

Department of Architecture and Media Technology

Medialogy, 10th Semester

Title: The Effect of Spatial Audio on	Abstract:
Immersion, Presence, and Physiological	With the increasing interest in virtual re-
Response in Games	ality games and the newly available op-
A Master's Thesis	tions of binaural simulated sound (3D
	audio), the effects of spatialized audio
Project period:	on players has become of interest. This
02.02.2015 - 27.05.2015	study focuses on the effect that 3D au-
	dio has on players' levels of immersion
Semester theme:	and presence through the use of ques-
Master's Thesis	tionnaires, phasic electrodermal activity
	(EDA) and heart rate variability (HRV).
Supervisor:	In this study participants played a game
Martin Kraus	exposing them to different horrific audi-
	tory stimuli either rendered with stereo or
	3D audio. Questionnaires showed no sig-
Projectgroup no.:	nificant differences in level of immersion
151031	or presence, but a significant difference
	was observed for EDA responses. This
Members:	could suggest that 3D audio induces a
	higher level of immersion or presence on
	a subconscious level compared to stereo,
	though a higher sound intensity was ob-
Jacob Junker Larsen	served for 3D audio which could also ex-
	plain the results.

Marc Pilgaard

Editions: 2 **Number of pages:** 85 **Number of appendices:** 2 CDs **Delivery date:** 27th May 2015

Copyright@2015. This report and/or app ended material may not be partly or completely published or copied without prior written approval from the authors. Neither may the contents be used for commercial purposes without this written approval.

The Effect of Spatial Audio on Immersion, Presence, and Physiological Response in Games A Master's Thesis

Jacob J. Larsen and Marc Pilgaard Group 151031

27. Maj 2015

Preface

This report uses a few terms which we have defined ourselves. *3D Audio* refers to an audio system that utilizes head-related transfer functions (HRTFs) when rendering audio, while *panning audio* or *stereo* refers to an audio system that uses stereo panning and *monaural audio* or *mono* refers to an audio system using mono sound. When referring to any of the terms stated above, we simple refer to them as *audio*, though they are audio rendering methods. In this report we generally refer to audio rendering as if it is done over headphones, since this is the only playback method we apply in our study. We differ between audio and sound by, audio is referring to a rendering method, while sound is the audible product of such rendering.

With this master's thesis, an appendix CD is available. The CD contains the following:

- 'AV Production' contains a video presentation of this study.
- 'Participant Data' contains plots of EDA recorded from participants.
- 'Questionnaires' contains the questionnaires used in this study.
- 'Thesis' contains this thesis.

Contents

1	Intr	Introduction			
2	Pre	Previous Work			
	2.1	Binau	ral Hearing		
		2.1.1	Cocktail-party Effect		
		2.1.2	Localization Errors		
	2.2	2.2 3D Audio			
		2.2.1	Spatial Audio Rendering in Virtual Environments		
		2.2.2	Alternative Methods for Spatial Audio		
		2.2.3	Individualized and Non-individualized HRTFs		
		2.2.4	Our Previous Work with 3D Audio		
	2.3	Psycho	$physiology \dots \dots$		
		2.3.1	Emotional Reactions in Games		
		2.3.2	Physiological Measurements		
		2.3.3	Games and Psychophysiology		
2.4 Immersion and Presence		sion and Presence $\ldots \ldots 16$			
		2.4.1	Flow		
		2.4.2	Cognitive Absorption		
		2.4.3	Immersion		
		2.4.4	Presence		
3	Goa	oal 23			
4	Exp	erime	nt Design 25		
	4.1	Mater	ials		
		4.1.1	Virtual Environment		
		4.1.2	Sound Rendering		
		4.1.3	Data Collection		
		4.1.4	Questionnaires		
5	Data Extraction 3				
		5.0.5	Deriving Sound-Related Events		
		5.0.6	Deriving Heart Rate Variability		
	5.1	Observ	vations and Participant Discussions		

6	Res	ults	40			
7	Discussion					
8	Conclusion					
9	Future Work					
Aj	ppen	dices	59			
\mathbf{A}	Pilo	t Testing	60			
	A.1	First Pilot Test	60			
		A.1.1 Observations	62			
	A.2	Second Pilot Test	63			
		A.2.1 Observations	63			
	A.3	Third Pilot Test	64			
		A.3.1 Observations	64			
в	Pro	blems Encountered	67			
	B.1	Shimmer3 and Unity	67			
	B.2	Black Smearing	68			
\mathbf{C}	Additional Information 6					
	C.1	Tools for Developers	69			
	C.2	Rendering of Mono Audio	70			
	C.3	Consent Form Description	70			
	C.4	Sounds	73			
	C.5	Participant Data	77			
	C.6	Participants' Immersion Questionnaire Scores	79			
	C.7	Participants' Presence Questionnaire Scores	82			



1. Introduction

For a number of years the field of game studies has gained an increase interest in spatial audio. Previously one of the most researched area was graphics. This was a reflection from the game industry which for a long time has seen progress in graphical fidelity as the most important contributor to technological advancements within games. Though there have been improvements into the fidelity of sound, moving from 8-bit to 16-bit and later to 32-bit as well as the amount of simultaneous sources that can be present at one time, little attention has been paid to spatialization of sound. For the past two decades panning audio has been the standard in almost every game released. Though, in the late 80's external sound cards with the ability to render 3D audio (not to be confused with surround sound) was available on the consumer market, which provided 3D audio in a number of games. But during the 90's the sales of these sound cards slowly began to fall as manufactures started to incorporate sound cards onto their motherboards, but these on-board sound cards did not support 3D audio.

