited Richard J. Chacon, and Ruben G. Mendoza, 15-33. Springer, 2017. Accessed February 20, 2018.
https://doi.org/10.1007/978-3-319-48402-0_2.
- "Listen to The World's Oldest Known Secular Norse Song." *Realm of History*. Last modified October 13, 2017. <https://www.realmofhistory.com/2017/08/02/oldest-norse-song-codex-runicus/>.
- "Loudness (PLOUD)." *EBU Technology & Innovation*. Accessed February 8, 2018.
<https://tech.ebu.ch/loudness>.
- McMahan, Alison. "Immersion, Engagement, and Presence, A method for Analyzing 3-D Video Games." In *the Video Game Theory Reader*, edited Bernard Perron, and Mark J. P. Wolf, 67-86. New York: Routledge, 2009. Accessed January 15, 2018.
<http://www.alisonmcmahan.com/sites/default/files/pdfs/papers/ch3-McMahanrev.pdf>
- Meshram, Alok, Ravish Mehra, Hongsheng Yang, Enrique Dunn, Jan-Michael Franm, and Dinesh Manocha. "P-HRTF: Efficient Personalized HRTF Computation for High-fidelity Spatial Sound." In *proceedings of 2014 IEEE International Symposium on Mixed and Augmented Reality - ISMAR, 2014*. Munich, German. Extended abstracts: 53-61.
<https://doi.org/10.1109/ISMAR.2014.6948409>.
- Muratovski, Gjoko. *Research for Designers: A Guide to Methods and Practice*. London: Sage Publications, 2016.

- Murphy, David, and Flaithrí Neff. "Spatial Sound for Computer Games and Virtual Reality." In *Game Sound Technology and Player Interaction: Concepts and Developments*, edited by Mark Grimshaw, 287-312. IGI Global, 2011. Accessed December 18, 2017. <http://dx.doi.org/10.4018/978-1-61692-828-5.ch014>.
- Nazemi, Mark, and Diane Gromala. "Sound Design: a procedural communication model for VE." In *Proceedings of the 7th Audio Mostly Conference on A Conference on Interaction with Sound - AM '12*, 2012, Corfu, Greece. Extended abstracts: 16-23. New York, NY: ACM, 2012. <http://dx.doi.org/10.1145/506443.506629>.
- "Nordic Bronze Age." Wikipedia. Last modified January 22, 2018. https://en.wikipedia.org/wiki/Nordic_Bronze_Age#Climate.
- Owen, James. "Bone Flute Is Oldest Instrument, Study Says." National Geographic News. Accessed February 18, 2018. <https://news.nationalgeographic.com/news/2009/06/090624-bone-flute-oldest-instrument.html>.
- Paterson Natasa, Katsiaryna Naliuka, Soren Kristian Jensen, Tara Carrigy, Mads Haahr, and Fionnuala Conway. "Spatial Audio and Reverberation in an Augmented Reality Game Design." In *Proceedings of the AES 40th International Conference*, 2010, Tokyo, Japan. 9 pages. <http://www.aes.org/e-lib/browse.cfm?elib=15532>.
- Pelachaud, Catherine. *Emotion-Oriented Systems*. London: ISTE, 2012. <https://www.safaribooksonline.com/library/view/emotion-oriented-systems/9781118603321/>.
- Pentikäinen, Juha. *Saamelaiset: Pohjoisen Kansan Mytologia*. Helsinki: Suomalaisen Kirjallisuuden Seura, 1995.
- Pougnadoresse, F. De Sorbier De, P. Bouvier, A. Herubel, and V. Biri. "From Research on the Virtual Reality Installation." *Human-Computer Systems Interaction Advances in Intelligent and Soft Computing*, 2009, 335-45. Accessed January 9, 2018. <https://hal-uepec-upem.archives-ouvertes.fr/hal-00681737/document>.
- "Presence | Definition of Presence in English by Oxford Dictionaries." Oxford Dictionaries. Accessed January 21, 2018. <https://en.oxforddictionaries.com/definition/presence>.
- "QSound Labs Inc. - Leaders in 3D Audio Enhancement & Virtual Surround Sound Technology." QSound Labs: Virtual Barber Shop. Accessed on Feb 2, 2018. http://www.qsound.com/demos/virtualbarbershop_long.htm.
- "Real World Terrain." Asset Store. Accessed November 10, 2017. <https://assetstore.unity.com/packages/tools/terrain/real-world-terrain-8752>.
- "Reality of VR/AR Growth." TechCrunch. Last modified January 12, 2017. <https://techcrunch.com/2017/01/11/the-reality-of-vr-ar-growth/>.

