Bachelor's Thesis

How Spatial Audio affects Motion Sickness in Virtual Reality

Author:
Tim Philipp
Matricel no.: 32230

Supervisors:
Prof. Dr. Stefan Radicke
Prof. Oliver Curdt

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Engineering

in the

Faculty Electronic Media
Study Programme Audio Visual Media

January 9, 2020

Declaration of Authorship

Hiermit versichere ich, Tim PHILIPP, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „How Spatial Audio affects Motion Sickness in Virtual Reality“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichenVersicherung zur Kenntnis genommen.

I hereby affirm, Tim PHILIPP, on my honor, that I have written this bachelor thesis with the title: "How Spatial Audio affects Motion Sickness in Virtual Reality" independently and without outside help and did not use any other aids than those stated. The parts of the work that were taken from the wording or the meaning of other works are in any case marked with the source. The work has not yet been published or submitted in any other form as an examination performance.

I have taken note of the significance of the assurance on honor and the consequences under examination law (§26 paragraph 2 Bachelor-SPO (6 semesters), §24 paragraph 2 Bachelor-SPO (7 semesters), §23 paragraph 2 Master-SPO (3 semesters) or §19 paragraph 2 Master-SPO (4 semesters and part-time) of the HdM) of an incorrect or incomplete assurance on honor.

Signed:

Date:

"A sure cure for seasickness is to sit under a tree."

Lord Nelson (1758 - 1805)

HOCHSCHULE DER MEDIEN

Abstract

Electronic Media
Study Programme Audio Visual Media

Bachelor of Engineering

How Spatial Audio affects Motion Sickness in Virtual Reality

by Tim PHILIPP

In this research work I examine the effects of spatially perceptible ambient sound on visually induced motion sickness (VIMS) in Virtual Reality (VR).
For this purpose, I will let a total of 38 participants, divided into two test groups, experience the same VR application twice, one with stereo and one with spatial audio setup. While measuring their pulse, I let the users travel in a moving vehicle a 5-minute ride along a winding track with ascents and descents at medium speed.
Before the journey, the general susceptibility for motion sickness is recorded, between and after the test units the current symptoms are queried by means of a questionnaire developed especially for simulator sickness.
The test results then give me information whether the use of surround sound has a positive effect on the VIMS in virtual environments.

In dieser Forschungsarbeit untersuche ich die Auswirkungen von räumlich wahrnehmbaren Umgebungsgeräuschen auf die visuell induzierte Bewegungskrankheit in Virtual Reality (VR).
Dazu lasse ich insgesamt 38 Teilnehmer, aufgeteilt in zwei Testgruppen, die gleiche VR-Anwendung zweimal erleben, einmal mit Stereo und einmal mit räumlichem Audio-Setup. Während der Pulsmessung lasse ich die Benutzer in einem fahrenden Fahrzeug eine 5-minütige Fahrt auf einer kurvenreichen Strecke mit Auf- und Abstiegen bei mittlerer Geschwindigkeit durchführen. Vor der Fahrt wird die allgemeine Anfälligkeit für die Bewegungskrankheit erfasst, zwischen und nach den Testeinheiten werden die aktuellen Symptome mit Hilfe eines speziell für die Simulatorkrankheit entwickelten Fragebogens abgefragt.
Die Testergebnisse geben mir dann Auskunft, ob sich der Einsatz von Surround-Sound in virtuellen Umgebungen positiv auf die Bewegungskrankheit auswirkt.

Acknowledgements

Prof. Dr. Stefan Radicke
Prof. Oliver Curdt
Lisette Vonk
Dr. Jurriaan Mulder
Stuttgart Media University (HdM)
Amsterdam University of Applied Sciences (HvA)
Dennis Boon

Contents

Declaration of Authorship iii

Abstract vii

Acknowledgements ix

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Related Research Papers on Motion Sickness and Sound . . . 3
  - 1.2.1 Demonstrating the Potential for Dynamic Auditory Stimulation to Contribute to Motion Sickness . . . 3
  - 1.2.2 Pleasant music as a countermeasure against visually induced motion sickness . . . 3
  - 1.2.3 Visual-vestibular conflict induced by Virtual Reality in humans 4
- 1.3 Preview of the Thesis . . . 4

2 Stereoscopic Vision and Displays 5
- 2.1 Human visual system and depth perception . . . 5
- 2.2 Technical characteristics of Head Mounted Displays . . . 10
- 2.3 Overview and comparison of current Head Mounted Displays . . . 15
  - 2.3.1 Product choice and explanation . . . 16

3 Audio and Spatial Sound 19
- 3.1 History and recent Development . . . 19
- 3.2 Sound Waves . . . 23
- 3.3 Human Audio Perception . . . 24
  - 3.3.1 Interaural Time Difference . . . 25
  - 3.3.2 Interaural Intensity Difference . . . 25
  - 3.3.3 Blauert Bands . . . 26
  - 3.3.4 Head-Related Transfer Function . . . 27
- 3.4 Spatial Game Audio . . . 29

4 Kinetosis (Motion Sickness) 39
- 4.1 Vestibular system . . . 40
- 4.2 Motion Sickness Susceptibility Questionnaires (MSSQ) . . . 41
- 4.3 Simulator Sickness Questionnaire (SSQ) . . . 41

5 Experimental Setup Description **45**
5.1 Research Components 45
5.2 Metrics 45
5.3 Hypothesis 50
5.4 Execution 51
5.5 Test Results 51

6 Evaluation **55**
6.1 Metrics Assignment 55
6.2 Main Conclusion 61
6.3 Room for Interpretation 61
6.4 Further Research Potential 62

List of Abbreviations

VIMS	Visually induced Motion Sickness
VR	Virtual Rreality
AR	Augmented Rreality
MR	Mixed Rreality
e.g.	exempli gratia (Latin), meaning "for example"
VVC	visual-vestibular conflict
CGI	Computer-Generated imagery
MSQ	Motion Sickness Questionnaire
SSQ	Simulator Sickness Questionnaire
HMD	Head Mounted Display
DoF	Depth-of-Field
LCD	Liquid Crystal Display
OLED	Organic Light Emitting Diode
AMOLED	Active Matrix Organic Light Emitting Diode
TFT	thin-film-transistor
i.e.	id est (Latin), meaning „that is to say"
fps	frames per second
CFF	critical flicker frequency
HFR	higher frame rate
IT	Information Technology
DoF	Degrees of Freedom
HDMI	High-Definition Multimedia Interface
USB	Universal Serial Bus
ppi	pixels per inch
USP	Unique Selling Point
HiRes	High Resolution
CD	Compact Disc
PCM	pulse code modulation
SNR	signal-to-noise-ratio
DAW	Digital Audio Workstation
MIDI	Musical Instrument Digital Interface
SQNR	signal-to-quantization-noise-ratio
ITD	Interaural time difference
IID	Interaural intensity difference
API	Application Programming Interface

SDK	**S**oftware **D**evelopment **K**IT
DSP	**D**igital **S**ignal **P**rocessor
CNS	**c**entral **n**ervous **s**ysytem
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers

Physical Constants

Speed of Light	$c_0 = 2.99792458e8 \times 10^8 \frac{m}{s}$ (exact)
Speed of Sound	$v_{sound\,in\,air} = 340.29 \frac{m}{s}$ (approx)

List of Symbols

c_0	Speed of Light	$\frac{m}{s}$
f	Frequency	Hz
T	Period	Hz = $\frac{1}{s}$
λ	Wavelength	m
A	Amplitude	m
Φ	Phase	rad
x	Distance	m
v	Velocity	$\frac{m}{s}$
t	Time	s

Chapter 1

Introduction

1.1 Motivation

Virtual Reality, VR for short, is currently on everyone's lips. Although the Gartner Hype Cycle 2016 still predicted a breakthrough for VR ("Slope of Enlightenment"), the technology is not included in this year's forecast of the market research company, as are Augmented Reality (AR) and Mixed Reality (MR). The reason for this is the fact that VR is already almost mature and the technology is predicted to reach the high level of maturity ("Plateau of Productivty") in three to eight years from now. Partly this new technology is seen as an opportunity for some areas but it is not yet possible to estimate what effect a continuous consumption will have on the human psyche and body. Long-term studies on this topic are not available yet. In my following thesis I will therefore explore how different stereo or spatial sound sources affect the human sense of balance in VR.

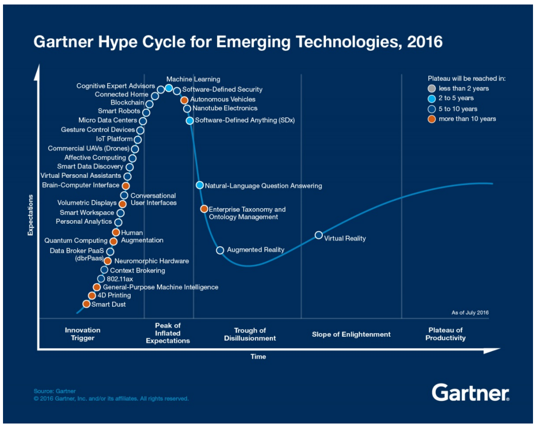


Fig. 1.1 Gartner (2016): *Hype Cycle For Emerging Technologies*

Before I begin to dive further into the test scenario, it is good to take a step back and understand the illusions of presence in order to achieve the intended VR experience for my test audience. Presence in a synthesized multi-sensory environment is generally defined as a user's subjective sensation of "being there" in a scene depicted by a medium (Barfield and Furness 1995), usually virtual in nature. In the field of presence research, VR is therefore located between real experience and hallucination and characterized by a sensory or non-sensory experience of para-authentic or artificial objects (Lee 2004).

> *Presence is a **psychological** state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's **perception** fails to accurately acknowledge the role of the technology in the experience.* (Presence Research[online] 2000)

For me personally, a research work in this intermediate space of human senses represents a special fascination. It's not only the borderline experience of human resilience that plays a role but also the approach and gradual fusion of human individuals with computer technology built by him.
Much has been said about the positives of VR technology that can reshape reality or even create a new one, be it creating new methods of communication, of medical treatment or of military training. But in my opinion we should also be aware of the potential negatives of reality technologies. The high speed with which new technologies develop may be the basic problem here: As soon as new possibilities in science or technology become public, the industry and economy immediately insists on pouring them into products and investment options as quickly as possible in order to withstand the pressure of our commercialized financial world. The problem isn't the technology; it's that its adoption rate has been so relatively quick that people, society and potential thoughts on consequences haven't had a chance to catch up. In early technologies, be it the vehicle or the color TV, we can see a steep but still relatively prolonged adoption rate. But when it comes to the smartphone, the internet or mobile AR the lines are nearly vertical.

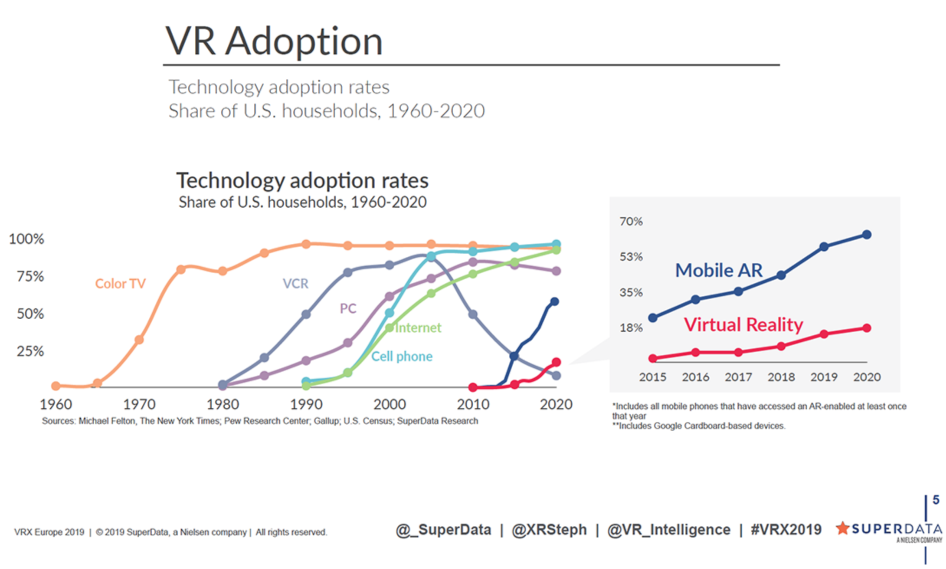


Fig. 1.2 Stephanie Llamas, Nielsen/SuperData Research (2019): *Technology adoption rates, share of U.S. households, 1960 - 2020*

So, questions about ethics, sense and last but not least the consequences for nature, people and health fall by the wayside. I would like to oppose this compulsion for high-speed development by taking a step backwards and discuss the question of an immersive, visual digital burden on human health.

Indeed, VR is perfect for creating environments in which to test, train, teach and treat people. That's because this technology can be used to create the ultimate Skinner box (Skinner 1935): In this experiment rats were conditioned (so-called *operant conditioning*) differently in a box with the help of a lever. Pressing the lever was the desired behavior. After pressing the lever, different consequences followed in different test series. Depending on the amplifier, the behavior of the rat changed as follows: Positive/negative amplification or positive/negative punishment. Accordingly, it is possible to put someone in a (virtual) environment, let him interact in 3D space and actually change how his brain functions, e.g. reducing the peoples' implicit racial bias (Hasler, Spanlang, and Slater 2017).
I would like to highlight the physical effects of VR usage and try to establish a relation between the auditory spatiality and the influence on the sense of balance during an experience. My assumption here is that spatially occurring sounds have a considerable influence on the VIMS. My theory is that sound - similar to the strength of anxiety in a horror movie - has a significant and, in my opinion, underestimated value. For this very reason, I will prove in this present work that 3D auditory sensory impressions have a remarkable influence on the human balance perception.

1.2 Related Research Papers on Motion Sickness and Sound

1.2.1 Demonstrating the Potential for Dynamic Auditory Stimulation to Contribute to Motion Sickness

The study found that motion sickness was not affected by the inclusion of corresponding auditory information such that visual-only and combined auditory plus visual cues elicited comparable sickness results. In contrast, auditory cues increased the level of vection when they were added to the corresponding visual stimulus whereas vection elicited by pure auditory stimulation was rather weak. Most interestingly, motion sickness symptoms induced by pure auditory stimulation were observed in a sub-set of participants, demonstrating the existence of auditorily induced motion (Keshavarz, Hettinger, et al. 2014). With the help of this basic consideration I prepare my application with sound sources that correspond to what is currently seen as well as those that can only be heard in the background. The resulting mix of visible and subtle audio sources will cause motion sickness symptoms and make my results more meaningful.

1.2.2 Pleasant music as a countermeasure against visually induced motion sickness

In this article (Keshavarz and Hecht 2014) , the researchers found that when factoring the subjective pleasantness of music, a significant reduction of VIMS occurred only when the presented music was perceived as pleasant, regardless of the music type. On the basis of these results I decide to build my test application with a dark auditory basic mood and to extend it at some places with rhythmically pleasing but also heroically epochal music. I hope to create a broad emotional arc between tension, excitement and joy for the testers in order to achieve significant and meaningful results at certain points of the test ride.

1.2.3 Visual-vestibular conflict induced by Virtual Reality in humans

In this study, the researchers examined the development of subjective sickness- and balance related symptoms and objective equilibrium ataxia induced by visual–vestibular conflict (VVC) stimulation using VR. The study suggests that the VVC inputs are processed in different pathways causing subjective autonomic symptoms and postural instability in humans (Akiduki et al. 2003). I take this caused instability as a basis for my approach to visually induce the participants motion sickness in VR.

1.3 Preview of the Thesis

In this study, I used a VR-based 3D environment to imitate a real-world situation. I put a moving vehicle on a track to mimic realistic motion stimuli, surrounded the scene with computer-generated imagery (CGI) and implemented different audio sources the enhance the level of realism and immersion. Subjects are asked to sit in the vehicle to explore the ride. It was found that this scene is sufficient enough to induce simulator sickness symptoms almost identical to real motion sickness. First, the Motion Sickness Questionnaire [(Golding 1998), (Golding 2006a)] (MSSQ-Short) is used to indicate the vulnerability of the test subjects, then the heart rate is observed during the experience and the Simulator Sickness Questionnaire (R. S. Kennedy et al. 1993) (SSQ) is filled in between and afterwards to determine typical symptoms of simulator sickness.

Chapter 2

Stereoscopic Vision and Displays

Head-mounted displays, or HMDs, are an almost ancient piece of tech which have begun to see a reboot in the past few years as computers get more powerful and the games inside them more visually spectacular by the day. They are probably the most instantly recognizable objects associated with VR. As such, they are also referred to sometimes as 'VR headsets' or 'VR glasses'. HMDs attach straight to the head and present visuals directly to the eyes, and perhaps most excitingly, to the peripheral vision as well. As humans, the vision systems have evolved to a point where we have very high resolution in the front, wherever we're looking. Also, humans have very low resolution on the periphery but motion in the periphery is something that we pick up on because there could be a threat.

Fig. 2.1 Sutherland, I.E. (1968): *A Head -Mounted Three-Dimensional Display*

HMDs are not only used in VR gaming, they've also been utilized in military, medical and engineering contexts to name a few. As seen in films such as *The Terminator* (1984), these devices can be used to create AR which overlays digital information through an HMD filter onto the real world. They also have an incredible range of use. In order to allow the best possible experience with an HMD, I listed a number of basic principles and technologies below that need to be incorporated.

2.1 Human visual system and depth perception

Light-sensitive sensory cells on the retina are responsible for **color vision**. There are two types of these photoreceptors: Rods and cones. Both receptor types contain visual pigments that react to a specific wavelength between 780 nm (red) and 380 nm (violet). The visual pigment embedded in the rods is called rhodopsin, the visual pigment of the cones is called iodopsin. Rhodopsin and iodopsin are light-sensitive protein molecules with different dye additives. They change their

chemical composition as soon as light enters the receptor cells. The retina of the human eye is covered with about 120 million rods that are optimized for twilight vision. They mainly perceive **light-dark contrasts** and convey the perception of grey tones (*detection sensitivity*). The approximately seven million cones are responsible for high-resolution color vision during the day and the perception of movement. They occur in three specialized types which differ in the composition of the visual pigment and the corresponding observed wavelength. There are cones whose visual pigment responds most strongly to red, green or blue whereas green contains the highest brightness information. The chemical process that is triggered in the photoreceptors by the absorbed light energy converts the optical signal into an electrical nerve impulse. Subsequently, the so-called ganglion cells convert this electrical stimulus into a neural signal that reaches the brain via different switching points and pathways.
Humans have two eyes which means we can receive monocular and binocular depth cue information. **Monocular** depth cues require the use of only one eye whereas **binocular** depth cues require the use of both eyes working together in order to provide information to the brain about depth and distance. The following signal processing factors are important for **monocular stimuli**:

Motion parallax refers to the fact that objects moving at a constant speed across the frame will appear to move a greater amount if they are closer to an observer (or camera) than they would if they were at a greater distance.