Because computational resources was sparse back then, 3D audio was rendered on a sound card due to its a relatively resource-heavy computational process. But since the power of computers is increasing every year, at an exponential rate, the possibility of rendering 3D audio using the central processing unit (CPU) has been a possibility while still running a game. This possibility has existed for a couple of years, though it is only recently that 3D audio has come into the attention of the video games industry again. See Appendix C.1 for further information on solutions utilizing 3D audio. The attention towards virtual reality displays or head mounted displays (HMDs) might be an explanation for this interest [1].

HMDs allow for a intuitive control of camera movement using only ones own head movement, which traditionally has been done using a mouse or keyboard. With HMDs in virtual environments (VEs), games now appear more realistic and induce more immersion than ever. The industry has begun to see 3D audio as an essential factor for virtual reality (VR) technology due to its more realistic features compared to stereo. It is to our knowledge hard to find any research done on the effect of better spatialization of audio that revolves around comparing the traditional stereo panning with 3D audio.

In this study the effect of stereo and 3D audio on immersion and presence levels, as well as physiological responses for players of a VR game, was investigated. This study includes an experiment where participants played a horror game where sound were either rendered



with stereo or 3D audio. Participants' self-evaluated immersion and presence was measured, together with their electrodermal activity (EDA) and heart rate variability (HRV) during the experiment. The results showed no significant difference in the questionnaire responses, but a significant difference were found in the EDA events, suggesting that 3D audio have had a larger effect than stereo audio, and we believe this is a product of a subconscious difference for induced immersion or presence. Though, we also found an intensity difference between the audio conditions which might explain this effect.

Applications for this study are games or VR. Knowing the effect of spatial audio can both help developers and designers to identify arguments when deciding which spatial auditory rendering to use for their systems and the consequences of their selected choice. Additionally, this study also contributes to the field of spatial audio, were often researches simply assume that 3D audio will induce more immersion or presence without any studies to support this claim [2]–[5].



2. Previous Work

This chapter is a presentation of previous work that is related to 3D audio, immersion, presence, and physiological measurements in games and VR.

2.1 Binaural Hearing

When we as humans perceive sound in the real world we are able to determine the position of the sound in terms of direction and distance. This ability is called binaural hearing and can only occur because we have two ears. Binaural hearing exists due to the physical properties of sound and the properties of the human body such as pinna, head and shoulders. Binaural hearing functions due to three auditory components: interaural time difference (ITD), interaural intensity difference (IID), and spectral differences (SD). [6]–[11]

ITD is caused by the distance between the two ears and the fact that sound waves move through space over time. This results in a sound wave reaching one ear before the other, if the sound source is not placed in the median plane. This results in a difference in arrival time between the two ears, see Figure 2.1 for an illustration of ITD. Humans can perceive ITDs of 10 microseconds which is a difference of approximately 1°[12] and the maximum ITD, based on an average head size, is around 650 microseconds [7].

IID is the difference in intensity from one ear to the other which is primarily caused by head shadow, a phenomenon caused by the head absorbing energy of the sound signal, see Figure 2.2. IID is used to determine the location of an audio source based on the difference in intensity between the two ears. As an example, if an audio source is located next to the right ear, the intensity of the sound in the right ear would be higher than the intensity in the left ear. It is especially frequencies above 1.5 kHz that are attenuated mostly due to their physical features being more prone to absorption because of head size [7].

SD is based on reflections and absorptions that occur on the body of the listener more specifically the shoulder head and pinna [7], [9]. When a sound is played at a point in space the body reflects and absorbs the signal differently for each location based on horizontal position (azimuth) and vertical position (elevation). This alters the sound's intensity at different frequencies which results in a cue that can help determine the





Figure 2.1: Illustration of ITD formed by a sound signal arriving at each ear at different times as the sound move through space.



Listener

Figure 2.2: Illustration of the IID caused by head shadow.



position of the sound relative to the listener. SD is the only cue that contains information about the sound's elevation [13], [14] where ITD and IID only contains information about azimuth. It should be noted that we distinguish between IID and SD in that IID is mostly concerned with the overall amplitude of sound, while SD is concerned with amplitude of each frequency.