- Regenbrecht, Holger, and Thomas Schubert. "Real and Illusory Interactions Enhance Presence in Virtual Environments." In *Presence: Teleoperators and Virtual Environments* 11, no. 4 (08 2002): 425-34. Accessed February 28, 2018. <https://doi.org/10.1162/105474602760204318>.
- Roginska, Agnieszka, and Paul Geluso. *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. New York: Routledge, 2018. <https://www.safaribooksonline.com/library/view/immersive-sound/9781317480105/>.
- Rossing, Thomas D., F. Richard Moore, and Paul Wheeler. *The Science of Sound*. San Francisco: Pearson, 2002.
- Rozwadowski, Andrzej. "Historic and proto-historic shamanic rock art in Siberia: a view from the Altai." In *proto-historical and Historic Rock Art in the Iberian Peninsula International Conference, 2014, Castellón, Spain*. Extended abstracts 195-204. Accessed February 14, 2018. http://repositori.uji.es/xmlui/bitstream/handle/10234/156585/Quaderns_33_12.pdf?sequence=1.
- Rumsey, Francis. *Spatial Audio*. Oxford: CRC Press, 2017.
- Sachs, Curt. *The History of Musical Instruments*. New York: Norton & Company, 1940. https://archive.org/details/the_history_of_musical_instruments_curt_sachs.
- "SÁMI GRAND PRIX 2011 | All the yoiks." YouTube. Accessed 16 February 2018. www.youtube.com/watch?v=cg0TQyjdHJ0.
- Sax, Leonard. "Sex Differences in Hearing: Implications for best practice in the classroom." In *Advances in Gender and Education, 2* (2010). Montgomery Center for Research in Child Adolescent Development, 2010. 13-21. Accessed February 26, 2018. <http://www.leonardsax.com/sex-differences-in-hearing/>.
- Schlemmer, Eliane, and Luciana Backes. "Immersion, Telepresence, and Digital Virtual Presence in Metaverses." In *Learning in Metaverses: Co-Existing in Real Virtuality*, 102-126. IGI Global, 2015. Accessed April 03, 2018. doi:10.4018/978-1-4666-6351-0.ch005.
- "Sigur Rós - Frequently Asked Questions." Return to the Front Page. Accessed November 22, 2017. <https://sigur-ros.co.uk/band/faq.php#08>.
- Slater, Mel. "A Note on Presence Terminology." Presence Connect, 2003. Accessed January 6, 2018. www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology.
- "Sound Quality." Wikipedia. Last modify March 05, 2018. https://en.wikipedia.org/wiki/Sound_quality.

- Staff, Live Science. "Babies Are Born to Dance." LiveScience. Last modified March 15, 2010. <https://www.livescience.com/6228-babies-born-dance.html>.
- "Stereo Imaging." Wikipedia. February 18, 2018. https://en.wikipedia.org/wiki/Stereo_imaging.
- Swallow, Dave. *Live Audio: The Art of Mixing a Show*. Oxford, Focal Press, 2011. www.safaribooksonline.com/library/view/live-audio-the/9780240816043/.
- Talbot-Smith, Michael. *Sound Engineering Explained*. Burlington: Focal Press, 2013. www.safaribooksonline.com/library/view/sound-engineering-explained/9780240516677/.
- Teatro Altosf. "DEL DESTIERRO..." YouTube. Last modified June 16, 2010. https://www.youtube.com/watch?v=_xYAtJXXyY.
- Technologies, Unity. "Audio Source Overview." Unity - Manual: Audio Source. Accessed March 22, 2018. <https://docs.unity3d.com/Manual/class-AudioSource.html>.
- Technologies, Unity. "Timeline Overview." Unity - Manual: Timeline Overview. Accessed February 7, 2018. <https://docs.unity3d.com/Manual/TimelineOverview.html>.
- Théberge, Paul, Kyle Devine, and Tom Everrett. *Living Stereo: Histories and Cultures of Multichannel Sound*. London: Bloomsbury Academic, 2015.
- Thorn, Alan. *Mastering Unity 2017 Game Development with C# - Second Edition*. Birmingham: Packt Publishing, 2017. <https://www.safaribooksonline.com/library/view/mastering-unity-2017/9781788479837/>.
- Triantafyllou, David, Angeliki Antoniou, and George Lepouras. "Sound and Kinesthesia in Virtual Environments." In *Proceedings of the 20th Pan-Hellenic Conference on Informatics – PCI '16*, 2016, Patras, Greece. Article No.55. NY: ACM, 2016. <https://doi.org/10.1145/3003733.3003737>.
- Uusi-Seppä, Niina. *Satakunnan Kulttuuriympäristöt - eilen, tänään, huomenna*. Pori: Satakunnan Museo, 2012. https://issuu.com/satakunnan_museo/docs/eilen_tanaan_huomenna_web.
- Vesikansa, Jouko. *Nykysuomen Sanavarat*. Helsinki: WSOY, 1989.
- "Wardruna, Aurora and Oslo Fagottkor: "HELVEGEN"." YouTube. Last modified July 08, 2017. <https://www.youtube.com/watch?v=cg0TQyjdHJ0>.
- "Why Do We like Our Music Loud?" MusicRadar. Last modified July 13, 2009. <https://www.musicradar.com/tuition/tech/why-do-we-like-our-music-loud-212790>.