Fig. 2.2 Krantz, J. H., Schwartz, B.L. (2015): *ISLE 7.5 (a1): Motion Parallax Illustrated*

Perspective: On a drawing in central perspective two lines converge at a point in the horizon. Between the lines there are two bars, one further down on the paper (perspectively in front) and one further up (perspectively further away). The upper bar is perceived as larger although both lines have the same length.

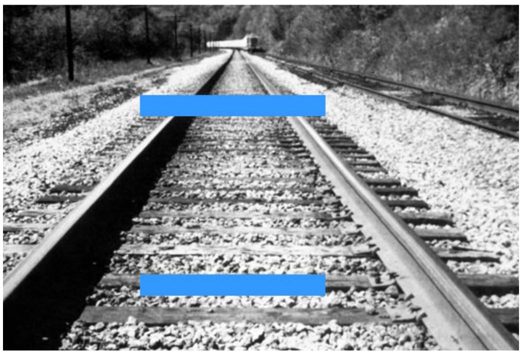


Fig. 2.3 Ponzo, M. (1913): *rail illusion*

Object size: Relative object size refers to the tendency to visually perceive the object that produces the largest image on the retina as being closer and the object that produces the smallest image on the retina as being further away. However, the objects being perceived must be expected to be about the same size in real life.

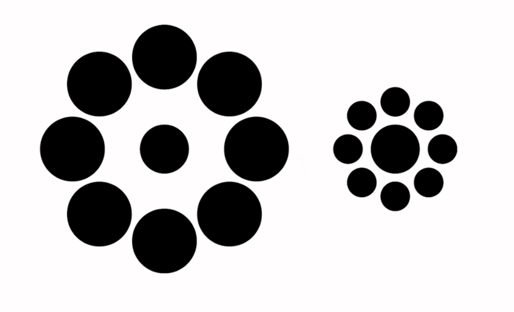


Fig. 2.4 Ebbinghaus, H. (1890): *Ebbinghaus illusion* (l)., Delboeuf, J. R. L. (1865): *Delboeuf illusion* (r.)

Accommodation involves the automatic adjustment of the shape of the lens to focus an object in response to changes in how far away the object is. It bulges to focus on nearby objects and flattens to focus on objects further away. The brain then judges distances by how much the lens bulges or elongates.

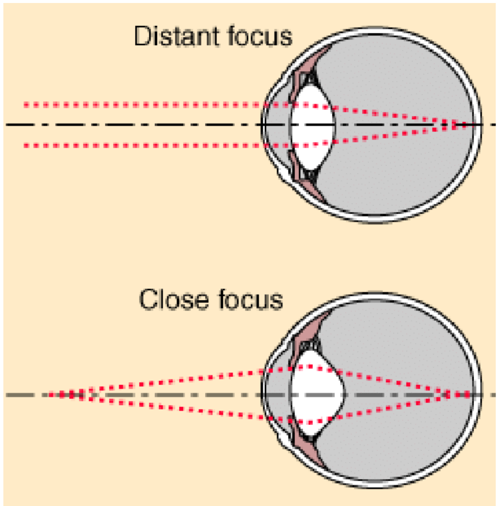


Fig. 2.5 Sweet, B., Kaiser, M. (2011): *Depth Perception, Cueing, and Control*

Depth-of-Field (DoF): Blur in images can create the sensation of depth because it emulates an optical property of the eye; namely, the limited depth of field created by the eye's lens. When the human eye looks at an object, this object appears sharp on the retina but objects at different distances appear blurred.

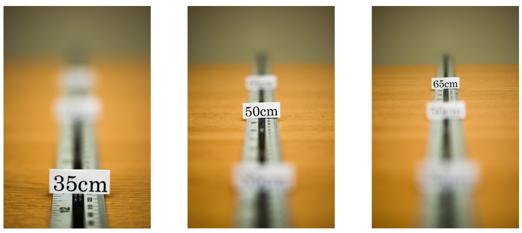


Fig. 2.6 Mauderer, M. , Conte, S. , Nacenta, M. A., Vishwanath, D. (2014): *Depth perception with gaze-contingent depth of field*

Occlusion is produced by partially overlapping objects: Objects that partially block other parts of the scene are perceived to be closer to an observer than the blocked objects (and vice versa).

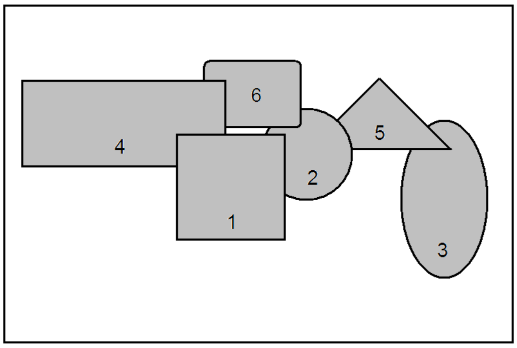


Fig. 2.7 McIntire, J., Havig, P. R., McIntire L. (2009): *Ideas on authenticating humanness in collaborative systems using AI-hard problems in perception and cognition*

Lighting (shading, reflection and illumination): Brightness of a surface depends on its orientation with respect to the light source. The visual system assumes that the light comes from above (sun). Brighter patches appear to be tilted up facing the light. Also, the interpretation of shape from shading interacts with the interpretation of shape from contours.

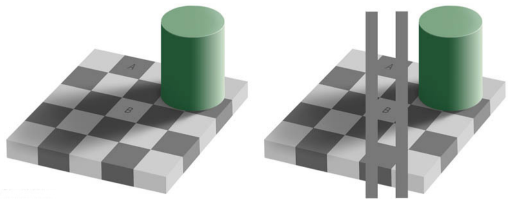


Fig. 2.8 Adelson, E. H. (1995): *The checker shadow illusion*

The following signal processing factors are important for **binocular stimuli**:

Retinal Disparity (Stereopsis) refers to the very slight difference in the location of the visual images on the retinae which enables humans to make judgements about the depth or distance of an object.

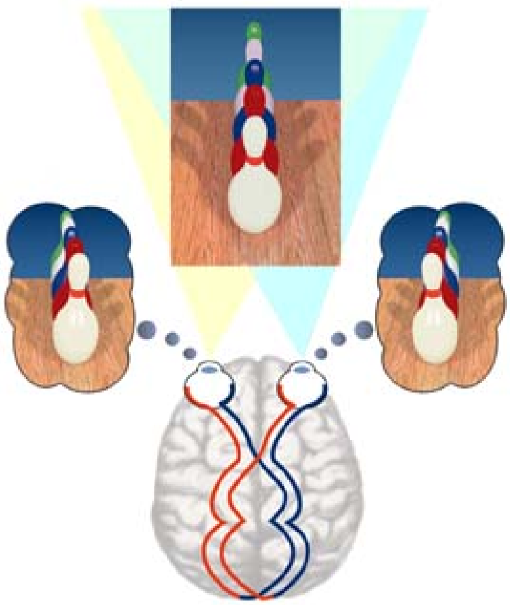


Fig. 2.9 Cooper, R. (1995): *How to See 3D*

Convergence involves the brain detecting and interpreting depth or distance from changes in tension in the eye muscles that occur when the two eyes turn inwards to focus on objects that are close. The brain interprets greater tension in the eye muscles as an object gets closer and less tension as an object gets further away. As a person ages, the ability to accommodate decreases resulting in the condition called *presbyopia*.

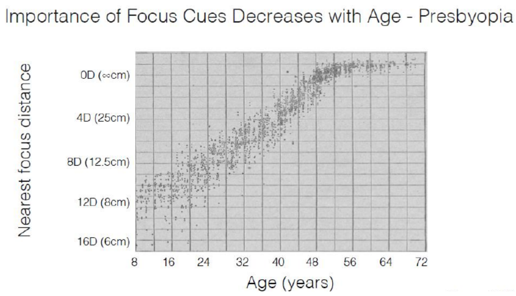


Fig. 2.10 Duane, A. (1912): *Normal values of the accommodation at all ages* Journal of the American Medical Association, pp. 1010-1013

Stereoscopic Rendering simply renders two views, one for each exact eye position and provides a single, correct 3D view relative to the display.

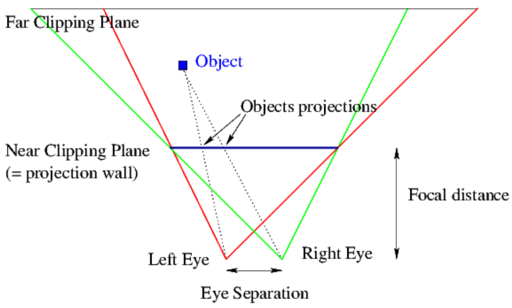


Fig. 2.11 Marten, C. (2011): *Stereoscopic Rendering* tinySG

Technically there are two types of stereoscopic displays:

- **Simultaneous (parallel)**, e.g. HMDs, polarization displays (linear or circular) or autostereoscopic displays (lenticular or parallax barrier): Providing each eye its **own set of pixels**, **color** and **display**.

- **Time sequential (alternate)**, e.g. Active Shutter or Passive Polarized Glasses: Providing each eye its **own time slot**.

2.2 Technical characteristics of Head Mounted Displays

Display Technologies: Today's HMDs are usually built with two LCD, OLED or AMOLED monitors integrated into the glasses.

LCD, short for Liquid Crystal Displays, use polarizing filters and emit linearly polarized light. Liquid crystals do not produce light themselves. LCDs require backlighting. The spiral liquid crystals wind up to a greater or lesser extent depending on the amperage and allow more or less light to pass through. Between two crossed polarizing filters the liquid crystals change the polarization direction of the light depending on their orientation.

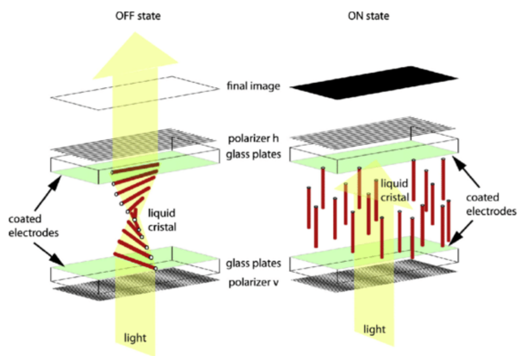


Fig. 2.12 Wuerger, S. M. (2011): *Visual gamma correction for LCD displays*

Organic Light Emitting Diode (**OLED**) is a further development of the light emitting diode (LED). In contrast to LEDs, coloured self-luminous OLEDs consist of organic semiconductors that emit light in an electric field. The cathode generates the electrons which are injected via the Electron Injection Layer (EIL) and drift via the Electron Transport Layer (ETL) to the Emission Layer (EML). In the EML, an organic light-emitting polymer layer, the electrons and the holes recombine to form an electron-hole pair, an exciton. The holes are injected from the hole injection layer (HIL) and drift through the hole transport layer (HTL) to the emission layer (EML) due to the voltage potential between anode and cathode. When the electrons are recombined with the holes, they emit photons as visible light.

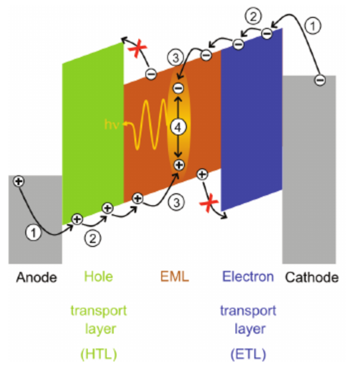


Fig. 2.13 Ruhstaller, B. (2011): *Advanced Numerical Simulation of Organic Light-emitting Devices*

AMOLED means Active Matrix Organic Light Emitting Diode. The single pixels are controlled by an Active Matrix. Each pixel has its own power connection and is thus made to glow. The active matrix is responsible for the electronics and control of the TFT (thin-film transistor) layer. An image is displayed line by line, i.e. sequentially. This technology uses a TFT with a capacitor memory which stores the pixel lines.

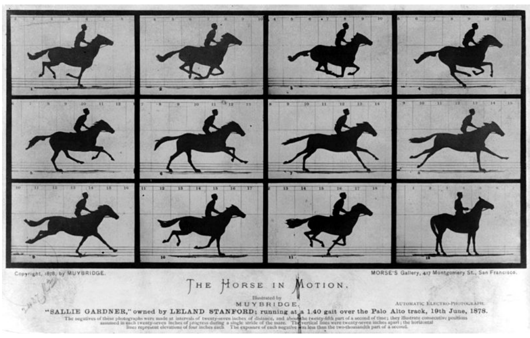


Fig. 2.14 Muggeridge, E. J. (1878): *The Horse in Motion*

The **refresh rate** indicates how many images are displayed in one time unit. It is a characteristic value for the sequence of moving image sequences and for the flicker-free representation. The image repetition in fast sequences is necessary to bypass the sluggishness of the eye and to make moving objects in images appear continuous. The number of frames per second (fps) is calculated from the refresh rate. The critical flicker frequency (CFF) depends on age and lies between 22 Hz and 90 Hz. This is the cut-off frequency at which the human eye can no longer distinguish a periodic light stimulus from a uniformly luminous one. As a comparison: Cinematic film generally runs at a uniform rate of 24 fps. However, *The Hobbit* (2012) was the first film to transition to a higher frame rate (HFR) of 48 fps. This caused the film to look very smooth or 'hyper real'. But it has had a mixed reception from audiences – some people report that this newer frame rate feels less immersive. In VR, a refresh rate of 90 Hz is recommended. Setups that operate below 90 fps are more likely to induce nausea and disorientation (IrisVR[online] 2009). If the frame rate drops down to the 30 to 60 fps range it can lead to jumping and stuttering images and therefore break the immersion even more.

Latency is the delay time that elapses between the occurrence of a particular event and the subsequent event expected to occur. In the information technologies (IT), latency is the travel time required by data from the source to its destination. Although the quantized IT world works at almost the speed of light ($c_0 = 2.99 \times 10^8 \frac{m}{s}$), circumstances such as material processing, shielding, translation of individual components or worn can affect the latency. In order for VR devices to provide an immersive world, the setup requires very low latencies. An absolutely top-notch experience usually equals a latency of 20 ms or less.

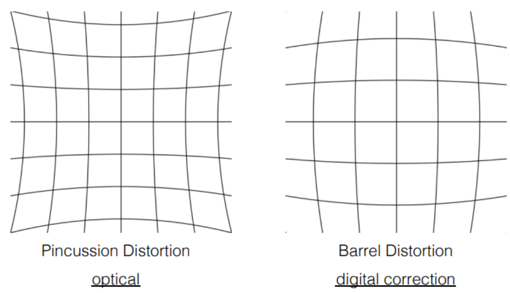


Fig. 2.15 Tykkälä, T. (2013): *Real-time image-based RGB-D camera motion tracking and environment tracking*

Optics: In order to create that immersive feeling of inhabiting a virtual world, a predistorted image has to be stretched (Pincussion distortion) to entirely fill the visual field. This effect is achieved by using a Fresnel lens. It consists of a concentric series of simple lens sections and results in a thin lens with short focal length and large diameter. In doing so, the lens distributes more even resolution area and less distortion.

The quality of the lens is also significant to avoid poor picture quality, low clarity, and unwanted distortions. The pixel-based predistorted image must then be corrected in graphics rendering (Barrel distortion) using linear interpolation.

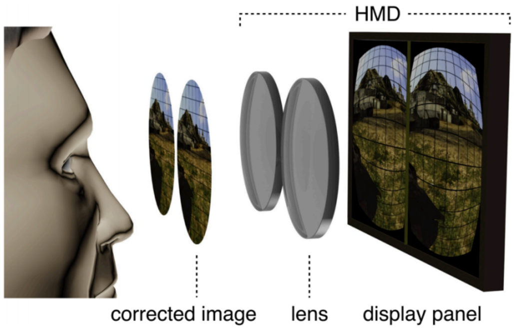


Fig. 2.16 Kilgard, M. (2016): *NVIDIA OpenGL in 2016*

Head Tracking is important for viewing and interaction in VR. In general, there are two different approaches for tracking head movement: 3 Degrees of Freedom (DoF) and 6 DoF. Where 3 DoF only allows turning and rotating (rotation around 3 axes: orientation - yaw, pitch, roll) of the head, 6 DoF allows to walk through (rotation around 3 axes: orientation + translation along 3 axes: position - yaw, pitch, roll, up/down, left/right, forward/back) and explore the environment even more precise. The accuracy of the parameters can be calculated absolute and relative. HMDs accomplish this using an optical process with infrared, gyroscope and accelerometers. The software's speed in turning the head position information into visual information is vital to the immersion of the virtual world. To acknowledge the head's position and transform that into data, the HMD needs an accurate head tracking technology. A distinction is made between two procedures: Inside-Out and Outside-In tracking. It is defined by where the photons originate and where they are received. The Inside-Out procedure places the corresponding sensor on the user whereas the Outside-In tracking procedure places the sensor in the immediate environment. In addition, these procedures can be supported by so-called fiducial markers. These markers are visible objects placed in an imaging system (e.g. camera) as reference points or measuring units in order to improve tracking accuracy.

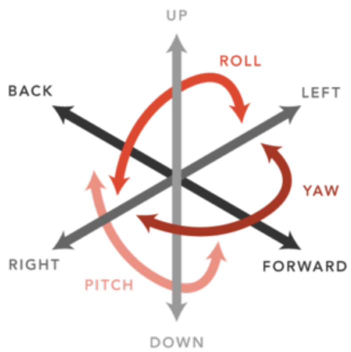


Fig. 2.17 Kilograph.com (2017): *Virtual Reality and 6 degrees of freedom*

Audio Hardware: Some HMDs currently attach their own headphones while others only use a corresponding signal input socket and offer the corresponding headphones as an optional purchase option. Multi-speaker audio that gives the illusion of a 3-dimensional world is a technology that already very much exists as well as the availability of high resolution audio quality.