2.1.1 Cocktail-party Effect

Besides the localization of a single sound source, binaural hearing creates an effect called the cocktail-party effect. The name of this effect is based on the ability that when one attend a cocktail party and speak with a person, one are able to focus on that persons voice even though there are a larger number of people around speaking. The cocktailparty effect helps us to filter individual sound sources in a multi-source environment, thereby increasing spatial awareness of sources. [11], [15]

2.1.2 Localization Errors

People can make errors when localizing sound both in terms of direction and distance. These errors are often caused by a common phenomenon known as cone of confusion, see Figure 2.3. This is a term describing when sounds are equal in ITD and IID and therefore only SD can be used to distinguish between what is up and down or front and back [8]. One of the most common errors that is seen is front-back errors. This is an error that occurs when the listener determines that a sound originates from the front is perceived as originating from behind or vice versa [14]. This phenomenon can however, be almost eliminated by introducing head movement [16], [17].

2.2 3D Audio

This section provides an explanation of 3D audio and how it distinguishes itself from other methods of audio rendering.

2.2.1 Spatial Audio Rendering in Virtual Environments

There exists a number of ways to render audio in real-time and some introduce more complexity than others. The most simple method of rendering audio is mono. Using headphones this method is applied by simply playing the same signal from both speakers, which is perceived as originating from within the head, hence it gives no spatial cues to the listener as it uses neither IID, ITD or SD.

A higher level of spatial fidelity for sound is stereo audio. Stereo audio allows for two different audio signals to be perceived at each ear, as an example if listening to music a guitar could be heard at the left ear while the right ear perceives drums. However, often





Figure 2.3: The cone of confusion is a range of positions where ITD and IID will be equal. On this figure the sound sources A and B and sources C and D will have an equal IID and ITD to the listener. Image retrieved from [18].

the signal is mixed and there is only a small diversity between the two signals. Stereo audio can be used to convey positional information about an audio source. This is often achieved through stereo panning, a method where the signal in the two audio channels are amplified or attenuated based on the position relative to the listener. For example if a sound source is present to the listener's right, the sound for the right ear is amplified while the left ear is attenuated, this effect is based on the use of IID. This relatively simple method of rendering audio is sufficient to give spatial cues about azimuth though fails to convey information about elevation. Depending on the implementation of stereo audio, ITD can also be used, though SD is not used. [11], [13], [19]

To reach an even higher level of spatial fidelity than stereo, one has to reproduce the effect of binaural hearing. This is what we refer to as 3D audio, and is a product of real-life ear recordings and a simulation based on these.

Using either a human listener or a dummy head one can obtain recordings of binaural audio. In both cases a pair of small microphones is placed inside the listeners ears [20], [21]. If not intended for simulation, one can use this setup to capture any type of audio and use it for binaural playback purposes. If intended for simulation, one can perform a systematically recording at specific elevations and azimuths capable of capturing a SD model. This is achieved by aiming a speaker at the listener and then play a sound impulse, whose response will be recorded by the microphones, hence the SD will be captured. Often a vertical array of speakers are arranged, and then rotated around the listener. The difference between the original impulse and the recorded impulse is called a head-related impulse response (HRIR). These recordings are often made in an anechoic chamber where reflections of walls are eliminated to get a simple and clear set of HRIRs with no room reflections. [6]



When these HRIRs have been recorded they are transformed with a Fourier transform to get a set of head-related transfer functions (HRTF). Applying these HRTFs to a sound signal, the signal can be perceived as originating from a point in space [6], [9], [22]. Often the number of HRTFs is up-sampled to get more HRTFs. A commonly used method for this up-sampling is bilinear interpolation where an extra HRTF is generated from an average of the four closest HRTF points but other methods also exist [19], [23].

Alternative methods have been applied in order to obtain HRTFs faster and easier. The original method of measuring HRIRs requires some very specialized tools and environments, and have therefore primarily been used for research purposes. For those who do not have access to such resources, a number of the HRTF-databases has been made publicly available [24]. One alternative method is a reciprocal configuration for HRTF measurements. This means that by applying the acoustic principle of reciprocity one can swap around the positions of microphones and speakers when acquiring the measurements. In practice this means that a small loudspeaker is placed inside the listener's ears and an array of microphones is placed around him. This allow of simultaneous collecting information from different angles based on the amount of microphones. [25]

Enzner et al. [24] have proposed a method which is based on continuously turning the subject while at the same time continuously playing the stimulus. This method has a very short duration of approximately one minute per person.

Another method to get a set of HRTFs is from 3D models of the shoulder, head and pinna of a person [21], [24]. By simulating the physical behavior of sound around the model the HRIRs can be calculated. Meshram et al. [9] proposed a method for obtaining HRTFs using a standard digital camera. They presented a technique where taking an amount of pictures around the head, shoulders and pinna can be used to reproduce the 3D model which can then be used for calculating a set of HRTFs. This has the potential of creating personal HRTFs for consumers, without the need of expensive and specialized equipment, see Section 2.2.3 on personal HRTFs.

After one has obtained a set of HRTFs, applying them is the next step for achieving 3D sound for real-time applications. To this end, the HRTF that corresponds to the azimuthal and elevated angles from the sound source's position relative to the listener is applied to the sound signal. Since there is only finite number of HRTFs, there will often not be a HRTF that exactly matches the azimuth or elevation. One could simply apply the HRTF which lies closest, but an undesired audible change will be apparent when azimuthal and elevated angles are recalculated and a new HRTF is the best fit. Different methods have been applied for interpolating between HRTFs. [23], [26]

2.2.2 Alternative Methods for Spatial Audio

The most common utilization of 3D audio is through the use of headphones, though this is not the only method. One other method that also utilize HRTFs is crosstalk cancellation, which is a technique that allows for individual left and right channel listening through a pair of normal speakers. The technique creates an auditory wall between signals from



the two loudspeakers, rendering the right speaker's sound to the right ear and the left speaker's sound to the left ear. This is achieved by a filter which sends counter waves to cancel out the sound before reaching the opposite ear. However, this technique only works in a *sweet spot*, a confined space, which limits the listener to be positioned within the sweet spot to perceive the effect. [27]

Another technique for obtaining highly spatialized sound reproduction is Ambisonics. This is a setup of loudspeakers placed all around the listener in a systematic arrangement, both vertically and horizontally, not to be confused with surround sound. Ambisonics is capable of reproducing a full spherical sound field compared to surround which is limited to a horizontal sound field. Ambisonics needs a special audio format to reproduce sound. This format is called B-format and consists in its first order of a W component which holds the amplitude of the sound (mono sound signal) and XYZ components which is pressure gradients of the spherical harmonics (directional information). A higher order of Ambisonics, more channels representing more spherical harmonics leading to better spatialization of the sound, though it also requires a higher amount of speakers to reproduce. [28], [29]

2.2.3 Individualized and Non-individualized HRTFs

Shoulders, head shape and pinna are different between individuals which means that HRTFs also differs. It is not always practical to obtain a new set of HRTFs and therefore often one can be required to make use of another individual's set of HRTFs. This introduces the problem of using non-individualized HRTFs which can be interpreted as listening with another individual's ears. This introduces more localization errors compared to using individualized HRTFs [8], [9], [21] though non-individualized HRTFs have been found to be sufficient for audio localization [8], [10], [13]. In particular localization errors are introduced in the median plane where the ITD is zero and only SD shapes the sound. [22]

Mendonça et al. [30]–[32] did a series of studies on how localization using non-individualized HRTFs was affected by training. The authors found that localization performance was increased by training. They also found that the effect was persistent after a month which suggests that our brain learn and stores information (plasticity) regarding auditory localization using another persons ears.

The Practical Problem of Individualized HRTFs

The practical problem when obtaining HRTFs the need of specialized materials [9], [20]. Besides, if one measures by the traditional method, it requires the subject of the measurement to sit still between 30 minutes to an hour. This problem has been subject to discussions regarding the potential of 3D audio in a commercial context, as this is a limitation to the consumer. One can think of a scenario where you in order to use an 3D audio solution, would have to go to a laboratory to obtain your own set of HRTFs in



order to get the best experience. This is not a viable solution, and before individualized HRTFs can be introduced to the mass market this problem has to be solved [2].

Romigh and Simpson [14] did an investigation into what components of a HRTF are the most critical to localization using individualized or non-individualized HRTFs. They found that the only significant component is the vertical SD because all other components such as ITD and IID are the ones that are most alike across individuals. Also, our hearing seems to compensate for some components like equalization of headphones, which makes it insignificant to equalize them for each individual. This suggests that if researchers are able to find an easy and minimal time consuming approach to obtaining only the vertical SD this could be sufficient for creating individualized HRTFs.

2.2.4 Our Previous Work with 3D Audio

Prior to this study, both authors have been involved in other projects regarding the comparison of 3D audio and stereo. The interest into the subject started as a 4th semester project, where we were involved in creating an audio game that investigated the effect of different controllers on player experience. This project was mainly aimed at visually impaired people, and during this project we investigated how to incorporate vertical game elements only utilizing sound. During this investigation we came across HRTFs and spatialized sound. Without fully understanding the concept of 3D audio, we implemented a simple low-pass filtering based on angle to the sound source attempting to imitate HRTFs.

On our 6th semester we decided to take on the subject again, and here we were also introduced to a local company, AM3D, which made the technology that we needed to render 3D audio. AM3D had a 3D audio engine, called Diesel Power Mobile at the time now known as Zirene 3D [33]. We integrated the engine into Unity [34] for our study. The study investigated the performance in navigational and localization tasks using only audio. Here we found that both performance in navigation and localization was better when using 3D audio comparing it to stereo panning. The results were published at Audio Mostly 2013 [10].

We continued our investigation into localization performance on the 7th semester where we built upon the previous study introducing visuals. In this study the task was to localize a visual target amongst an amount of visual distractors where the target was aided by a sound rendered either with 3D audio or stereo. This experiment did also show that 3D sound improved performance over stereo sound, and had an even large effect in environments with numerous visual distractors. The results were published at the 137th Audio Engineering Society Convention [13].

On our 8th semester we investigated the effect of spatialized audio in voice communication, where the task was for a pair of participants to navigate a virtual building while collaborating in both finding each other and locating a target within a building. All sound was either rendered using 3D, stereo or mono audio. At times they could use voice communication and at other times they were restricted to pinging. We found that the use



of communication was significant when the task was to find each other, though spatialization did not have any effect here. We also found that 3D and stereo was significantly better when the task was to find the target, but no difference was found between 3D and stereo audio, though a tendency was seen.