Computer Hardware: There are two types of HMDs: The first operates on a totally self-contained basis. It possesses all the necessary computer hardware required to function as a VR headset, allowing it to collect and display the input it receives. These standalone devices are generally mobile, battery-powered systems. The second type of HMD does not have any computing power. This means it must interact with an external computer. Usually these PC- or console-tethered HMDs accept an High-Definition Multimedia Interface (HDMI) input and use an Universal Serial Bus (USB) connector to send head tracking data through. Although smartphone hardware is at the point where it can provide reasonable VR experiences, they are still in notable distance behind what is possible with powerful computers and major video games consoles. This means that, in terms of pure visual fidelity and frame rate, dedicated external computers are still the best choice. Also, wireless display links for computers do exist but getting these to work within the ideal latency requirements needs some effort.

Other Hardware: HMDs are made from all sorts of materials: Cardboard, plastic, metal and anything else that will hold the parts together. It's important to consider what adjustments are available on a particular HMD. The adjustment range of the headstrap is important in this regard. If a user wears glasses the HMD should accommodate them or allow for lens adjustments that makes them unnecessary. Finally, the comfort padding and ergonomics of the HMD are often overlooked but very important in my opinion because the device spends a lot of time strapped to the user's head.

Additional Input Devices: At present, the most common way of navigating virtual worlds is with existing video game peripheral devices such as game pads, flight sticks, racing wheels or keyboard and mouse. More immersive devices meant specifically for VR are available such as specialized devices like the HTC Vive or the Valve Index controllers which are even capable of capturing finger movements. At the very high-end there are also full-body suspension and motion tracking systems, like the Xsens 3D suite, hydraulic vehicle simulation rigs, and active mechanical force feedback technology which reacts kin-aesthetically to the virtual environment, like the rumble pad on console controllers.

With the listing of these features there are also some issues which can occur:

- **Leakage**
  Provided images reach the wrong eye.
- **Flicker**
  Insufficient refresh rate is given.
- **Latency**
  A lag measured by the difference between the time there is a signal input and the time it takes the input to display on the screen.

- **Resolution**
  Human visual acuity is maximum between 0.02° and 0.03° possible (Ratnam et al. 2013).

- **Field of view**
  Humans have about 180°.

- **Restricted head movement**
  Wired setups, auto-stereoscopic or linear polarization displays.

- **Accommodation-convergence relation**
  Our focus is always in near field on the screen.

- **3D fusion problems**
  Not everyone can easily fuse stereoscopic images.

- **Depth perception/judgement**
  Underestimation can occur.

2.3 Overview and comparison of current Head Mounted Displays

The range of HMD eyewear on the market has increased continuously in recent years. In order not to lose sight of the bigger picture, I would like to present three devices, each representing a specific product category of VR headsets:

Product category	Standalone	Tethered (console-based)	Tethered (PC-based)
VR device	Oculus Quest	Sony PlayStation VR (ZVR2)	HTC Vive Pro

For a comparison it is worthwhile to investigate the specifications:

	Oculus Quest	Sony PlayStation VR	HTC Vive Pro
Display technology	OLED	OLED	AMOLED
Resolution (in pixels, per eye)	1440 x 1600	960 x 1080	1440 x 1600
Pixel density (in pixels per inch)	538	386	615
Refresh rate (in Hz)	72	120, 90	90
Field of view (in °)	100	100	110
Headphone type	In-Ear	In-Ear	On-Ear
Headphone socket	3.5 mm audio jack	3.5 mm audio jack	USB-C 3.0
Sensors	Accelerometer, gyroscope, internal cameras	Accelerometer, gyroscope	SteamVR tracking, G-sensor, gyroscope, proximity, interpupillary distance sensor
Tracking technology	wireless Inside-Out	Outside-In	marker-based Inside-Out
Additional requirements	-	PlayStation 4	PC with advanced *VR ready*™ specs
Operating platform	Oculus Mobile	PlayStation 4	SteamVR
Release date	2019-05-21	2016-10-01	2018-04-05

Table 2.1 Specification overview of current HMDs

Considering the technical features listed above, I will discuss now which device is most suitable for my research purpose.

2.3.1 Product choice and explanation

First of all, it should be noted that the current HMDs have not yet fully reached the mass market. This is partly due to the high product prices but also to the lack of mass-suited application scenarios and stress symptoms after staying too long in virtual worlds. Thus, I would like to give my test audience an experience as close to reality as possible but only let them wear the glasses for so long that my research series still becomes meaningful and verifiable.
The testers should therefore not only experience the highest possible wearing comfort but also the maximum visual clarity during the test series. The headset by HTC with their Vive Pro with built-in AMOLED screen technology meets this visual demand better than the competitor from Oculus and Sony. Respectively, AMOLEDs are even brighter and more luminous than OLEDs and therefore better suited to maintain a realistic claim.
If I take a look at the per eye resolution of the glasses, I notice that both the Quest and the Vive Pro have the same amount with a total pixel number of 1440 in width and 1600 in height. The VR headset by Sony clearly falls behind at this point with 960×1080 pixels per eye. Here as well, it is important to offer the user a high level of realism. Accordingly, it is worth taking a look at the pixel density of the devices: The Vive Pro with 615 ppi comes out as the clear winner ahead of the Quest with 538 ppi and the PlayStation VR with 386 ppi.
Looking at the refresh rate, it's clear to see that both Sony and HTC guarantee 90 Hz on their devices. Oculus takes a different approach with only 72 Hz. I consider this a risky step because the framerate can also be a reason for enhanced motion sickness symptoms. Sony promises besides regular 90 Hz a frequency of 120 Hz with its headset. Since this is usually not native 120 Hz, - it depends on the imaging application - 60 Hz is operated in the application and 120 Hz can be displayed on the glasses by interpolation and reprojection. The additional 60 frames are generated by taking the HMD tracking data and adjust or warp the last rendered image to fit the movement done since then. So, in order to offer the best possible immersion I will stick to Sony and HTC when it comes to refresh rate.
In terms of audio, the manufacturers are somewhat covered with information. This may be due to the fact that the obvious really visible improvements can be seen in the spatial and temporal resolution of the glasses and therefore be advertised better. On the other hand sound for most potential buyers is a minor argument to purchase. Furthermore, the manufacturers are very similar in their specifications which probably does not produce a USP (unique selling point). In particular, they all use the advertising term HiRes (High Resolution) what stands for a higher sound quality in audio technology. These are digital music formats that have a sampling rate of at least 96 kHz or a sampling depth of at least 24 bits. However, from a scientific point of view an audible improvement in sound quality cannot be proven by a higher sampling rate [(Lüke 1999), (Meyer and Moran 2007)]. The audibility of a sampling depth greater than 16 bits is only given for music with quiet

passages that is reproduced at a very high volume. But what all manufacturers without exception provide is a connection for headphone components. In particular, the Quest and PlayStation VR have a standard 3.5 mm audio jack. The Quest doesn't come with headphones but the PlayStation VR in the 2018 version (CUH-ZVR2) already does. Oculus recommends their costumers to buy their in-ear headphones, Sony already includes them at purchase. The Vive Pro takes a different approach: It comes with fixed on-ear headphones connected to a USB-C 3.0 socket which can also be detached and replaced.

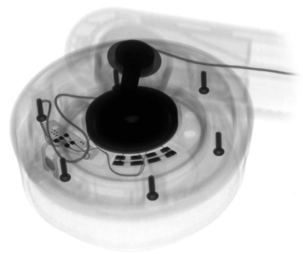


Fig. 2.18 Suovanen, J. (Creative Electron, 2018): *Radiography of the HTC Vive Pro headphone*

A look at the tracking methods of the device manufacturers reveals that they all use a different technology. While Oculus and HTC choose the way of Inside-Out tracking for their products, Sony's headset offers the Outside-In method. Thus, the PSVR uses an external camera connected to the PlayStation 4 via cable to scan the room for the headset and accordingly send the position data to the console. Sony has also recognized that this tracking method is dying out more and more and is already working on an Inside-Out method, as a recent patent application shows (Sony/USPTO[online] 2019). Although both the Quest and the Vive Pro use the Inside-Out technology for tracking, they differ in one characteristic: While the Quest uses a markerless variant, the Vive uses markings on the glasses themselves which are continuously scanned by the Lighthouses via infrared. So, while markerless Inside-Out tracking "intelligently" detects the glasses' position in space with integrated sensors, marker-based Inside-Out tracking does therefore not need a bright environment but tracking markings on the glasses and Lighthouse sensors to detect positional change. In order to achieve the most accurate head tracking possible, I believe that the marker-based tracking variant of HTC works more accurately here.

For the quest, since it's a standalone headset, no further equipment is necessary for a proper use. Whereas the PSVR requires a compatible console such as a PlayStation 4 or newer and the HTC Vive requires a very powerful computer or laptop with special *VR ready*™ certification. However, if I take a look at the operating system, it becomes clear that for a custom-built application to run on the Quest or PSVR, a special build would have to be made for each platform. This is much easier with SteamVR because an application which is already running in the game engine (I use the Unity Editor) can be exported as an executable file (*.exe *Build*) and can be easily accessed via the already installed SteamVR headset configuration. In return, I save a lot of time and nerves not having to generate a special build for different appropriate platforms.

The bottom line is that the performance and accessibility advantages of the Vive Pro dominate those of the competing products. I would have liked to see a build of my application on the Oculus Quest or another comparable standalone device. Nevertheless, due to the limited time and the already existing knowledge regarding VR build workflow in Unity, I put myself in safety with the HTC Vive Pro and its benefits.

Chapter 3

Audio and Spatial Sound

3.1 History and recent Development

Hardly any other specialist area has such a great effect on our everyday life but nevertheless such a low active awareness. For most, sound, both analog and digital, is simply present and completely normal. The first sound recording known to us dates back to 1857 and began with the idea of making sound waves visible. Édouard-Léon Scott de Martinville set himself this goal and developed the first "phonautograph". To make sound visible the phonautograph used a horn connected to a membrane that made an image on a hand-wound cylinder with a pig's bristle. However, the device was still unable to play anything. Not until 1877 Thomas Alva Edison, who became better known for his patent for the first light bulb, succeeded in building a device that could also reproduce sound recordings. The sound was recorded via a membrane with a blunt needle onto a wax roller in deep writing, thus with vertical needle deflection and played back again via a funnel. He received the patent for his "speaking machine" a year later. Only 10 years after this remarkable achievement, two German researchers, Emil Berliner and Werner Suess, recognized some weak points and optimized the device. From now on more easily stackable sound pads made of hard rubber, later of shellac and from 1930 of PVC, were used which was much more cost effective and durable. In addition, there were some changes to the sound box which served to bundle the tones produced in order to make an even better sound possible. With the birth of the gramophone 1987 it became possible to use such devices commercially. Through improvements such as stereophony or quadrophony, records could be found in every household, even today the record still has many lovers. A further development was taken in 1921 with the invention of the optical-sound-processing which enforced against the magnetic sound-recording system because of the lower technical complexity and better durability in continuous operation. For the first time it was possible to record

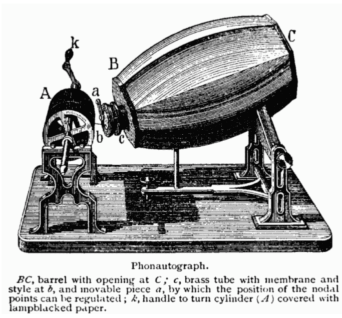


Fig. 3.1 Scott de Martinville, É.-L. (1911): *The Century Dictionary, Vol. VI, Page 4450,* Pholadidae to Phonikon

sound and images simultaneously. The audio track is recorded parallel to the image and later scanned with a photodiode. However, this happens with a delay in time, since images are always passed by the lamp step by step and the sound must be sampled continuously. Today, in addition to analog audio tracks, digital audio tracks are also written on the film. This method was used until digital film projection was established and is still standard in cinemas with film projectors today.
The magnetic sound and image recording was presented in the current sense around 1935 by the AEG-Telefunken AG Berlin. The medium used in their magnetophone is a plastic tape coated with a magnetic emulsion. Due to the poor sound quality, the device was only intended as a voice recorder. However, from the beginning there was a great interest by radio stations which until then still recorded with wax rollers. The frequency range at that time was already 50 to 10.000 Hz and had a dynamic range of about 35 dB. By pre-magnetization this process offered from 1940 a completely new sound quality and was standard because of improvements such as Dolby A or the 48 multi-channel technology in every recording studio for a very long time until they were replaced by digital hard disk recording. From 1976 onwards magnetophones were replaced by cassettes in the consumer sector although they functioned according to the same principle.
The first attempts to digitalize audio signals were already there 1970 before the Compact Disk (CD) was introduced in 1983. Hereby, the audio signal is sampled and stored digitally via PCM (pulse code modulation). PCM is a coded data stream that is generated when analog audio information is converted into digital signals. The sampling rate determines how often the analog signal is picked up and converted into a digital value in one second. The quantization (also called *resolution* or *word width*) indicates how many different digital values a signal can be converted into and therefore it's possible to let different amount of sound data information to be displayed. Examples of common sample rates are:

Medium	sample rate (in Hz)
CD	44.100
DVD	48.000
Blu-Ray and High Definition Recording	96.000
High Resolution Rate	192.000

Table 3.1 Examples of common sample rates

Examples of common bit depths are:

Bit depth	dynamic range (in dB)
16	ca. 96
24	ca. 144
32	ca. 192

Table 3.2 Examples of common bit depths

More dynamic range means a better signal-to-noise ratio (SNR), better precision when mixing and less worrying about headroom. Professional Digital Audio Workstations (DAW) are using an internal bit depth of 32 or 64 bits nowadays by default.

The Australians Peter Vogel and Kim Ryrie were pioneers in this field with their Fairlight audio workstation: The CMI (Computer Musical Instrument), the first digital synthesizer with sampling technology and the MFX 3, the world's first fully digital 24-track disc recorder. The greatest advantage of the digital workflow is the non-linear way of operating recordings and audio data and also the consistency in quality when multiplying and playing back. The last big improvement is the DAW, so to say an integrated recording studio. The digital audio recording made it possible to cut audio tracks directly in the computer. By continuous improvement of the computers it was also possible to emulate more and more studio devices (e.g. via MIDI, short for Musical Instrument Digital Interface, devices), so that today apart from a powerful computer and an interface no other devices are necessary anymore to make good recordings.

Fig. 3.2 Bizzle, J. (2011): *Fairlight CMI II*, National Association of Music Merchants (NAMM) in Anaheim, California (U.S.)

To convert analog signals to digital data values some basic principles need to be explained: The sampling or Nyquist-Shannon theorem says that a signal can be completely reconstructed from its samples taken at a sampling frequency F if it contains no frequencies higher than $\frac{F}{2}$, referred to as the Nyquist frequency:

$$f_{max} < f_{Nyquist} = \frac{F}{2}; \quad \text{i.e.} \quad F > 2f_{max}$$

This equation is referred to as the *Nyquist condition* for perfect signal reconstruction.

The SQNR (signal-to-quantization-noise ratio) reflects the relationship between the maximum nominal signal strength and the quantization error (also known as *quantization noise*) introduced in the analog-to-digital conversion.

$$SNR = \frac{3 \times 2^{2n}}{1 + 4P_e \times (2^{2n} - 1)} \frac{m_m(t)^2}{m_p(t)^2}$$

The SQNR formula is derived from the general SNR formula for the binary PCM communication channel where

P_e is the probability of received bit error,
$m_p(t)$ is the peak message signal level and
$m_m(t)$ is the mean message signal level.

As SQNR, like SNR, is a ratio of signal power to some noise power and can be calculated as:

$$SQNR = \frac{P_{signal}}{P_{noise}} = \frac{E[x^2]}{E[\tilde{x}^2]}$$

The following table uses the signal-to-quantization-noise ratio formula, the Nyquist frequency and bitrate formulas to determine various properties of uncompressed PCM digital audio:

Format	sample rate (in Hz)	bit depth (in dB)	channels	max. dynamic range (in dB)	unique values	folding frequency (in Hz)
CD-DA	44.100	16	2	96.329	65536	22.05 (total bit rate 1411.2 kbps)
DVD-A stereo	192.000	24	2	144.494	16777216	96 (total bit rate 9216 kbps)
DVD-A surround	96.000	24	6	144.494	16777216	48 (total bit rate 13824 kbps)
DTS-HD Master Audio	192.000	24	8	144.494	16777216	96 (total bit rate 36864 kbps)

Table 3.3 Various properties of uncompressed PCM digital audio

3.2 Sound Waves

In physics, sound is a vibration that typically propagates as an audible wave of pressure through a transmission medium such as a gas, liquid or solid. In human physiology and psychology, sound is the reception of such waves and their perception by the brain.
Sound waves are often simplified to a description in terms of sinusoidal plane waves which are characterized by these generic properties:

Period T
Frequency f
Wavelength λ
Amplitude A
Phase Φ

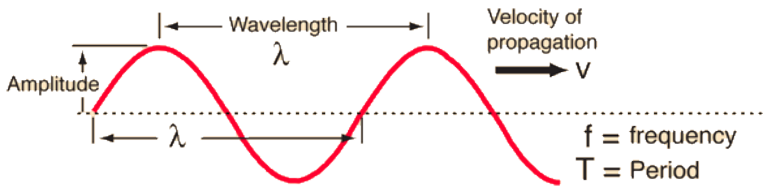


Fig. 3.3 Nave, C.R. (2016): *HyperPhysics:* Traveling Wave Relationship

A single frequency traveling wave will take the form of a sine wave. A snapshot of the wave in space at an instant of time can be used to show the relationship of the wave properties frequency, wavelength and propagation velocity. Sound is a wave of pressure and the basic wave is a sinusoid which means each sound signal can be constructed out of (or decomposed into) multiple sinusoids. The motion relationship $x = v \times t$ is the key to the basic wave relationship. With the wavelength as distance, this relationship becomes $\lambda = v \times T$. Then using $f = \frac{1}{T}$ gives the standard wave relationship

$$v = f \times \lambda.$$

This is a general wave relationship which applies to sound and light waves, other electromagnetic waves and waves in mechanical media.

3.3 Human Audio Perception

The sound pressure waves enter the ear canal and make the ear drum vibrate. This action moves the tiny chain of bones (ossicles – malleus, incus, stapes) in the middle ear. The last bone in this chain "knocks" on the membrane window of the cochlea and makes the fluid in the cochlea move. The fluid movement then triggers a response in the hearing nerve. This is passed on to the brain. The main sound perception sensations are:

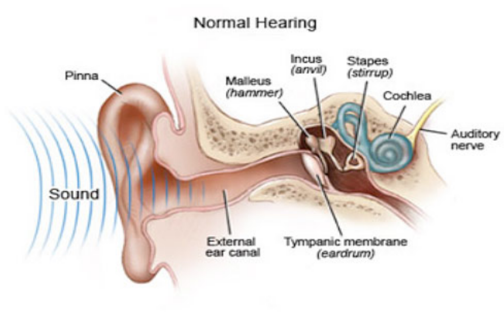


Fig. 3.4 Osborne Head and Neck Institute (2016): *Ear, Nose and Throat Doctors* Los Angeles

- **Frequency f (pitch):** Humans can hear sound waves with frequencies between about 20 Hz and 20.000 Hz and can differentiate between approx. 3000 different tone levels.

- **Amplitude A (volume):** The sound pressure level or volume perception is presented by a logarithmic scale according to the Weber-Fechnersches law [(Weber 1834) (Fechner 1860)]. This basic psychophysical law states that there is a precise numerical relationship between the intensity of a physically measurable sensory stimulus and the strength of the sensation triggered by that stimulus. They found that the physical and the experienced stimulus strength are not in a linear relation to each other but that the intensity of the sensation (I) is proportional to the logarithm of the strength of the triggering stimulus (S):

$$I = c \times \log \frac{S}{S_0}$$

where c = constant of proportionality.

Humans auditory threshold is at 2×10^{-5} Pa = 0 dB at 1000 Hz while the pain threshold is at 120 dB $\widehat{=}$ 20 Pa. Within the limits of the audible range, the sensitivity of the ear depends on frequencies. Normal equal-loudness-level contours (Fletcher and Munson 1933) show that this dependence is particularly pronounced at low frequencies and low sound pressure levels between 2 and 5 kHz (Standardization/TC 43 Acoustics[online] 2003).

- **Phase Φ:** The increase and decrease in pressure cycle of any single vibration.

- **Timbre (tone color):** The perceived quality of any sounds' multiple frequencies changing through time.

- **Localization (3D/spatial sound):** Humans can localize sound, directly in front of them most accurately while to the sides and behind their heads least accurately.

Spatial Sound Perception:
An explanation for binaural hearing, or hearing with two ears, is the duplex theory of Lord Rayleigh from 1907 (Plack 2005). The duplex theory is based on the relationship between the physical properties of a sound reaching the ears and the geometry of the head. Humans have two ears, each positioned at a different location. This causes interaural differences in

- time: **Interaural time difference (ITD)**, in
- intensity: **Interaural intensity difference (IID)** and in
- **spectral differences:** Spectral change of the signal dependent on the angle of incidence due to shape of auricle and external auditory canal. But these differences are not perceived consciously by human.

3.3.1 Interaural Time Difference

The speed of sound in dry air is given approximately by

$$v_{sound\,in\,air} \approx 331.4 + 0.6 \times T_C \,\frac{m}{s}$$

for temperatures reasonably close to room temperature where T_C is the celsius temperature. It would take approximately 0.6 ms for sound to travel the width of an average head [(Feddersen et al. 1957), (Green 1976)]. ITD occurs because sound from a sound source closer to one ear than to the other reaches the facing ear earlier than the averted ear. The position of the head plays an important role in recognizing the transit time differences. In our environment several sounds are perceived simultaneously by our ear. A "cone of confusion" (Goldstein 1996) occurs when several sounds with the same time differences arrive at the ear. In this case, humans cannot locate the individual sound sources. In order to locate the sound source correctly the head has to be moved.

3.3.2 Interaural Intensity Difference

If the wavelengths of the sound are smaller in comparison to the head, they are reflected by it and a sound shadow is formed. This shadowing on the opposite side of the head results in different sound intensities on the left and right ear. These are called interaural level differences. This happens because the head is an obstacle to sound propagation. The averted ear lies therefore in an acoustic shadow and reaches rather lower frequencies and a less intense sound. This only applies to high frequencies because they are slowed down, less intense and do not travel around the head as much as low frequencies do. That's why low frequencies (e.g. base drums) from a concert at a certain distance are still recognizable compared to high frequencies.

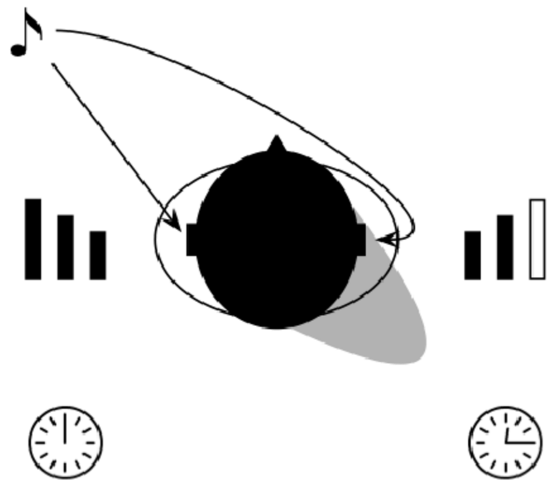


Fig. 3.5 Cohen, M. (2016): *Schematic Summary of Binaural Effects: ITD and IID,* Applications of Audio Augmented Reality: Wearware, Everyware, Anyware, and Awareware

3.3.3 Blauert Bands

Jens Blauert found out that in the natural hearing of sound sources in the median plane, only certain frequency bands raised in level are decisive for the spatial directional localization in front, above and behind because there are no evaluable level and time differences at the ears (Blauert 1974).

Our ears are able to obtain information about the signal spectrum from the wavefront, i.e. within a very short evaluation time. He himself called this psychoacoustic phenomenon *directional bands* or *directional frequency bands*. The illustration shows the frequency-dependent relative commonness of information given by test persons. It can be seen, for example, that in the frequency range around 1 kHz the direction "rear" ("h" ≙ 180°) is detected most reliably with a probability of over 80&.

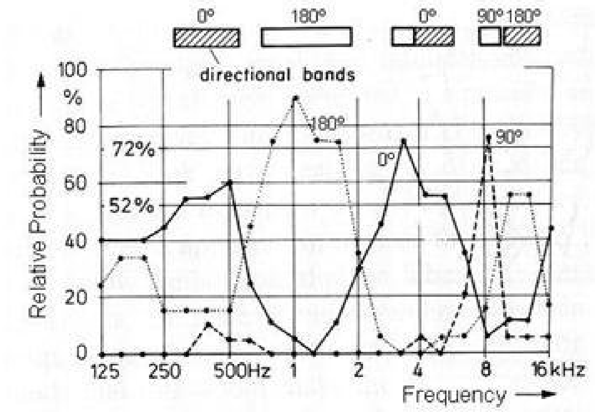


Fig. 3.6 Blauert, J. (1974): *Relative frequencies of observers who give one of the judgements "v", "o" or "h" more frequently than both of the others together (at a 95% level of significance).* Sound Localization in the Median Plane, pp. 209

Together with the auditory canal, the numerous raising and deepenings of the outer ear form an acoustic resonator system that is stimulated depending on the direction of sound incidence. This creates a directional filter whose spectral maxima and minima depend on the direction of sound incidence. These spectral maxima and minima are evaluated by the auditory system and the directions front, top, rear are derived from them. This localization in the median plane is, however, considerably less accurate than, for example, the horizontal localization via time and intensity differences between the two ears. The scientific literature mentions a localization sharpness of about 10° (Makous and Middlebrooks 1990) compared to 1° (Mills 1958) for horizontal localization.

3.3.4 Head-Related Transfer Function

This outer ear transfer function describes for a specific sound incidence angle the sound transmission from the free field at an entrance point of the auditory ear canal. Knowledge of HRTFs are particularly important for further development of binaural signals for VR applications, for acoustic measurements in room designs by architects or for headphone development. HRTF is also used in Sound Retrieval Systems (SRS) where a comprehensive sound field is generated by two loudspeakers which gives the listener the impression of spatial reproduction with the possibility of locating individual instruments

HRTFs are not calculated but measured by putting small microphones in the ears of a dummy head and comparing the recorded with the emitted sound at various locations. These functions refer to the sound that reaches the ear directly or indirectly as reflected sound from walls, ceilings and furniture. The brain also evaluates the reflected, diffuse sound as well as the sound sensation caused by turning, lowering or lifting the head. All sound waves, direct, indirect, reflected and diffuse, determine binaural hearing. The reason why the listener recognizes the direction of the sound source is that the sound on its way to the two ears is filtered according to the direction of incidence. This filtering comes from the reflections and diffractions on the human torso, head and auricles which are described by the HRTFs. The auditory system "recognizes" this filtering and can use it to localize the direction of the sound source. The measured sound pressure amplitudes are different in both ears depending on frequency and sound travel time - and for humans in general similar but different, like fingerprints. Certain computer programs, like the MS HRTF Spatializer plugin, can now simulate such ear signals. To do this, the computer must be able to access averaged data from HRTFs for a large number of sound incidence directions. The figure shows how different the HRTF curves are for different people. A simple arithmetic averaging of the measured values would be wrong because then many subtleties of the curves would be ignored.

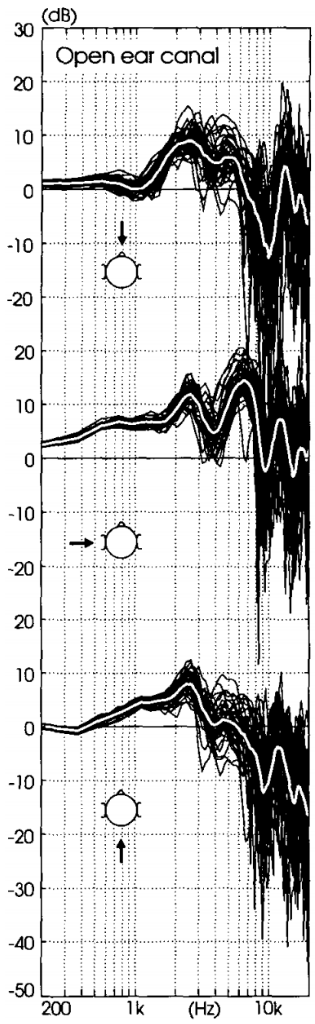


Fig. 3.7 Sørensen, M, Hammershøi, J., (1995): *Left ear HRTFs for open ear canal. White curves represent the evaluated mean values.* AES Vol.43/5, pp. 300-321

3.4 Spatial Game Audio

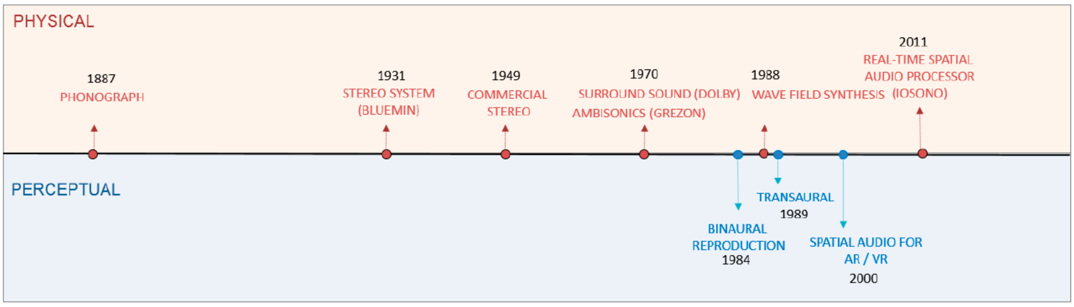


Fig. 3.8 (Hong 2017): *Evolution reproduction systems,* Spatial Audio for Soundscape Design

There are numerous audio technologies of platform companies offering a new dimension into which game sound designers can now work in terms of bringing depth and immersion to the players.
Some of the best known are:

- Binaural recording is an extended form of stereo recording and consists of two microphones that form a system and whose generated intensity differences $_\Delta I$ and/or propagation time differences $_\Delta t$ provide the required audio width on a stereo loudspeaker base. Therefore, binaural recording is the format that is the closest to the human hearing when it's played through calibrated headphones. By recording sound at the eardrum positions, it automatically embeds all the cues needed for sound localization in 3D space and natural alteration of the sound timbre due to propagation, reflection and scattering of sound as it interacts with the human body.

  – **XY-stereophony:** Two microphones with identical directional characteristics, mostly cardioid, for the left and right channel. The sound hits the common axis of rotation as vertically as possible. Different opening angles can be used. The directional characteristics of the microphones - with exception of the "8" - are not ideal but frequency dependent which is unfavorable for the medium recording range (0°).

  – With **coincidental stereophony** a stereo panorama is realized by level differences between left and right. There are no time differences, i.e. no phase differences which means high correlation, good mono compatibility and a good localization sharpness (up to ±5°). The opening angle is the total included angle between the microphone axes. The recording angle is ideally between the center axis and maximum left. In reality this is 18 dB between the center axis and maximum left which makes extreme left-right localization with up to 100% auditory event direction of the loudspeaker base possible.

  – With **MS-stereophony**, the directional characteristic of the center is freely selectable. Contrary to the XY-method, the alignment is made to the sound source which ensures a very good mono compatibility. The directional characteristic of the side is always an "8" and 90° to the center capsule with the in-phase (positive) side facing left. A "pseudo-stereo" is achieved by matrixing:

$$Left = Center + Side$$

$$Right = Center - Side$$

The level of the side signal determines the recording angle and the width of the stereo panorama. The spatiality remains changeable during remixing.

– **Polymicrophony** is based on individually arranged microphones. Very different mixes are possible for different target applications (e.g. CD, surround or television). The main reason for this is the large amount of post-processing possibilities and at the same time good signal separation with little crosstalk (>15 dB) by manipulating the following properties:
  * Volume, Balance
  * Amplification
  * Panorama
  * Equalizer, effects
  * Delay, timing
  * Pitch

  However, a disadvantage is an unnatural depth graduation ("pearl cord effect").

– With **delay stereophony**, a stereo sound is almost exclusively achieved by time differences $_{\Delta}t_{LR}$ between Left and Right. The decisive psychoacoustical factor here is the law of the first wavefront in combination with the Haas effect (Haas 1951): If two identical sound sources emitting coherent sound signals are at the same distance from the listener, a mid-range impression is created. As a result of this so-called sum localization effect, only one fictitious sound source is perceived on the angle bisectors. 60 μs transit time difference $\widehat{=}$ 1 dB level difference whereas a delay difference of approx. 1.5 ms ($\widehat{=}$ 50 cm) is required for an extreme left-right mapping. The use of pressure receivers (spherical characteristic) minimizes distance-related level differences. This results in a better spatial reproduction than with coincident stereophony. However, the disadvantage is a worse localization and monocompatibility than with coincident stereophony.

– **Mixed stereophony**, such as the ORTF (*Office de Radiodiffusion Télévision Francaise*) or the NOS (*Nederlandsche Omroep Stichting*) procedures, offer a fusion of time and intensity differences:

	ORTF	NOS
Base width (in cm)	17,5	30
Opening angle (in °)	110	90
Intensity difference (in %)	60	42
Delay difference (in %)	40	58
Max. recording range for 100% localization (in °)	96	81

Table 3.4 Görne, T. (1994): *Mixed stereophony techniques* Mikrofone: Theorie und Praxis

- A streaming-media Dolby Atmos mix supports up to 128 simultaneous audio objects and includes metadata that describes the location of each one in space. HRTFs are used to generate positional audio cues in a two-channel output signal. A finite impulse response filter is used to process the audio with lower latency. "Dolby Atmos For Headphones" converts this technology into full binaural surround. This virtual headphone surround sound technology was initially developed and marketed by Lake Technology in 1997 and sold in 2003 to Dolby Laboratories.

- For an Ambisonics microphone, a compact tetrahedron arrangement is used in which four microphone capsules with cardioid characteristics point spherically in all directions. This signal is recorded by a recorder with at least four channels which should have as little self-noise as possible and a digital gain control to allow identical pre-amplification of all microphones. In post-production, when converting from A to B format, the audio signal remains four channeled but instead of microphone capsules 1, 2, 3 and 4, the B format signals are now converted to channels W, X, Y and Z. For a given source signal S with azimuth angle θ and elevation angle ϕ, Ambisonics pans the desired four components as:

$$W = \frac{S}{\sqrt{2}}$$

$$X = S \times \cos\theta \times \cos\phi$$

$$Y = S \times \sin\theta \times \cos\phi$$

$$Z = S \times \sin\phi$$

  X, Y and Z are spatial axes while W is a mono-compatible signal that contains all signal components and is therefore omni-directional. One advantage is certainly the spherical microphone arrangement which, in contrast to most surround or immersive audio formats using a hemispherical microphone, also projects height information "from below". In addition, the patents have already expired, so the technology is freely accessible. In VR, a binaural stereo signal is calculated in real time from the Ambisonics signal via HRTF. However, Ambisonic is scene based. Moving away from the starting position is only possible to a limited extent and therefore less suitable for VR than an object-based implementation. Another big disadvantage is that without a decoder the audio will not play back correctly, so the same audio may sound different on different platforms. There is a low compatibility for static stereo sounds (e.g. music), here a workaround is necessary.

- HRTF: DirectX 7.0 already implemented 3D sound in Windows 98. **DirectSound 3D** is part of this package and the interface for hardware-independent spatial sound reproduction. The positions that can be generated with this function library are calculated on a software basis and are of rather moderate quality. If only two speakers are connected to the PC, HRTF leads to inaccurate localization due to the lack of phase rigidity and delay distortion. In conjunction with four speakers, HRTF is an extremely precise model with correspondingly good results. The advantage of the spatial sound systems based on DirectSound 3D is the high level of compatibility with any sound hardware. Nevertheless, with two loudspeakers it's only powerful through additional integrated APIs that create a realistic spatial impression.

  Since July 2016, with version release 1607, Windows 10 supports spatial audio playback via the **AudioGraph** API (application programming interface). It allows to specify a position in 3D space where audio data is put out from an input or submix node.

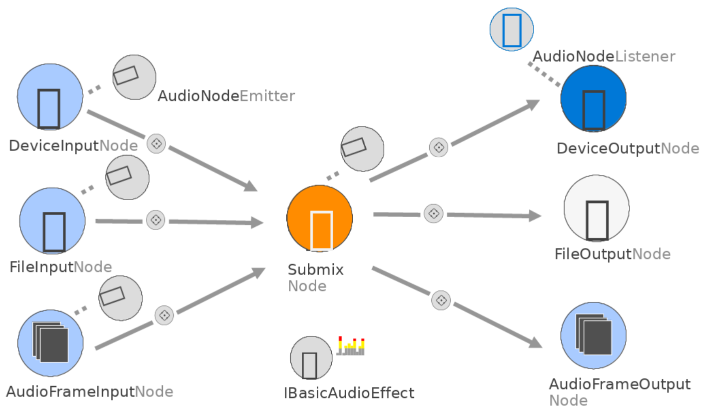


Fig. 3.9 Wilssens, S., Tuliper, A. (2016): *Spatial Audio in Windows 10 RS,* Windows Developer P527 #Build2016

  It's also possible to specify the shape and direction of audio output, specify a speed used for Doppler shifting (Doppler 1842) the node's audio data and define a decay model that describes how sound is attenuated with increasing distance. The class `SpatialAudioModel Enum` specifies the spatial audio `AudioNodeEmitter` class:

```
public enum SpatialAudioModel {1;0;};
```

The class `AudioNodeEmitter.SpatialAudioModel Property` gets or sets a value that specifies the spatial audio processing model used by the emitter:

```
public SpatialAudioModel SpatialAudioModel {get;set;};
```

Fields		Spatial audio processing model
FoldDown	1	Spatial audio is processed using non-spatial methods, such as panning between stereo channels. This provides less realistic spatial sound that preserves the X and Z coordinates of the emitter but effectively ignores the Y coordinate. This model has a lower CPU and memory cost than object-based processing.
ObjectBased	0	Spatial audio is processed using the Microsoft HRTF algorithm. This provides a more realistic spatial sound that preserves X, Y, and Z coordinates of the emitter but at a higher CPU and memory cost than fold-down processing.