2.3 Psychophysiology

Due to the correlation between physiological reactions and psychology (psychophysiology), this section will cover some of the work done within this field. We are particularly interested in how it is possible to invoke physiological responses by inducing emotions in subjects. We also believe that immersion and presence has an interaction between how convincing an emotion is and how immersed or present one is [35], [36].

2.3.1 Emotional Reactions in Games

It is well known amongst gamers that games can give you an emotional response to the content presented by the game [37]. This emotional response can manifest itself in a wide range of emotions from fear and anger to happiness and calmness and everything in between.

Perron [38] outlines three types of emotional categories that describes the origin of an emotion. First is *fiction emotions* which are caused by the story of a game. Secondly there is *artifact emotions* which are elicited by the environment similar as art can, where the visuals or sounds are contributing factors. Thirdly he introduced the term *gameplay emotions*, which are created by gameplay and interactions with a game.

Emotions caused by gameplay might manifest themselves as any other emotions caused in real life, though the distinction can be seen in behavior. One example of this is when the game is causing fear in a player, he does not suddenly jump and run away from the computer as if he would in real life if approached by a horrifying monster or dangerous animal. Even though the player might be so engaged with the game that he portrays himself as the character of the game (immersion and presence) there is still some cognitive awareness of the situation that the player is actually sitting in front of a computer or TV. Instead the emotional response is often transfered into an action within the game as a *fight-or-flight* response. In an encounter with a monster the player also appraises the emotion based on the possibilities of gameplay, so if the player is able to beat the monster in any way the player might feel anger rather than fear, as the appraisal is based on how he can cope with the situation.

Ekman notes that:

"In games with a protagonist, gameplay emotions may equal care for the protagonist, but this care is essentially different from empathetic emotion: From the perspective of gameplay, the protagonist is a means, a tool, for playing the game." - Ekman [39, p. 22]



This suggests that emotions towards the protagonist is not only caused by empathy, but also looking at the protagonist as a tool for achieving personal goals.

Players also seek to play games which gives the player negative emotions, such as fear and sadness, because even negative emotions can to some extent give the player a positive experience. Another example of this is that sometimes a puzzle game can feel frustrating, but this might also contribute to a feeling of relief and joy when the player finally solves it. This is what Perron describes as Motive-Inconsistent emotions. [38]

It has also been argued whether these emotions felt during gameplay or watching movies are just fictional emotions or if they are real emotions at all [39]. This is also known as the *paradox of fiction*. The paradox lies in that our limbic system can have emotions that our higher brain functions do not share; thus we can and cannot have an emotion at the same time [40], [41]. In other words we can have a feeling of sadness because we emphasize with a piece of fiction, though we know that it is caused by fictional events. Therefore, we argue that since the emotions feels real they must be real, but since they are caused by fictional events we find that these emotions in most cases are more vague, both in terms of intensity and duration, than if they were caused by real life events.

Valence and Arousal

When speaking of emotions one of the common mappings is valence and arousal. Valence referrers to a positive or negative feeling in regard to pleasantness, so the feeling of being happy, calm and enthusiastic has a positive valence (pleasant), while feelings like sadness, boredom and fear are of negative valence (unpleasant). Arousal refer to a level of cognitive and bodily activation so feelings like stress and excitement have high arousal while sleepy and fatigued are of low arousal [42]–[45]. Russell [43] proposed the circumplex model of emotions as a tool for mapping emotions in relation to valence and arousal see Figure 2.4.

Emotions and Sound

The connection between emotions and sound has been made in both games and movies [38], [47]. Toprac and Meguid [36] argue that sound is one of the most important factors for evoking emotions within games and they also argue that immersion is created by sound, and that this to some extent is a precursor for emotions. See Section 2.4 for more on immersion. In the study of Toprac and Meguid they used sound to elicit fear in participants by playing sounds of different attributes. They found that loud sounds elicited the highest fear responses.

Garner and Grimshaw [48] presents a framework for eliciting fear in games using sound. They suggest that the level of fear based on sound is affected by the fear or anxiety already manifested in the player. Their model suggests that sounds of immediate relevance (close to the player) induces higher levels of fear than sounds of non-immediate relevance (far away from the player).



Figure 2.4: The circumplex model of emotions which presents a range of emotions and their relation to valence and arousal. This is a modified version retrieved from Ellerm [46] originally proposed by Russell [43].

Ekman and Kajastila [47] did a study on how location and/or spread of sound affected the scariness of sound. They found that point sounds from behind were judged to be scarier than if played from the front, but spread sounds increased the scariness. They suggest that this is caused due to the ambiguous location of the sound source.

Games without sound can be a good experience, though adding sound can create a stronger connection to the action or the characters. Sound can be in terms of music and/or sound effects (SFX), and both add to the experience of the game. Speaking of gameplay emotions especially SFX are important to convey feedback of actions to the player [39].