Table 3.5 Field selection of the `SpatialAudioModel Enum` class, Windows AudioGraph API

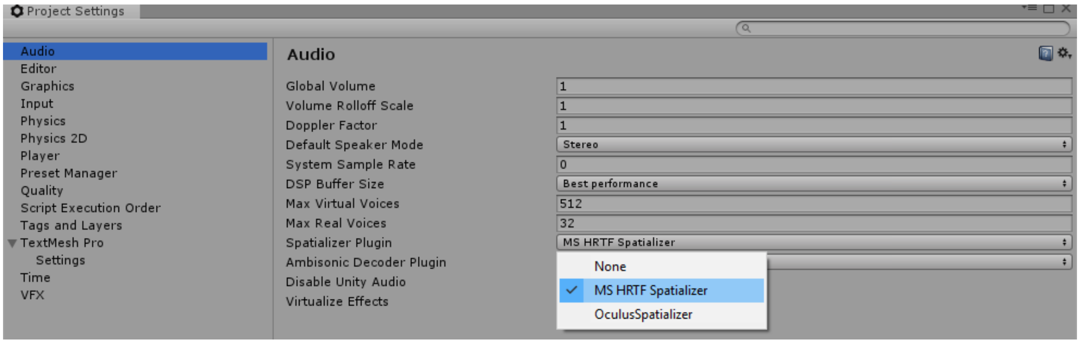


Fig. 3.10 *MS HRTF Spatializer plugin*, Unity Editor audio settings

This API is implemented in Unity via the MS HRTF Spatializer plugin as part of the optional Windows Mixed Reality package. It's possible to apply meta-data effects on the spatializer directly after the AudioSource has decoded audio data. Therefore each instance of the spatializer effect has an own instance of `UnityAudioSpatializerData` associated with mainly data about the AudioSource.

```
struct UnityAudioSpatializerData {
    float listenermatrix[16]; // Matrix that transforms source
                              // position into the local space
                              // of the listener
    float sourcematrix[16];   // Transform matrix of audio source
    float spatialblend;       // Distance-controlled spatial blend
    float reverbzonemix;      // Reverb zone mix level parameter
                              // (and curve) on audio source
    float spread;             // Spread parameter of the audio
                              // source (0 to 360 degrees)
    float stereopan;          // Stereo panning parameter of the
                              // audio source (-1: fully left,
                              // 1: fully right)
    UnityAudioEffect_DistanceAttenuationCallback distanceattenuationcallback;
};
```

3D sound settings that are applied proportionally to the Spatial Blend parameter:

`dopplerLevel`	Determines how much Doppler effect will be applied to the audio source, if it's set to 0 no effect is applied.
`spread`	Sets the spread angle (in °) of a 3D stereo or multichannel sound in speaker space.
`rolloffMode`	Sets how fast the sound fades. The higher the value, the closer the Listener has to be before hearing the sound. (This is determined by a graph: logarithmic, linear or custom).
`minDistance`	Within the `minDistance`, the sound will stay at loudest possible. Outside `minDistance` it will begin to attenuate. Increasing `minDistance` of a sound makes it 'louder', decreasing makes it 'quieter' in a 3D world.
`maxDistance`	The distance where the sound stops attenuating at. Beyond this point it will stay at the volume it would be at `maxDistance` units from the listener and will not attenuate any more.

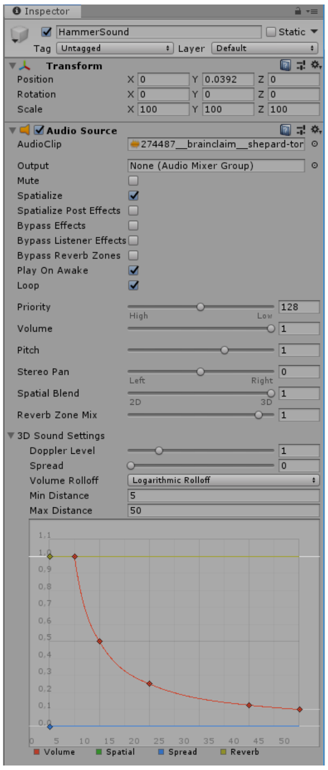


Fig. 3.11 *Audio source 3D sound settings*, Unity Editor GameObject inspector

```
public AudioSource Play(AudioClip clip , float volume, float pitch) {
            AudioSource hammerSound = this.AddComponent<AudioSource >();
            hammerSound.rolloffMode = AudioRolloffMode.Logarithmic;
            hammerSound.spatialBlend = 1;
            hammerSound.minDistance = 5;
            hammerSound.maxDistance = 50;
            hammerSound.spread = 0;
            hammerSound.dopplerLevel = 1f;
            hammerSound.reverbZoneMix = 1;
            hammerSound.clip = clip;
            hammerSound.volume = volume;
            hammerSound.pitch = pitch;
            hammerSound.Play ();
            Destroy (hammerSound, clip.length);
            return hammerSound;
        }
```

- Spatial Audio SDKs, like **Google Resonance Audio** which offers reference implementation of Ambisonics formats (Ambix ACN/SN3D), encoding, sound field manipulation and decoding techniques as well as HRTFs. Also, an entire library of optimized DSP (digital signal processor) classes and functions which includes resamplers, convolvers, filters, delay lines, spectral reverb and other DSP capabilities. Valve's **Steam Audio** SDK offers more or less the same in different wording: Realtime sound propagation, occlusion and reflection, HRTF-based low latency 3D audio and physics-based Ambisonics rendering. Supported platforms are Windows, macOS, Linux and Android.

In reality, surround sound formats and 3D sound exist as long as games consoles and PCs themselves do (e.g. *QSound* by Sega, 1993). So, that these newer immersive formats offer something revolutionary is easy to dismiss.
Traditionally, surround sound, at least in cinema, has always been about spectacle, particularly in a theatrical setting. From innovative 5.1 surround movies like *Apocalypse Now* (1979) to full Atmos mixes like *Gravity* (2013), all have celebrated the visceral sound in a shared fictional space. By physically moving the air around the audience, most notably in the case of low frequencies but also through positioning or moving sound around them, sound brings real spatial dimensionality to a cinematic experience that is simply not possible through the screen light alone – the more physical, impactful and convincing that sound 'scenery' is, the greater the potential emotional reactions are. With Dolby Atmos, the biggest innovation is the inclusion of overhead speakers into the playback array which creates a much more immersive sphere effect around the listener. So, now the audience no longer hears a flat disc of sound like in a 7.1 setup, they hear positional sounds in a full 3D arc above them too. There are obvious some challenges to this kind of spatial sound approach when done in cinema:

> *I assume that* [director] *Cuarón was going for a certain kind of "immersive realism" by using this aggressive approach to surrounds, but for me it sometimes backfires, and actually makes scenes less "immersive" by yanking the listener out of the water every few minutes so that an unseen and unmotivated cat can yeowl, car horn can honk, or a dialog line can come from the surrounds.* (Thom[online] 2019)

If a sound comes from above, this is a specific hot zone in which humans have some very hard-wired instincts, exactly the same is the case for a sound behind them. And this is the fundamental difference between video games and cinema. If a gamer hears a sound behind or from above, he actually can move the camera and look up or behind. In this sense for games, spatial sound has a deep meaning and added value, it serves gameplay, story, and encourages exploration. I see video games as a major driver of this new aesthetic of immersive and spatial surround sound.
Since it is sometimes a bit intransparent which game uses which spatialization technology on which platform, I came across a collection thread in the reddit community which is listed on the following pages:

Game	Initial Release Date	Platforms	Notes
Hellblade: Senua's Sacrifice	August 8, 2017	PlayStation 4, Xbox One, Nintendo Switch, Microsoft Windows	Binaural Recordings of Dialogues and maybe Ambisonics Game Audio SDK for environmental sounds
PlayerUnknown's Battlegrounds	March 23, 2017	PlayStation 4, Xbox One, Microsoft Windows, Android, iOS	Gunshots and main elements use HRTF for 3D Spatial Effect (Only On PC and Console; later added in updates)
Counter Strike: Global Offense	August 21, 2012	PlayStation 3 , Xbox 360, Microsoft Windows, Linux, macOS	Gunshots and main elements use HRTF for 3D Spatial Effect (one of the first; later added in updates)
Minecraft	May 17, 2009	Microsoft Windows, Linux	Uses OpenAL library for sound, so HRTF can be enabled
Mass Effect: Andromeda	March 21, 2017	PlayStation 4, Xbox One, Microsoft Windows	Dolby Atmos for Headphones
Battlefield V	November 9, 2018	PlayStation 4, Xbox One, Microsoft Windows	enable audio renderer ('3D Headphones' preset in the audio options)
Sniper Elite 4	February 14, 2017	PlayStation 4, Xbox One, Microsoft Windows	Optional Binaural HRTF Conversion in Sound Settings
Shadow of Tomb Raider	September 14, 2018	PlayStation 4, Xbox One, Microsoft Windows	Mixed in Dolby Atmos setup but HRTF conversion for 3D sound on headphones is super realistic and amazing
Ring of Elysium	September 19, 2018	Microsoft Windows	Enable HRTF and 3D Audio Spatialization In Sound Options
Final Fantasy XV	November 29, 2016	PlayStation 4, Xbox One, Microsoft Windows	Dolby Atmos for Headphones
Battlefield 1	October 21, 2016	PlayStation 4, Xbox One, Microsoft Windows	Dolby Atmos for Headphones

Table 3.6 *Non-VR Games with 3D Spatial Audio* (r/GameAudio[online] 2019)

Game	Initial Release Date	Platforms	Notes
Budget Cuts	June 14, 2018	Microsoft Windows	Uses Steam Audio
PROZE: Prologue	28.08.2018	Oculus Rift, HTC Vive	Uses Auro 3D
Rick and Morty: Virtual Rickality		PlayStation 4, Microsoft Windows (HTC Vive, Oculus Rift)	
VR Battle Grid		HTC Vive	
Job Simulator		PlayStation 4, Microsoft Windows (HTC Vive, Oculus Rift)	
Far Beyond: A Space Odyssey VR		Microsoft Windows (HTC Vive)	
P·O·L·L·E·N		Microsoft Windows (Oculus Rift, HTC Vive)	
KFC The Hard Way		Oculus	
Cosmic Awakening VR		Oculus Rift, HTC Vive	
Cargo Breach		HTC Vive	
VR Battle Grid		HTC Vive	

Table 3.7 *VR Games with 3D Spatial Audio* (r/GameAudio[online] 2019)

Chapter 4

Kinetosis (Motion Sickness)

The trigger for the motion sickness is a sensory insufficiency between the sensory impressions of the eyes and the organ of balance which signalize movement and rest at the same time in contradiction to each other. It is not necessary for movement to actually take place: Occuring nausea can also be triggered by visual stimuli in an observatory, a cinema or even VR. Between the sensory organs, their emotional coloring and processing, the brain stem and the nerve plexuses of the gastrointestinal tract, there are complex control loops that can easily be disturbed [(Golding 2006b), (Grundy 2006), (Lewis, Hitchcock, and Sullivan 2004)]. In the first phase of nausea, conflicting sensations lead to the following symptoms:

- Increase in skin conductivity and heart rate
- Sweating, cold sweat, dizziness, tiredness, nausea
- Dry mouth
- Breathing difficulties, anxiety

Furthermore, kinetosis is also aggravated by:

- Negative expectations like fear
- Alcohol and drugs
- Food containing calories and protein
- Menses or pregnancy
- Smoke and kitchen smells
- Uncomfortable sitting position where the eyes cannot follow the outer movement.

Keeping balance and orientating oneself in a world consisting of up and down, front and back, right or left is not an easy task for the human body. For this achievement, the human body therefore makes use of different information which are provided by different senses:

- Optical system: The information provided by the vestibular system is supplemented and balanced by the stimuli that the eyes receive and send to the brain.
- Proprioceptive system: The receptors of the joints and muscles constantly transmit information about the position of the arms and legs but also of all other parts of the body and thus help to regulate the body's statics. Part of this proprioceptive system are also pressure sensors in the skin which inform the brain about the posture of the body, the ground and the speed of movement.
- Vestibular system: Organ of balance (see below)

The higher authority, the centre of balance, is located in the brain stem. Here, with the help of incoming information, the current position in space is calculated and compared with movement sequences that humans have learned and stored the evolution. Commands that ensure that the balance is maintained are sent to the muscles and eyes which then make the necessary corrective movements and thus assure that the balance is maintained.

4.1 Vestibular system

The information that the sense of balance receives from the optical, proprioceptive and vestibular systems in the central nervous system (CNS) and processes is supplied to it by the vestibular organs which are located in the inner ear at the petrous bone, one each right and left. This is a small "apparatus" with one large and one small atrial sac and three arch-shaped ducts.

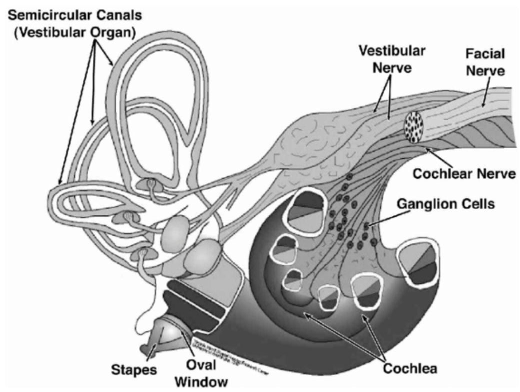


Fig. 4.1 Virginia Merrill Bloedel Hearing Research Center (2008): *Diagram of the inner ear*

These are partly filled with lymphatic fluid and are in right angles to each other - much like the three sides of a cube meeting in one of its corners. On the inner walls there are highly sensitive tactile or sensory hairs, each of those is connected to a nerve. With every change of position of the head, no matter how small, the liquid moves in the arcades. This stimulates the sensory hairs which transmit these stimuli to the center of balance and the cerebellum. These archways are responsible for the perception of rotational movements of the head (or the body including the head) in space. If a person rotates around his or her own axis for a longer period of time and suddenly stops, the fluid in the vestibular organ continues to rotate for a while, creating the impression of an opposite rotation.
Problems such as dizziness arise when the information that arrives in the CNS

appears to be or actually is contradictory and therefore does not fit together properly. Of course, they also occur when correct information arrives but the central switch point in the brain cannot process it properly due to dysfunction.

4.2 Motion Sickness Susceptibility Questionnaires (MSSQ)

With the MSSQ, each subject estimates his or her susceptibility to motion sickness in nine different situations with body motion: In cars, on buses, on trains, in airplanes, on small boats, on ships and rides, on swings and carousels on playgrounds, on roller coasters or similar vehicles in amusement parks. The data is collected for the last ten years of the test subject. These movement situations are measured with the scale criteria categorized from "n" (*not applicable*), "0" (*never felt sick*) to "4" (*always felt sick*).

	not applicable	never felt sick	rarely felt sick	sometimes felt sick	frequently felt sick
Cars					
Buses or coaches					
Trains					
Aircraft					
Small boats					
Ships					
Swings in playgrounds					
Roundabouts in playgr.					
Funfair rides					

Table 4.1 Vulnerability questionnaire of the MSSQ (Golding 1998).

In the original questionnaire according to Golding, this vulnerability is also asked separately for the stage of childhood until the age of 12. Since my test audience is very young anyway, I have decided not to query for the childhood and only draw conclusions about the susceptibility of the last 10 years. The total score is then calculated as follows:

$$MSSQ\,short\,raw\,score = (total\,sickness\,score) \times \frac{(9)}{(9-n)}$$

n = number of types not experienced

The possible range starts from minimum "0" (*never felt sick*) to maximum "27" (*always felt sick*) whereby it is rather unlikely that the maximum value will be reached.

4.3 Simulator Sickness Questionnaire (SSQ)

In the literature there are various theories on the origin of the simulator sickness. The most widely accepted is the so-called *Sensory Conflict Theory* (Reason 1975). It was first developed to explain motion sickness and later extended to simulator sickness. Its basic assumption is that the human orientation in 3D space is provided to the CNS by four sensory inputs:

- Archways
- Macular organs
- Visual system
- Proprioceptive system

The occurrence of simulator sickness is explained by a mismatch between what the sensory system expects on the basis of previous experiences and what is actually perceived in the simulator. This discrepancy causes an internal conflict that cannot be resolved and ultimately causes complaints. Thus, such a conflict will occur when my test subjects are "driving" in my fixed VR setup: The visual information on the projection screen indicates a linear acceleration but since the users themselves are not moving, the corresponding vestibular stimuli are missing.
In the *Postural Instability Theory* (Riccio and Stoffregen 1991), motion sickness is not attributed to a sensory conflict but to a longer lasting instability in postural control. Postural instability is defined as a state in which uncontrolled movements of the perceptual and action systems are reduced to a minimum.
The SSQ is the best known method for measuring simulator sickness. It is based on a version of the Motion Sickness Questionnaire [MSQ; (R. Kennedy and Graybiel 1965)] that covered 28 different symptoms. The underlying data set consisted of n = 1.119 pairs of pre- and post-measurements from different studies on ten different flight simulators. A first goal of the questionnaire developers was to identify the symptoms among the 28 used symptoms that lead to systematic changes of the initial state due to simulator exposure. For this purpose symptoms were excluded which

- occurred with a frequency of less than 1%,
- showed no changes in frequency or severity or
- proved to be irrelevant or unspecific (e.g. boredom).

A total of 12 of the 28 symptoms were excluded. The following table shows those symptoms that were included in the SSQ from the MSQ and which I also used to query my test participants:

	none	slight	moderate	severe
General discomfort				
Fatigue				
Headache				
Eyestrain				
Difficulty focusing				
Increased salivation				
Sweating				
Nausea				
Difficulty concentrating				
Fullness of head				
Blurred vision				
Dizziness (eyes open)				
Dizziness (eyes closed)				
Vertigo				
Stomach awareness				
Burping				

Table 4.2 Symptoms of the SSQ (R. S. Kennedy et al. 1993)

The 16 symptoms of the SSQ are assessed on a four-level scale. The degree of severity of each symptom is evaluated by the subject as "none", "slight", "moderate" or "severe" immediately before and after both test runs in the VR simulator application. The numerical value "0", "1", "2" or "3" is assigned in the same order. The total value is calculated by adding the respective number of points and dividing it by 16. Accordingly, three numerical values are calculated, one before the first run (*pre test*), one between both runs (*post stereo* or *post spatial*) and one after the whole VR experience (*post spatial* or *post stereo*). The calculated value then gives information about how strong the simulator sickness symptoms are after the experience in terms of stereo compared to spatial audio within the range of "0" (none) to "3" (severe).
The original SSQ categorized the individual symptoms into corresponding superior symptom groups. Each symptom receives a factor according to a charge matrix (main axis analysis; Varimax rotation). The three resulting factors - nausea, oculomotor and disorientation - describe the effects of simulator exposure on different target systems in humans. Furthermore, each factor forms the basis for a sub-scale of the SSQ. By summing up the symptom values in the corresponding factor groups a total value is obtained by multiplication with the respective factor scale weight. For the total score, the total values are then added and multiplied by the weighting factor 3.74. I did not use this final factorization grouping because I wanted to present a rather general picture of my test subjects' constitution without specifying into the corresponding categories.

A large number of studies show that simulator disease cannot be attributed to a single cause but that a large number of different factors contribute its occurrence. A distinction can be made between the following relevant influencing factors (Kolasinski 1995):

- Individual factors: Age, gender, disposition, concentration, ethnicity, experience with real task, simulator experience, flicker fusion frequency, ability to mental rotation, perception style, postural stability, health status, personality traits.
- Simulator-related factors: Calibration, field of view, viewing region, scenario content, fluorescent delays, repetition rate, flicker, brightness, with or without motion platform, color, contrast, image repetition rate, time delay, stereo- or monoscopic displays, screen resolution, interpupillary distance, positional detection error.
- Task- or exercise-related factors: Extent of control, duration of exposure, head movements, acceleration, height above ground, illusory self-movement, speed of self-movement, type of locomotion, unusual maneuvers, type of application, sitting or standing, global pattern of optical flow.

Chapter 5

Experimental Setup Description

5.1 Research Components

For my experimental setup I work with the following devices:

- Certified *VR ready*™ desktop computer:
  System type: 64-bit
  System manufacturer/model: Supermicro C7X99-OCE
  OS: Windows 10 Enterprise Build 17134, release version 1803
  CPU: Intel Core i7-5960X, 3 GHz (16 CPUs)
  RAM: 32.768 MB
  GPU: NVIDIA GeForce GTX 980 Ti, 6.097 MB GDDR5
  DirectX: Version 12
  Sound: Microsoft High-Definition digital audio device (S/PDIF)
- HTC Vive Pro
- Unity Editor (64-bit) personal version, release 2019.2.3f1 (August 28, 2019)
- PULOX PO-300 pulse oximeter

Fig. 5.1 Schleise, V. (2019): *Impression during test cycle I*

5.2 Metrics

I use the following metrics to make my test procedure empirically measurable and reproducible:

- Susceptibility indicators (MSSQ raw score): 9 à 0 to 3 (in total 0 to 27)
- Symptom indicators (SSQ score): 16 à 0 to 3
- Pulse rate measurement at rest and during the runs.

Fig. 5.2 *Subject data and MSSQ query,* research questionnaire page 1

Fragebogen zur Anfälligkeit von Motion Sickness

Alter: ………..   Geschlecht: [ ] männlich [ ] weiblich

Dieser Fragebogen wurde entwickelt, um herauszufinden, wie anfällig Sie für Bewegungskrankheiten sind und welche Arten von Bewegungen am effektivsten sind, um diese Krankheit zu verursachen. Krankheit bedeutet hier, sich unwohl oder angewidert zu fühlen oder sich tatsächlich zu erbrechen.

Bitte Ihre Erfahrung für jede der folgenden Transportarten angeben wie oft Sie sich krank oder angewidert in den vergangenen zehn Jahren fühlten:

	Niemals gereist	Niemals krank gefühlt	Selten krank gefühlt	Manchmal krank gefühlt	Häufig krank gefühlt
Autos					
Busse oder Reisebusse					
Züge					
Flugzeuge					
Kleine Boote					
Schiffe, z.B. Kanalfähren					
Schaukeln auf Spielplätzen					
Karussell auf Jahrmärkten					
Achterbahnfahrten					
	n	0	1	2	3

Fig. 5.3 *SSQ query*, research questionnaire page 2

Kreuzen Sie jene Symptome an, die im Augenblick auf Sie zutreffen:

	nicht	leicht	mittelschwer	schwer
Allgemeines Unwohlsein				
Ermüdung				
Kopfschmerzen				
angestrengte Augen				
Schwierigkeiten, scharf zu sehen				
erhöhter Speichelfluss				
Schwitzen				
Übelkeit				
Konzentrationsschwierigkeiten				
Kopfdruck				
verschwommene Sicht				
Schwindel (Augen offen)				
Schwindel (Augen geschlossen)				
Gleichgewichtsstörungen				
Magen macht sich bemerkbar				
Aufstoßen				

Fig. 5.4 *Closing questions and feedback*, research questionnaire page 3

Abschlussfragen

Welche Farbe hat das Auto zu Beginn des Testlaufs?	[ ] rot [ ] grün [ ] gelb [ ] blau
Wie viele Feuer haben Sie während der Fahrt wahrgenommen?	[ ] 1 [ ] 2 [ ] 3 [ ] 4
An wie viele musikalische Audioquellen können sie Sich erinnern?	[ ] 2 [ ] 3 [ ] 4 [ ] 5
Wie viele Minuten hat die Testfahrt in Anspruch genommen?	[ ] 2 [ ] 3 [ ] 4 [ ] 5
War das Ihre erste Erfahrung mit Virtual Reality?	[ ] ja [ ] nein

	....	....
Welchen Durchlauf haben Sie angenehmer wahrgenommen?	[ ] 1	[ ] 2
In welchem Test haben Sie eine höhere Immersion (Grad des Eintauchens) gefühlt?	[ ] 1	[ ] 2

Warum?

____

Gibt es noch Anmerkungen von Ihrer Seite?

____

Vielen Dank für Ihre Teilnahme!

Test track

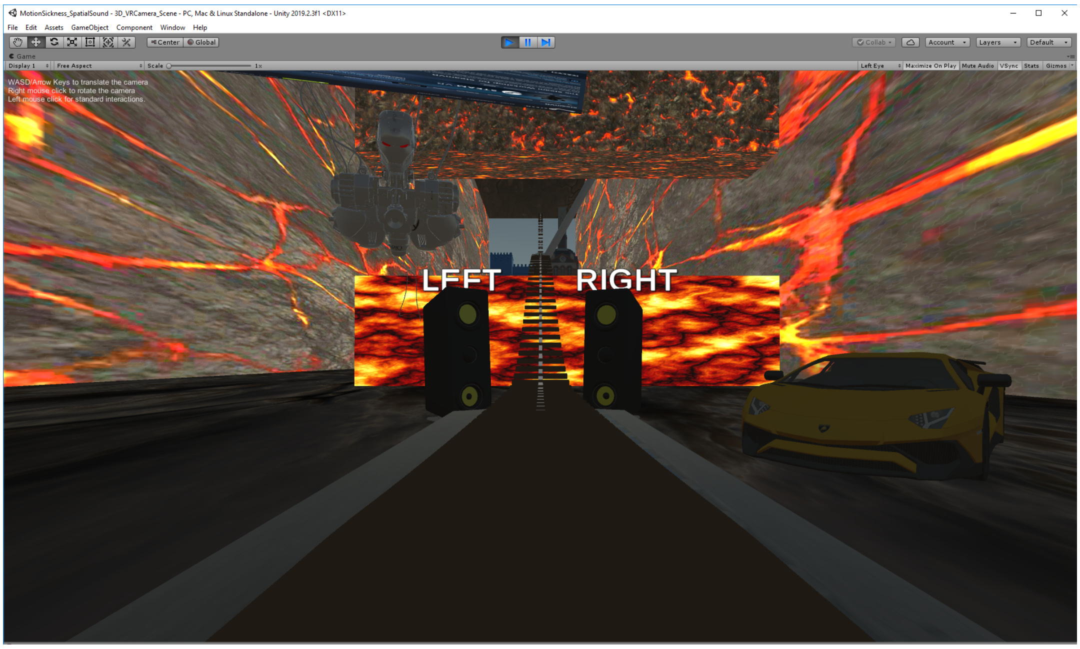


Fig. 5.5 *In-game POV sight*, test track

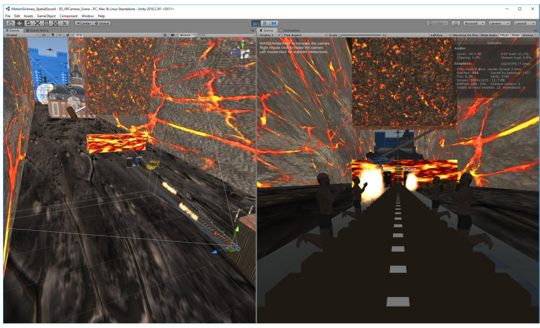


Fig. 5.6 *Scene/game view I*, test track

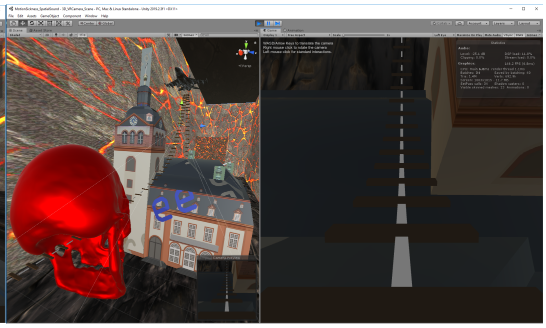


Fig. 5.7 *Scene/game view II*, test track

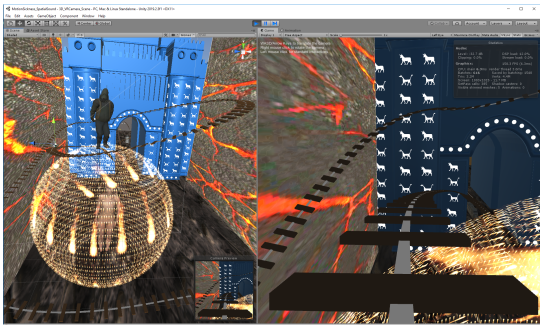


Fig. 5.8 *Scene/game view III*, test track

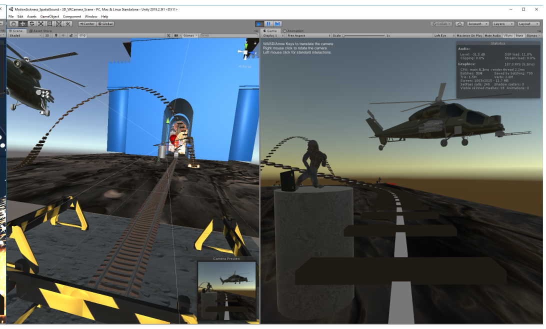


Fig. 5.9 *Scene/game view IV*, test track

During the research procedure the *VSync* option in the Unity built version was activated to ensure a constant refresh rate between approx. 89 and 92 fps.

Listing of audio sources used:
(in order of appearance)

No.	sound type*	description
1	dynamic ambient	Wind
2	dynamic GameObject	Engine
3	dynamic GameObject	Zombie sound
4	dynamic GameObject	Skeleton
5	static GameObject	Praying men whisper voice
6	static GameObject	Fire
7	dynamic GameObject	Cyborg voice
8	static GameObject	Lamborghini Aventador engine
9	static GameObject	Audio lineup & ID speaker sound
10	static ambient	Dark ambient gateway sound
11	dynamic GameObject	Hammer sound Shepard tone
12	static ambient	Castle church atmosphere
13	static GameObject	Choral chanting
14	dynamic GameObject	F16C combat aircraft engine
15	static GameObject	Dark halloween skull voice
16	static ambient	Picture This - One Drink (live music)
17	static GameObject	Police siren
18	dynamic GameObject	Train passing
19	static ambient	Smilla - Play (Worakls Remix) music
20	static GameObject	Rifle shots
21	dynamic GameObject	Helicopter engine rotor blades

Table 5.1 *Listing of audio sources used*, test track

5.3 Hypothesis

It seems humans live in a disorderly universe. But if we look behind the facade and start to numerically record naturally occurring phenomena, we find many forms and fundamental structures of nature which can be mathematically described with the help of fractals. Fractals are objects in which the whole resembles its components. This is especially visible in trees, crystals, long polymer molecules but also in clusters of galaxies or in a visualized heart frequency.

> *Clouds are not spheres, mountains are not cones, coastlines are not circles and bark is not smooth, just as thunderstorms do not travel in a straight line.*
>
> (B. Mandelbrot 2006)

Accordingly, I believe that humans are quite capable of differentiating between orderly and irregular sensory impressions - even if subconsciously. I hope to find

*static: fixed location, dynamic: positional change
GameObject: audio sounds from a specific item, ambient: audio sounds in a specific area

measurable evidence that the human balance perception in VR is more strongly affected by unnatural auditory stimuli in form of a fixed stereo audio setup than it is the case with dynamic environmental surround sound that comes closest to natural hearing.

5.4 Execution

After a call for test participation, a suitable time is determined with possible interested parties. My two test groups completed their tests in the period from November, 13 to December, 3 2019. When individually contacting the interested parties to make an appointment, I asked those to come sober, in a stable physical and mental condition and not to eat any food rich in calories or protein immediately before the start of the test scenario in order not to aggravate simulator sickness symptoms artificially as described in chapter 4.
The test procedure is divided into 9 time phases:

1. Welcoming and introduction.
2. Test subject indicates susceptibility via MSSQ (questionnaire page 1) and gives information about current symptoms via SSQ (questionnaire page 2).
3. Test subject is applied pulse oximeter, measurement of the pulse rate begins.
4. Test subject goes through first run in VR application (either stereo or spatial audio version).
5. Test subject indicates current symptoms via SSQ (questionnaire page 2) by highlighting differences to the initial ideal condition.
6. Test subject goes through second run in VR application (either spatial or stereo audio version).
7. Pulse oximeter is taken from test subject.
8. Test subject indicates current symptoms via SSQ (questionnaire page 2) by highlighting differences to the initial ideal condition and fills in the closing questions (questionnaire page 3).
9. Acknowledgment and farewell.

5.5 Test Results

In total my test procedure has 38 people, divided into two groups of 18 and 20 people. It was important for me to create a data pool as homogeneous as possible. Therefore, I always paid attention to have both genders equally represented. In addition, I have always paid attention to the balance of the test order. Thus, both

test groups had the same proportion of those who experienced either the stereo or the spatial variant first.
In the first test group the participants have an age range of 20 to 74, in the second 19 to 50 years.

	Research series I	Research series II
Participants	18	20
Enumeration*	A1 to A9, B1 to B9	X1 to X10, Y1 to Y10
Age	20 to 74	19 to 50
Female	8	10
Male	10	10
VR experienced	9	12
Avg. MSSQ raw score	7.33	6.15
Avg. post Spatial SSQ score	0.76	0.46
Avg. post Stereo SSQ score	0.93 ($\nearrow$ 5.56%)	0.55 ($\nearrow$ 2.92%)
Spatial more pleasant	13	15
Spatial higher immersion	10	15
Avg. correct answers	1.2	1.2

Table 5.2 Data record of both test series

In test group II, a total of 12 people have slightly more VR experience than 9 in the first group. Basically, the spatial audio setup was perceived more pleasant than the stereo setup. In test group I 13 (72%) of 18 subjectively found spatial audio more pleasant, 10 (55%) of 18 felt a higher immersion. In test group II, 15 people each (75%) felt spatial audio more pleasant and immersive. With regard to the motion sickness susceptibility, it is noticeable that test group I at 7.33 is somewhat more susceptible than test group II at 6.15. Consequential, I observe in test group I an increase of 5.56% comparing post stereo to post spatial symptoms whereas the clarity of symptoms in the second test group is only 2.92%. It is also interesting to note that in both test groups a value of about 1.2 correctly answered questions on four asked questions is recorded.