2.3.2 Physiological Measurements

The use of physiological measurements has been used to evaluate emotional states and game experiences in a number of previous works. [44], [49]–[55] Physiological responses occur autonomously from the participants, and therefore they create an insight into re-



actions that the participants might not be consciously aware of. Therefore, physiological measurements can become a tool for objective indication of emotional states. The physiological changes is caused by the autonomous nervous system (ANS) which is responsible for bodily reactions to situations based on their relevance. In a dangerous or stressful situation the sympathetic *fight-or-flight* system is activated, while the parasympathetic *rest-and-digest* is active during relaxed situations. [50], [51], [56]

Tonic and Phasic Response

When working with physiological measurements there are to two types of responses. First is the tonic response, which is looking at the data over a long period of time, typically the entire play session of a game. Tonic can however often be hard to measure because physiological data introduces a lot of noise over time, which can have a large effect especially for short lasting experiences. Tonic can often be used to determine the overall feeling that a player has experienced during a play session and can also help to identify a continuous increase or decrease in arousal. [50]

Secondly one can look at phasic responses, which are related to a specific event of short duration. These responses are often very immediate after an event, and provides an indication of whether a subject has reacted to an event or not, and can also help to determine how strong the reaction has been. Phasic responses do not, however, show the overall effect of a session of gameplay due to the short nature of arousal. [50]

Skin Conductance

Electrodermal activity (EDA), sometimes referred to as galvanic skin response, electrodermal response or skin conductance level, is a way of measuring activity of the sweat glands in the skin. The sweat gland activity changes if a person's level of arousal changes by physiological or psychological events. The level of sweat gland activity activates or deactivates the production of sweat and therefore changes in sweat gland activity results into changes of conductivity in the skin. EDA is often measured with the use of two electrodes placed on two different fingers, where one electrode sends a small current through the body, and the resistance is measured over the two electrodes where the body functions as a variable resistor. EDA has a rather slow response to an event and can be measured one to four seconds after the event has occurred. However, it is still a good indicator of phasic response of arousal, due to its sensitive nature to arousal. It can also indicate tonic levels of arousal. [42], [50], [56]

Cardiovascular Activity

Cardiovascular activity is caused by physical og psychological arousal which results in the heart beating faster to pump more blood. This results in more oxygen being transported around the body to feed cells. In the case of physical activity the muscle cells need



oxygen to continue the activity. In the case of psychological arousal it affects the body's preparation for physical activity, such as when a person gets scared the body goes into fight-or-flight mode. The cardiovascular activity can be measured in a couple of ways, one being electrocardiography (ECG) which involves a number of electrodes (minimum 3) being placed at certain parts of the body where the resistance between the electrodes manifests itself as the heart's activity. Another method is photoplethysmography (PPG) which is the measurement of blood flow through the veins. This is often measured at the fingers by placing a pulse oximeter that illuminates the skin with a diode, and a light sensor which detects changes in illumination. When the blood flows through the vein this absorbs the light and the change in reflected light can then be measured. [50]

From the ECG or the PPG one can extract various features such as heart rate (HR), interbeat (R-R) interval, heart rate variability (HRV), and blood flow all providing information about the activity of the cardiovascular system. Regarding HRV this refers to the variability in R-R intervals and is often associated with relaxation or stress. When one is relaxed the high frequencies for HRV (above 0.15 Hz) is more frequent due to activity in the parasympathetic system of the ANS. For example, the heart rate is influenced by respiration (Respiratory Sinus Arrhythmia) due to the exchange of gas between the lungs and the blood. This variability is decreased when the body is in a stressed state [51], [57]. To measure HRV one can obtain the power spectrum of an R-R interval signal, and compare the power density between the lower and higher frequencies [58]. One can also look at phasic HR events, but the change is often very small, the HR can also be measured over longer periods giving differences in tonic levels [50].

Facial Muscle Activity

When we get an emotional response this often reflects in activity in the facial muscles. It can be a smile when we experience something positive, or a contraction of the eyes when we experience the opposite. The activity of the facial muscles can sometimes be so small that the naked eye cannot detect it. To obtain this activity one can use electromyography (EMG), which is obtained by placing electrodes on top of the different facial muscles that are of interest. These electrodes can detect the small electrical impulses that are generated when muscle fibers contract. The EMG signal can be used to identify different emotions as these are associated with certain facial muscle activities [44]. The response of EMG can be measured almost instantly after a stimulus and therefore EMG is good for phasic responses, especially regarding valence [50]. Certain facial muscles have often been associated with mental effort [59].

Brain Activity

The most important organ in the body when speaking of emotional response is the brain, because it is from here that the rest of the organs and activities are controlled. It is the brain that interprets the response to the stimuli of e.g. the eyes and the ears and converts those into a bodily reaction. It is also possible to measure the activity of the



brain in several ways. One of the common methods within game research of emotions is electroencephalography (EEG). EEG is the measurement of electrical activity along the scalp that is a result of neurons firing within the brain. These responses can be measured almost instantaneously from an event to response and is therefore a good indication of phasic response, but it can also be measured for tonic levels of brain activity. Activity in certain areas of the brain corresponds to specific emotional states. [50]

2.3.3 Games and Psychophysiology

Raaijmakers et al. [51] used HRV and EDA in a biofeedback context where they used it for participants to control events in a series of games. As an example the participants had to control their HRV and EDA by following a breathing exercise. The games was used as a therapeutic exercise, where they found no evidence to support that treatment was affected by biofeedback.