*Information on test order:
A and X: 1. run Stereo Mixdown, 2. run Spatial HRTF
B and Y: 1. run Spatial HRTF, 2. run Stereo Mixdown

Fig. 5.10 *Mass distribution pulse rates*, Stereo Audio run test group I (n = 18)

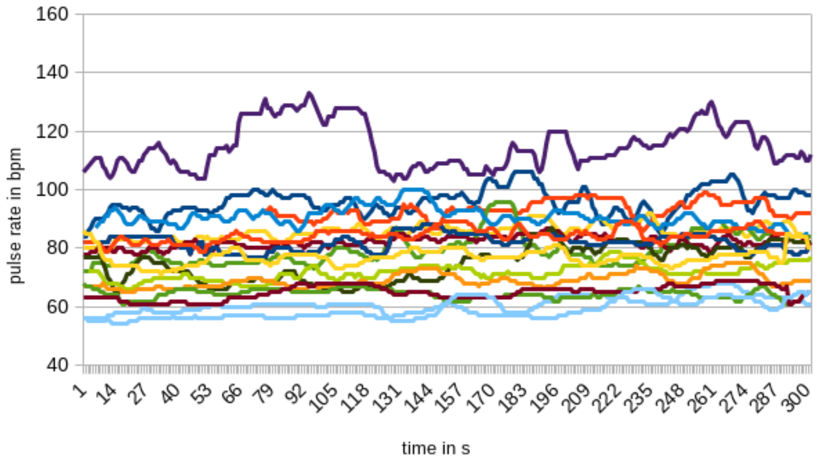


Fig. 5.11 *Mass distribution pulse rates*, Spatial Audio run test group I (n = 18)

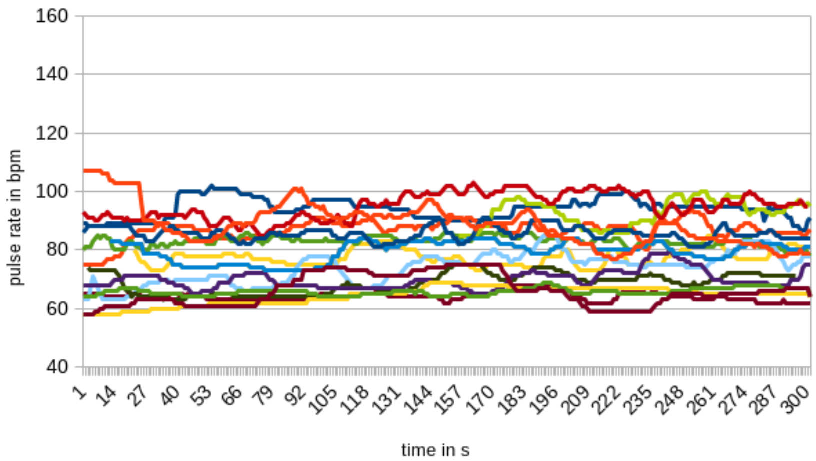


Fig. 5.12 *Mass distribution pulse rates*, Stereo Audio run test group II (n = 20)

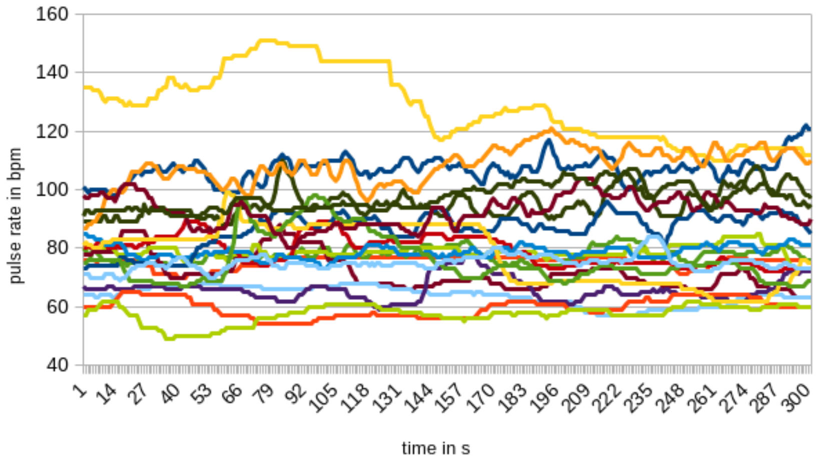
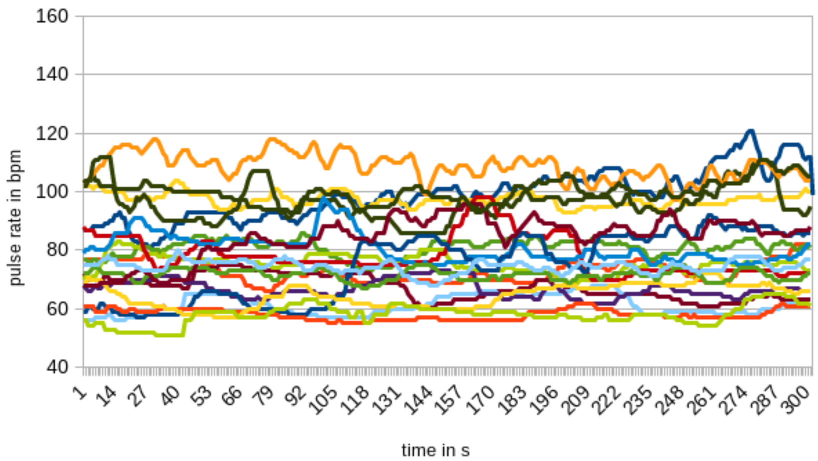


Fig. 5.13 *Mass distribution pulse rates*, Spatial Audio run test group II (n = 20)



Chapter 6

Evaluation

6.1 Metrics Assignment

What immediately catches the eye is the fact that both test groups show a slightly wider distribution range and higher pulse rates in the stereo version. In addition, the stereo graphs show very clearly that one outlier line with a significantly higher pulse rate was created in each group. So first, I would like to focus my attention on these two noticeable characteristics. Test subjects A9 of research series I and X3 of research series II show the following individual progression in their test runs:

A9: Female, 22

☑ VR experienced

MSSQ raw score	11
Pre Test SSQ score	0.438
Post Spatial SSQ score	0.813
Post Stereo SSQ score	0.563
1. run	Stereo
2. run	HRTF

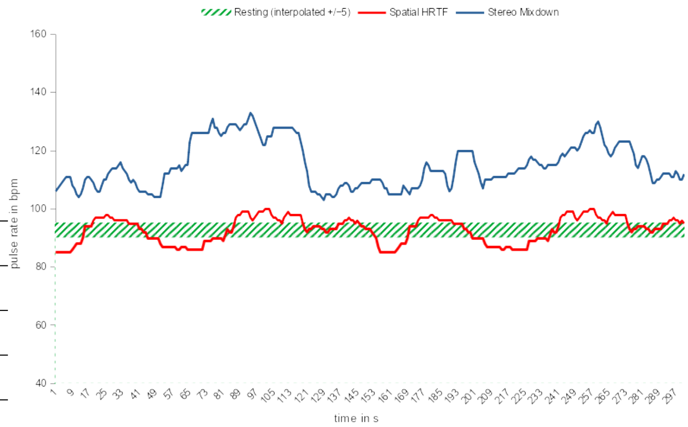


Fig. 6.1 *Pulse rate graph*, test subject A9 research series I

This test candidate is above average susceptibility of the test group with an MSSQ raw score of 11. This is also reflected in an increased resting pulse rate. The SSQ symptoms were already slightly present before the test (0.438). Although the SSQ stress symptoms are much higher in the spatial version with 0.813, it should be mentioned that this test subject had first gone through the stereo version and subjectively found the spatial test run more pleasant and immersive. I derive the significantly higher spatial SSQ value from the fact that she already indicated some SSQ symptoms before the test runs and, with an MSSQ value of 11, there was already a higher vulnerability to motion sickness existing. Due to repetitive habituation, the pulse rate of the spatial version runs with up to 30 bpm difference clearly below the first experienced stereo version and closer to the resting pulse rate. In addition, it is noticeable that the graph of the spatial version runs clearly smoother and more homogeneous, in contrast to the graph of the stereo version in which harsh jumps with suddenly occurring peaks are recorded. I attribute this to the auditory perception of the stereo version which is unnatural to humans.

X3: Female, 19

☐ no VR experience

MSSQ raw score	8
Pre Test SSQ score	0.188
Post Spatial SSQ score	0.313
Post Stereo SSQ score	0.625
1. run	Stereo
2. run	HRTF

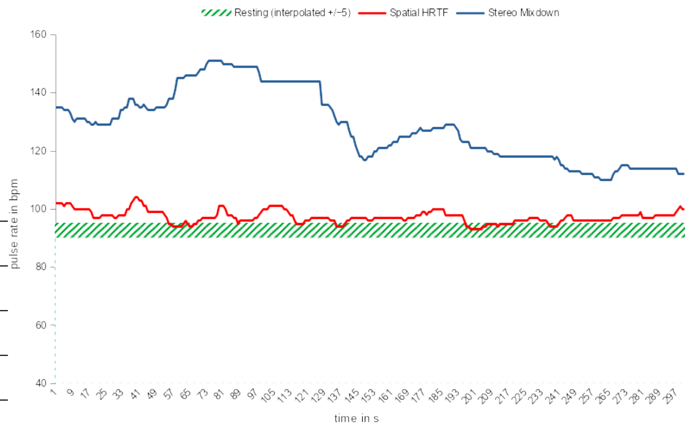


Fig. 6.2 *Pulse rate graph*, test subject X3 research series II

This test candidate has no VR experience and is in the normal MSSQ susceptibility midfield with an value of 8. Although she barely shows any SSQ pre-symptoms, she enters the stereo version with a pulse rate of almost 140 bpm and later even reaches the maximum value of 150 bpm, the highest value measured in both test groups so far. I attribute this to a lack of VR experience resulting in an increased excitement. In the further curve progression of the stereo variant, I notice a strongly unsteady curve similar to that of test subject A9 which slowly levels off towards the end around 120 bpm. Remarkable here is the gradient of the whole graph and the rather angular instead of pointed curve as with A9. The SSQ value for the stereo variant is also correspondingly high: 0.625, twice as high as for the spatial version. Due to the fact that the spatial variant is used to repetition, the curve is significantly lower than the stereo graph at up to 30 bpm difference. Again, there are ups and downs but these are not angular edges but rather gradual slopes and valleys, often within the resting pulse area indicating a more natural auditory stimulus perception.

In the following I will analyze further individual test runs which show significant characteristics and recurring patterns within both test series:

A5: Male, 27

☑ VR experienced

MSSQ raw score	7
Pre Test SSQ score	0.438
Post Spatial SSQ score	0.625
Post Stereo SSQ score	0.813
1. run	Stereo
2. run	HRTF

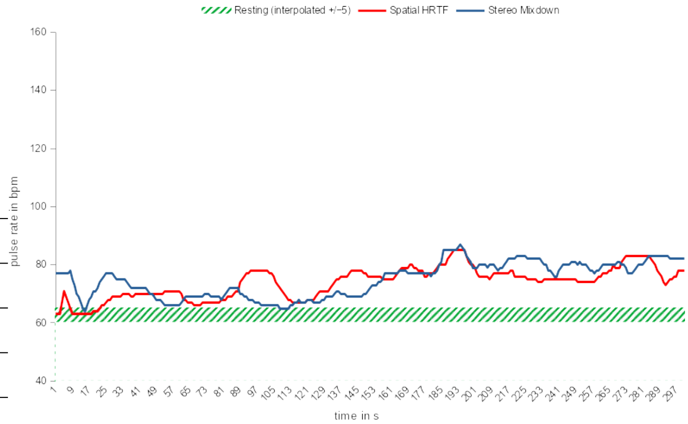


Fig. 6.3 *Pulse rate graph*, test subject A5 research series I

This test probant has VR experience and with a vulnerability value of 7 it is approximately in the MSSQ average of test group I. With his both pulse rate progressions slightly above the resting pulse he shows a recurring pattern of almost all participants in both research series. Also, like most of the participants, he indicates increased SSQ symptoms after the stereo run (0.813) but in his normal condition before the VR experience he also came to the test with slight SSQ signs (0.438). The lines of both variants are almost congruent in some areas of the graph but still show striking differences here and there. As with the previous test, the spatial version shows a more natural up and down curve corresponding nearly identical to the elevation of the test track.

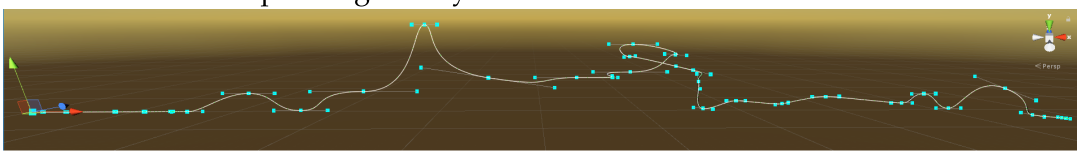


Fig. 6.4 *Side view with extracted elevation path*, test track

Surprisingly, despite his given VR experience, he subjectively indicated after the test the stereo version more comfortable and immersive although the graphs and symptom values tell different.

B7: Male, 64

☐ no VR experience

MSSQ raw score	0
Pre Test SSQ score	0
Post Spatial SSQ score	0.125
Post Stereo SSQ score	0.063
1. run	HRTF
2. run	Stereo

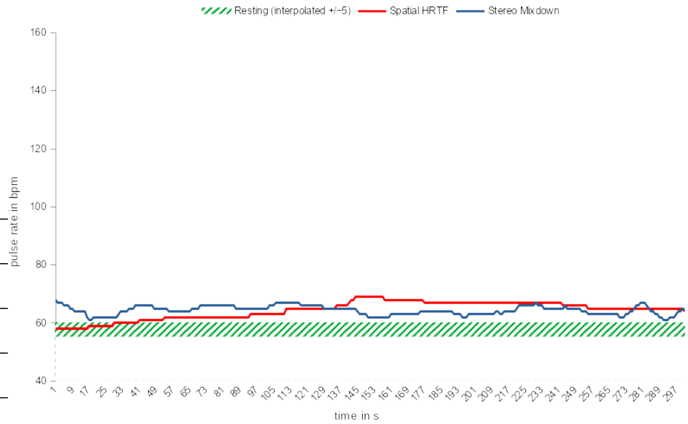


Fig. 6.5 *Pulse rate graph*, test subject B9 research series I

This tester is representative for the three other participants between 50 and 74 years, as the remaining participants are 30 years old or younger and show little in common with the younger ones. First of all, it should be mentioned that the pulse rates of the older participants are significantly lower and run in smaller amplitudes. A good example of this is the curve of participant B7 described above. In his first run with spatial audio setup I noticed a slow and steady ascent, almost linear until about the half of the application's time when the track reaches its steepest climb and largest altitude. Quite contrary to the course of the stereo pulse rate: Again, here I see the now known fluctuations but much less pronounced than it was partly the case with the testers before. Furthermore, there are no abrupt increases or decreases. Maybe this is also due to the fact that the older participants are not so much enthusiastic for new technologies than the younger ones.

A4: Male, 29

☐ no VR experience

MSSQ raw score	3
Pre Test SSQ score	1.375
Post Spatial SSQ score	1.438
Post Stereo SSQ score	1.813
1. run	Stereo
2. run	HRTF

Y1: Male, 24

☑ VR experienced

MSSQ raw score	3
Pre Test SSQ score	0.375
Post Spatial SSQ score	0.313
Post Stereo SSQ score	0.500
1. run	HRTF
2. run	Stereo

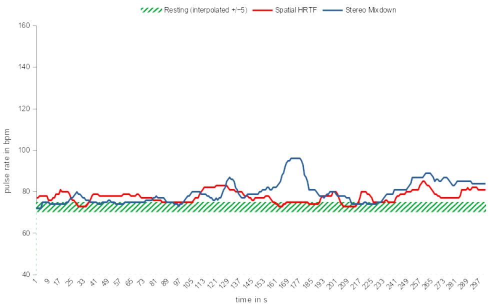


Fig. 6.6 *Pulse rate graph*, test subject A4

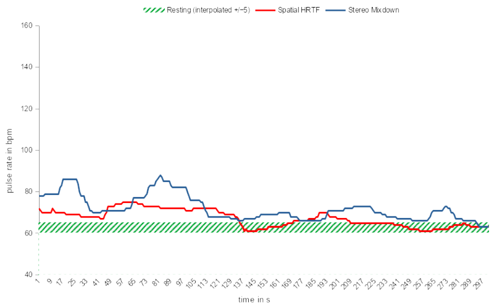


Fig. 6.7 *Pulse rate graph*, test subject Y1

Finally, I present two candidates who had the same MSSQ susceptibility values, a different audio setup test order but similarly striking features in their graphs and SSQ symptom values. Both testers did not only subjectively perceive the spatial version to be more pleasant and immersive but the symptom values of both participants were also more enhanced in the stereo version. It is also interesting to note that for test candidate Y1 there was even a minimal improvement in the symptom values by the spatial version compared to the initial situation. Although both had a different test order and also differ in the previous VR experience, the curves of the pulse rates show similarities. While the curves of the spatial variant are smoother for a longer period of time, the curves of the stereo variant often show unsteady changes of direction. In addition, sudden curve breakouts and value maxima can be seen at some points in both graphs whereas the curves of the spatial variant at the same places only illustrate a gradual and much more homogeneous in- or decrease.

Final analysis of both research series

When I summarize the data of both groups it is noticeable that the female participants illustrate my research results to a greater extent than the male testers do:

	Female	Male
Avg. post Spatial SSQ score	0.597	0.609
Avg. post Stereo SSQ score	0.733	0.725

Table 6.1 Avg. SSQ scores divided by gender

Another interesting observation I made during the data analysis is the fact that it is not important whether a person has previous VR experience or not. In both cases the stereo version is perceived as more disturbing and therefore a higher SSQ symptom value is assigned. However, the result is somewhat clearer for participants with VR experience:

	VR experienced	No VR experience
Avg. post Spatial SSQ score	0.634	0.566
Avg. post Stereo SSQ score	0.780	0.665

Table 6.2 Avg. SSQ scores divided by existing VR experience

I also think it's interesting to mention that, except for one person, none of the testers subjectively stated that the stereo version was both more pleasant and more immersive in the questionnaire. Furthermore, those who found the stereo version either more comfortable or immersive had no VR experience before and were distracted by the overstimulation to make an objective decision.

Another interesting observation I have made is the severity distribution of the individual SSQ symptoms:

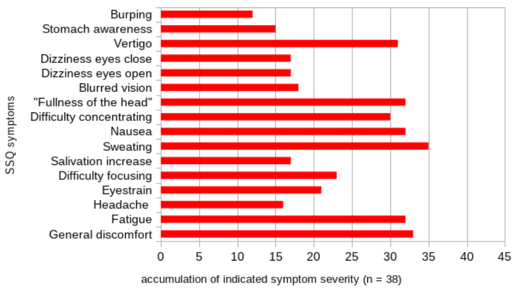


Fig. 6.8 *Symptom severity*, Spatial HRTF

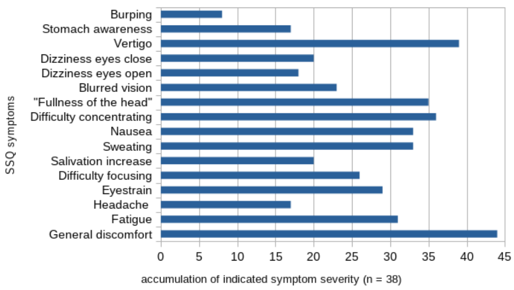


Fig. 6.9 *Symptom severity*, Stereo Mixdown

While testers in the stereo version indicated general discomfort and vertigo as the strongest symptoms, the spatial version shows a larger individual distribution with less severe symptoms. As many as seven symptoms have similar severity values and are therefore synonymous for the general consequences of simulator sickness.