Garner and Grimshaw [54] conducted a study where they used EDA and EMG as input for a game, where they changed the various properties of sound according to the signals received by the physiological measurements. The game was designed to alter the level of fear using different sound effects. They found that EDA provided as an reliable indicator of fear.

Mandryk and Atkins [44] tried to use EDA, EMG and HR in a fuzzy model to detect levels of arousal and valence within a game context. This was successfully done and they furthermore made another model to convert arousal and valence into five different emotions as a tool for measuring user experience.

Nacke et al. [53] did an investigation into the effect of game design on EEG. They designed three levels that should endorse boredom, immersion and flow as gameplay experience which they evaluated with the Game Experience Questionnaire [60]. They found that EEG was indeed affected by game design.

Salminen and Ravaja [61] investigated the effect of play on EEG when playing the game Monkey Ball 2. The game involved rolling on a ball picking up bananas, and one could fall off the edge of the map. When players picked up bananas they saw that the players EEG suggested arousal. When the players fell off the edge the areas of the brain connected to motor actions was activated. When the players completed a level the EEG suggested a relaxed state of the player.

Garner [52] tried to use EEG as a biofeedback loop in an audio-only game. He used the EEG signal to control the amount of fear elicited by the game. He found that EEG has a potential of differentiating between fear and calmness.

Isaac [55] notes in his book on fears and phobias, that EDA and blood flow has been successful measures of fear. Especially EDA correlates with self-reported fear

See [56] for a review psychophysiological methods in game research.



2.4 Immersion and Presence

Immersion has been a term which has held many definitions though previous studies, and has been a source of great discussion amongst scholars within different fields of study [62]. The term immersion is originated from the Latin word *immergere* which means "dip into" and this meaning also correlates well with many definitions. Our definition of immersion is based on Cairns et al.'s definition:

"Immersion is a cognitive state that is influenced both by activities with the game, the social connections made through the game and external factors around the game." - Cairns et al. [63, p. 28]

This cognitive state that Cairns et al. talks about is a focused state of mind where the player has the feeling of being in the game. This means that immersion is a subjective feeling where one is involved with the play and one looses sense of time and the immediate environment. It is important to note that Cairns et al. makes a distinction between the term of presence and immersion, where they refer to presence as *spatial presence*, which is the convincing feeling of being transported to the virtual reality; a feeling of "being there".

To our definition and understanding presence is a term used to describe the extent to which one feels as being transported into a virtual environment. In other words it is how much one feels as "being in" a virtual environment or as Witmer and Singer puts it:

"Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another." - Witmer and Singer [64, p. 225]

To distinguish this from immersion here is an example: If a VE is presented as a concert the technological fidelity can create the illusion as if you are at the concert, but because the music is not to your liking, the music does not immerse you.

In the following sections the terms flow, cognitive absorption (CA), immersion and presence will be described as these are often mistaken for one another. As this study more focuses on immersion and presence, these will be described more thoroughly than flow and CA. The sections on immersion and presence exchange some common ground, and one must read both sections fully to understand their differences.

2.4.1 Flow

Flow is a state of mind where one is receiving challenges an interactive experience where the challenge matches one's skills. Flow is a fleeting experience, and is often experienced in action sequences of an interactive experience, such as a game, where you feel in control and challenged. You can not experience the state of flow if the challenge is not hard enough and you become bored, neither if the challenge is too hard and you become anxious. It can sometimes be hard to find a difference between flow and immersion, but Jennett et al. finds that flow is achieved only through a positive experience, where



immersion can be achieved in both a positive and negative experience [65]. Flow was originally fit into the context of interactive experiences by Csikszentmihalyi [66].

2.4.2 Cognitive Absorption

Cognitive absorption (CA) is a state of mind where all energy and concentration is focused on a specific task, especially software. This experience is often known to be experienced in work related context, where one can get so involved with the task at hand, that you shut off your immediate environment. CA is very much similar to flow in that it revolves around temporal dissociation, attention focus, heightened enjoyment, control and curiosity. For example it can occur when you are working on your computer with spreadsheets and the only focus is on the task and you do not notice anything happening around you. It differs from immersion in that CA does not require a VE to occur. [37], [65], [67] Jennett et al. puts it like this:

"A clear distinction between CA and immersion is that CA is an attitude towards information technology in general whereas immersion is the actual experience of a particular occasion of playing a videogame." - Jennett et al. [65, p. 643]

2.4.3 Immersion

The term immersion has been subject to different definitions through the past three decades [62]. However, there still has not been found any consensus on how to define and understand the term. The term has often conflicted with the term presence and both terms are related to somehow feeling some sort of connection to a virtual environment where one becomes incorporated or engaged with this environment.