6.2 Main Conclusion

With the help of two test groups with a total of 38 participants I have proven that spatial audio in VR ensures less motion sickness. This is reflected on the one hand in simulator sickness symptom indicators, on the other hand in pulse rate progression graphs but also in subjective statements on the experienced VR test tracks. I have shown that unnatural auditory sensory impressions, as it is the case with static stereo sound, enhance symptoms of simulator sickness. This supports the hypothesis that people with their fine sensory processing are able to compare auditory impressions with their already known auditory sensations, to evaluate them and even to connect them to other reference systems such as the organ of balance. I have shown that the sensitive stimulus processing system of the human body is disturbed in its effect and function by unnatural sensory impressions. I assume that the insufficient coordination between the sensory auditory impressions of the ears and the organ of balance gets even worse with stereo audio. I showed that this contradiction rises when the perceived soundscape is in a static stereo format: Not only subjectively users evaluated the corresponding application worse than the one with spatial sound but also physical reactions in form of simulator sickness symptoms and a changed, mostly irregular heartbeat progression expressed this and support the fact that the organs responsible for the sense of balance are sensitive and closely connected to each other and that this depending relation can be easily disturbed by unnatural influences.
These results should be considered for the further development of VR applications in terms of content and technology. A continuous confrontation with averse auditory sensory impressions could lead to users rejecting an application or getting ill, both physically and mentally, in case of long-term consumption. A little state control could also be applied here in terms of regulation and human protection of psyche and body. In my opinion, new technologies should not be used and distributed without restrictions just because it's technically possible and economically profitable.

I advocate at least a bit more government scrutiny into the ways VR and other reality technologies are being used. I hope that the Institute of Electrical and Electronics Engineers (IEEE), the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity, addresses the uses of VR specifically in its code of ethics and health.

6.3 Room for Interpretation

Due to the permanently installed *VR ready*™ desktop computers at the university, I was always bound to one room. From time to time the heating system in this room became very warm which could not be regulated locally due to the central climate control of the university. Opening the windows increased the heating capacity during the cold test months even more. Although, I always made sure that there was proper air circulation in the test rooms, it cannot be denied that the occasional increase in room temperature favored the simulator sickness symptom of sweating.

6.4 Further Research Potential

My research application showed the test participants the same visual stimulus twice in form of an accelerated track ride in VR. Since this circumstance greatly favors the risks of getting used to repetition, it would be interesting to find out whether visually different VR experiences with different audio setups in stereo and spatial have an even stronger impact and thus contain a higher result significance. Anyway, it would be interesting to investigate whether long-term exposure under wrong conditions causes an aggravation of well-being and simulator sickness symptoms or whether a gradual adjustment of habits takes place in human habits.
In addition, I think it would be desirable to investigate whether an increasing steepness or differences in height compared to the in-game normal zero is reflected in a correlation of the measured pulse rate, as already assumed in the present work.
Since my VR world contains a very dark basic audio atmosphere, I would also be interested to know if the pitch, or higher frequencies in general, have an influence on the VIMS.
Also, like my results consider, research in recent years shows that young people and women in particular are more susceptible to kinetosis, including seasickness [(Lawther and Griffin 1986), (Lawther and Griffin 1988)], traveling in buses (Turner and Griffin 1999) and in the context of linear open-loop oscillation (Koslucher et al. 2015). Thus, gender differences in motion sickness are also a current and prospective research area for VR and the content producing industry.

Bibliography

[1] Woodrow Barfield and Thomas A. Furness III, eds. *Virtual Environments and Advanced Interface Design*. New York, NY, USA: Oxford University Press, Inc., 1995. ISBN: 0-19-507555-2.

[2] Kwan Lee. "Presence, Explicated". In: *Communication Theory* 14 (Feb. 2004), pp. 27–50. DOI: 10.1111/j.1468-2885.2004.tb00302.x.

[3] International Society for Presence Research[online]. *The Concept of Presence: Explication Statement*. 2000. URL: https://ispr.info/about-presence-2/about-presence/ (visited on 09/03/2019).

[4] B. F. Skinner. "Two Types of Conditioned Reflex and a Pseudo Type". In: *The Journal of General Psychology* 12.1 (1935), pp. 66–77. URL: https://doi.org/10.1080/00221309.1935.9920088.

[5] Béatrice S. Hasler, Bernhard Spanlang, and Mel Slater. "Virtual race transformation reverses racial in-group bias". In: *PLOS ONE* 12.4 (Apr. 2017), pp. 1–20. URL: https://doi.org/10.1371/journal.pone.0174965.

[6] Behrang Keshavarz, Lawrence Hettinger, et al. "Demonstrating the Potential for Dynamic Auditory Stimulation to Contribute to Motion Sickness". In: *PLOS ONE* 9.7 (July 2014), pp. 1–9. URL: https://doi.org/10.1371/journal.pone.0101016.

[7] Behrang Keshavarz and Heiko Hecht. "Pleasant music as a countermeasure against visually induced motion sickness." In: *Applied ergonomics* 45 3 (2014), pp. 521–7.

[8] Hironori Akiduki et al. "Visual-vestibular conflict induced by virtual reality in humans". In: *Neuroscience Letters* 340.3 (2003), pp. 197–200. ISSN: 0304-3940. URL: http://www.sciencedirect.com/science/article/pii/S0304394003000983.

[9] J. F. Golding. "Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness". In: *Brain Research Bulletin* 47 (1998), pp. 507–516.

[10] J. F. Golding. "Predicting individual differences in motion sickness susceptibility by questionnaire". In: *Personality and Individual Differences* 41 (July 2006), pp. 237–248. DOI: 10.1016/j.paid.2006.01.012.

[11] Robert S. Kennedy et al. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness". In: *The International Journal of Aviation Psychology* 3.3 (1993), pp. 203–220. DOI: 10.1207/s15327108ijap0303_3.

[12] IrisVR[online]. *The Importance of Frame Rates*. 2009. URL: https://help.irisvr.com/hc/en-us/articles/215884547-The-Importance-of-Frame-Rates (visited on 12/13/2019).

[13] Kavitha Ratnam et al. "Relationship Between Foveal Cone Structure and Clinical Measures of Visual Function in Patients With Inherited Retinal Degenerations". In: *Investigative ophthalmology and visual science* 54 (Aug. 2013). DOI: 10.1167/iovs.13-12557.

[14] Hans-Dieter Lüke. "The origins of the sampling theorem". In: *IEEE Communications Magazine* 37 (1999), pp. 106–108.

[15] E. Brad Meyer and David R. Moran. "Audibility of a CD-Standard A/DA/A Loop Inserted into High-Resolution Audio Playback". In: *J. Audio Eng. Soc* 55.9 (2007), pp. 775–779. URL: http://www.aes.org/e-lib/browse.cfm?elib=14195.

[16] Sony/USPTO[online]. *United States Patent Application Publication No. US 2019/0304101 A1*. 2019. URL: https://nl.letsgodigital.org/uploads/2019/10/sony-ps5-vr-headset.pdf (visited on 12/16/2019).

[17] Ernst Heinrich Weber. *De pulsu, resorptione, auditu et tactu: annotationes anatomicae et physiologicae, auctore*. prostat apud CF Koehler, 1834.

[18] G.T. Fechner, ed. *Elemente der Psychophysik*. Breitkopf Leipzig, 1860.

[19] Harvey Fletcher and W. A. Munson. "Loudness, Its Definition, Measurement and Calculation". In: *Bell System Technical Journal* 12.4 (1933), pp. 377–430. DOI: 10.1002/j.1538-7305.1933.tb00403.x.

[20] International Organization for Standardization/TC 43 Acoustics[online]. *Acoustics — Normal equal-loudness-level contours (ISO 226:2003)*. 2003. URL: https://www.iso.org/standard/34222.html (visited on 12/19/2019).

[21] C.J. Plack. *The Sense of Hearing*. Lawrence Erlbaum Associates, 2005, p. 182. ISBN: 9781848729872.

[22] W. E. Feddersen et al. "Localization of High-Frequency Tones". In: *The Journal of the Acoustical Society of America* 29.9 (1957), pp. 988–991. URL: https://doi.org/10.1121/1.1909356.

[23] D. M. Green. *An introduction to hearing*. Halsted Press, 1976. ISBN: 978-0470151884.

[24] E. B. Goldstein. *Sensation and Perception*. Pacific Grove, CA: Brooks / Cole, 1996, p. 401. ISBN: 0534096727.

[25] J. Blauert, ed. *Räumliches Hören*. Stuttgart, BW, Germany: Hirzel, 1974. ISBN: 3-7776-0250-7.

[26] J. Makous and J. Middlebrooks. "Two-dimensional sound localization by human listeners". In: *The Journal of the Acoustical Society of America* 87 (June 1990), pp. 2188–2200. DOI: 10.1121/1.399186.

[27] A. W. Mills. "On the Minimum Audible Angle". In: *The Journal of the Acoustical Society of America* 30.4 (1958), pp. 237–246. URL: https://doi.org/10.1121/1.1909553.

[28] Joo Young Hong. "Spatial Audio for Soundscape Design: Recording and Reproduction". In: (2017), p. 10. DOI: https://doi.org/10.3390/app7060627.

[29] H. Haas. "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache". In: *Acustica* (1951), pp. 49–58.

[30] Christian Doppler, ed. *Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels*. Königliche Gesellschaft der Wissenschaften zu Prag, 1842. ISBN: 978-1334011122.

[31] R. Thom[online]. *RandyThomBlog - Thoughts on film sound design: Roma*. 2019. URL: https://randythomblog.wordpress.com/2019/03/12/roma/ (visited on 12/20/2019).

[32] r/GameAudio[online]. *Games With True 3D Spatial Audio (Ambisonics, Binaural, HRTF, Dolby Atmos for Headphones)*. 2019. URL: https://docs.google.com/spreadsheets/d/1OGi40yAonUDL6UcQaIIfGbhT1XY44JVs9lipkOx6QRg/edit#gid=634347005 (visited on 12/26/2019).

[33] J. F. Golding. "Motion sickness susceptibility". In: *Autonomic neuroscience : basic and clinical* 129 (Nov. 2006), pp. 67–76. DOI: 10.1016/j.autneu.2006.07.019.

[34] D. Grundy. "Nausea and vomiting - an interdisciplinary approach". In: *Auton Neurosci-Basic* 129 (1-2 Oct. 2006), pp. 1–2. DOI: 10.1016/j.autneu.2006.07.008.

[35] M. Lewis, D. F. A. Hitchcock, and M. W. Sullivan. "Physiological and Emotional Reactivity to Learning and Frustration". In: *Infancy* 6.1 (2004), pp. 121–143. URL: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15327078in0601_6.

[36] J. J. Reason J. T. Brand. *Motion sickness*. Academic Press London; New York, 1975, pp. 102, 135, 210, 241. ISBN: 0125840500. URL: https://nla.gov.au/nla.cat-vn247634.

[37] Gary E. Riccio and Thomas A. Stoffregen. "An Ecological Theory of Motion Sickness and Postural Instability". In: *Ecological Psychology* 3.3 (1991), pp. 195–240. ISSN: 1040-7413. DOI: 10.1207/s15326969eco0303_2.

[38] Robert Kennedy and Ashton Graybiel. "The Dial Test - A standardized procedure for the experimental production of canal sickness symptomatology in a rotating environment". In: (June 1965), p. 30. URL: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19660006694.pdf.

[39] E. M. Kolasinski. *Simulator Sickness in Virtual Environments*. Tech. rep. Army Research Inst for the Behavioral and Social Sciences Alexandria Va, 1995. DOI: https://journals.sagepub.com/doi/10.1177/154193129804202110.

[40] R. L. Hudson B. Mandelbrot, ed. *The Misbehavior of Markets: A Fractal View of Financial Turbulence: A Fractal View of Risk, Ruin, and Reward*. Basic Books, 2006. ISBN: 978-0465043576.

[41] Anthony Lawther and Michael Griffin. "The motion of a ship at sea and the consequent motion sickness amongst passengers". In: *Ergonomics* 29.4 (1986). PMID: 3709507, pp. 535–552. DOI: 10.1080/00140138608968289.

[42] Anthony Lawther and Michael Griffin. "Motion sickness and motion characteristics of vessels at sea". In: *Ergonomics* 31.10 (1988). PMID: 3208731, pp. 1373–1394. DOI: 10.1080/00140138808966783. URL: https://doi.org/10.1080/00140138808966783.

[43] Mark Turner and Michael Griffin. "Motion sickness in public road transport: Passenger behaviour and susceptibility". In: *Ergonomics* 42 (Apr. 1999), pp. 444–61. DOI: 10.1080/001401399185586.

[44] Frank Koslucher et al. "Sex Differences in the Incidence of Motion Sickness Induced by Linear Visual Oscillation". In: *Aerospace Medicine and Human Performance* 86 (Sept. 2015), pp. 787–793. DOI: 10.3357/AMHP.4243.2015.

List of Figures

1.1 Gartner (2016): *Hype Cycle For Emerging Technologies* 1
1.2 Stephanie Llamas, Nielsen/SuperData Research (2019): *Technology adoption rates, share of U.S. households, 1960 - 2020 [12pt]* 2

2.1 Sutherland, I.E. (1968): *A Head -Mounted Three-Dimensional Display* . . . 5
2.2 Krantz, J. H., Schwartz, B.L. (2015): *ISLE 7.5 (a1): Motion Parallax Illustrated* 6
2.3 Ponzo, M. (1913): *rail illusion* 6
2.4 Ebbinghaus, H. (1890): *Ebbinghaus illusion* (l)., Delboeuf, J. R. L. (1865): *Delboeuf illusion* (r.) 7
2.5 Sweet, B., Kaiser, M. (2011): *Depth Perception, Cueing, and Control* 7
2.6 Mauderer, M. , Conte, S. , Nacenta, M. A., Vishwanath, D. (2014): *Depth perception with gaze-contingent depth of field* 7
2.7 McIntire, J., Havig, P. R., McIntire L. (2009): *Ideas on authenticating humanness in collaborative systems using AI-hard problems in perception and cognition* 8
2.8 Adelson, E. H. (1995): *The checker shadow illusion* 8
2.9 Cooper, R. (1995): *How to See 3D* 8
2.10 Duane, A. (1912): *Normal values of the accommodation at all ages* Journal of the American Medical Association, pp. 1010-1013 9
2.11 Marten, C. (2011): *Stereoscopic Rendering* tinySG 9
2.12 Wuerger, S. M. (2011): *Visual gamma correction for LCD displays* 10
2.13 Ruhstaller, B. (2011): *Advanced Numerical Simulation of Organic Light-emitting Devices* 10
2.14 Muggeridge, E. J. (1878): *The Horse in Motion* 11
2.15 Tykkälä, T. (2013): *Real-time image-based RGB-D camera motion tracking and environment tracking* 11
2.16 Kilgard, M. (2016): *NVIDIA OpenGL in 2016* 12
2.17 Kilograph.com (2017): *Virtual Reality and 6 degrees of freedom* 12
2.18 Suovanen, J. (Creative Electron, 2018): *Radiography of the HTC Vive Pro headphone* 17

3.1 Scott de Martinville, É.-L. (1911): *The Century Dictionary, Vol. VI, Page 4450,* Pholadidae to Phonikon 19

3.2 Bizzle, J. (2011): *Fairlight CMI II,* National Association of Music Merchants (NAMM) in Anaheim, California (U.S.) 21
3.3 Nave, C.R. (2016): *HyperPhysics:* Traveling Wave Relationship 23
3.4 Osborne Head and Neck Institute (2016): *Ear, Nose and Throat Doctors* Los Angeles 24
3.5 Cohen, M. (2016): *Schematic Summary of Binaural Effects: ITD and IID,* Applications of Audio Augmented Reality: Wearware, Everyware, Anyware, and Awareware 26
3.6 Blauert, J. (1974): *Relative frequencies of observers who give one of the judgements "v", "o" or "h" more frequently than both of the others together (at a 95% level of significance).* Sound Localization in the Median Plane, pp. 209 27
3.7 Sørensen, M, Hammershøi, J., (1995): *Left ear HRTFs for open ear canal. White curves represent the evaluated mean values.* AES Vol.43/5, pp. 300-321 28
3.8 (Hong 2017): *Evolution reproduction systems,* Spatial Audio for Soundscape Design 29
3.9 Wilssens, S., Tuliper, A. (2016): *Spatial Audio in Windows 10 RS,* Windows Developer P527 #Build2016 32
3.10 *MS HRTF Spatializer plugin*, Unity Editor audio settings 33
3.11 *Audio source 3D sound settings*, Unity Editor GameObject inspector . . . 34

4.1 Virginia Merrill Bloedel Hearing Research Center (2008): *Diagram of the inner ear* 40

5.1 Schleise, V. (2019): *Impression during test cycle I* 45
5.2 *Subject data and MSSQ query*, research questionnaire page 1 46
5.3 *SSQ query*, research questionnaire page 2 47
5.4 *Closing questions and feedback*, research questionnaire page 3 48
5.5 *In-game POV sight*, test track 49
5.6 *Scene/game view I*, test track 49
5.7 *Scene/game view II*, test track 49
5.8 *Scene/game view III*, test track 49
5.9 *Scene/game view IV*, test track 49
5.10 *Mass distribution pulse rates*, Stereo Audio run test group I ($n = 18$) . . . 53
5.11 *Mass distribution pulse rates*, Spatial Audio run test group I ($n = 18$) . . . 53
5.12 *Mass distribution pulse rates*, Stereo Audio run test group II ($n = 20$) . . . 54
5.13 *Mass distribution pulse rates*, Spatial Audio run test group II ($n = 20$) . . 54

6.1 *Pulse rate graph*, test subject A9 research series I 55
6.2 *Pulse rate graph*, test subject X3 research series II 56
6.3 *Pulse rate graph*, test subject A5 research series I 57
6.4 *Side view with extracted elevation path*, test track 57

6.5 *Pulse rate graph*, test subject B9 research series I 58
6.6 *Pulse rate graph*, test subject A4 59
6.7 *Pulse rate graph*, test subject Y1 59
6.8 *Symptom severity*, Spatial HRTF 60
6.9 *Symptom severity*, Stereo Mixdown 60

List of Tables

2.1 Specification overview of current HMDs 15

3.1 Examples of common sample rates 20
3.2 Examples of common bit depths 21
3.3 Various properties of uncompressed PCM digital audio 22
3.4 Görne, T. (1994): *Mixed stereophony techniques* Mikrofone: Theorie und Praxis 31
3.5 Field selection of the `SpatialAudioModel Enum` class, Windows AudioGraph API 33
3.6 *Non-VR Games with 3D Spatial Audio* (r/GameAudio[online] 2019) . . . 36
3.7 *VR Games with 3D Spatial Audio* (r/GameAudio[online] 2019) 37

4.1 Vulnerability questionnaire of the MSSQ (Golding 1998). 41
4.2 Symptoms of the SSQ (R. S. Kennedy et al. 1993) 43

5.1 *Listing of audio sources used*, test track 50
5.2 Data record of both test series 52

6.1 Avg. SSQ scores divided by gender 60
6.2 Avg. SSQ scores divided by existing VR experience 60