Immersion has been used in various fields from books and arts to television an games, and therefore it introduces a variety of problems when trying to cover all fields at once. We are mostly interested in games and therefore we will focus on this aspect when discussing immersion.

The base research for the definition we use of immersion, was created by Brown and Cairns [37]. They did a study where they wanted to investigate and define immersion based on how gamers experienced immersion. This study found that immersion is used to describe the degree of involvement with a game. They found that immersion can be broken into three levels: Engagement, engrossment and total immersion.

Engagement is the first and lowest level of immersion that can be achieved, and to reach this state the player has to dedicate time, effort, and attention to the game. The player has to focus on the game rather than other tasks, and often if the player gets engaged he can lose track of time and wants to keep playing the game. To reach engagement it should also be fulfilled that the controls of the game are not a barrier anymore, and the player simply knows the controls to a degree where conscious mental effort is not allocated toward figuring out which buttons to push to execute a certain action.



The next level is engrossment, which besides being engaged as described by the first level, the player has to become emotionally involved with the game. To reach this level the game has to be constructed meaningfully for the player, which means that not only the controls and feel of the game has to be right, but also the tasks, graphics and plot of the game has to connect with the player. At this state the player is so emotionally involved with the game that his emotional state is affected by in-game actions and/or consequences thereof. This level is precursor for the last and highest level of immersion, total immersion.

Total immersion is the experience of the previous levels of immersion combined with presence. At this level players describe it as:

"[...] being cut off from reality and detachment to such an extent that the game was all that mattered." - Brown and Cairns [37, p. 1299]

This state requires the player to become transfered into this game world to an extent where the game is all that matters in the player's mind. To become totally immersed the player has to empathize with the character in the game and the actions that this character is taking, and the player feels as if he or she is that character. The player also has to identify with the game environment, therefore Brown and Cairns found that this total immersion is also often achieved in first person perspective games. These games allow the player to see through the main character's eyes into the world that the character is experiencing. This feeling of total immersion is however a fleeting experience, and is only felt in short periods of time, often in intense moments within the game.

Calleja [67] looks at immersion as a term which can be split into two sub terms, *immersion as transportation* and *immersion as absorption*. Immersion as transportation is the experience of being present in the environment where the context of the virtual environment convinces your senses that you are no longer in your immediate environment but have been transported to the virtual world. Immersion as absorption is the notion of one being engrossed within the game world, and lose sense of time.

Calleja's notion of immersion as transportation fits very well with our understanding of presence, where his notion of immersion as absorption fits into our definition of immersion. However, he does find the whole discussion of immersion, presence, flow and the correlation between these terms so confusing that he defines a new term called *involvement*. He states that involvement encapsulates both immersion as transportation and immersion as absorption into a single definition.

Ermi and Mäyrä [68] created a model to describe immersion called the SCI-model which consists of three different components of immersion. *Sensory immersion* which is the experience of immersion based on sensory information such as graphics and sound. The second component is *challenge-based immersion* which is the experience of challenge in a game. The last component is *imaginative immersion*, which is how the world, together with its characters and story, are able to make the player immersed. The component of imaginative immersion is the one that fits best with our definition of immersion, though the other two components do also partly apply. Challenge-based immersion is very much related to flow, though Ermi and Mäyrä argue that challenge based immersion is not



necessarily achieved by being in flow but is rather an expression of how hard a game is, when looking at their evaluation. However, one can argue that challenge-based immersion also relates to the first two levels of immersion (engagement and engrossment) of Brown and Cairns' findings as they propose that the controls of a game have to require no selective attention. The last component is imaginative immersion and fits into the third level of Brown and Cairns' model as the engrossment is achieved when one's state of mind empathizes with the characters and the world.

One of the main differences between Ermi and Mäyrä [68] and Brown and Cairns' [37] definitions of immersion is that we find that the SCI-model tries to somehow incorporate flow and presence into their model, instead Brown and Cairns try to exclude these components as much as possible. One could argue that their model of immersion is another description of selective attention though this is not exactly the case. Jennett [69] argues that one of the main differences between selective attention and immersion is feedback, where one can be engaged in an activity with almost no feedback such as attending a lecture. Immersion requires feedback from the activity, which is often a part of games. Jennett argues that:

"[...] immersion is a result of self-motivated attention which is enhanced through feedback from the game" - Jennett [69, p. 192]

Poels et al. [70] did an investigation into how gamers experience games. Their study involved 19 participants split into six focus groups based on gender, age, occupation and gaming frequency. They then did qualitative interviews with the different groups on how they experienced playing computer games, and they found that player experience can be divided into nine categories: Enjoyment, flow, imaginative immersion, sensory immersion, suspense, competence, tension, control and social experience. Here the terms imaginative and sensory immersion match our definition of immersion and presence respectively.

Slater et al. [71] have another view on immersion, as they depict immersion as a quantifiable description of a technology. Slater et al. sees immersion as something that describes how immersive a technology is, and they divide this description into five sub-categories: extensive, surrounding, inclusive, vivid and matching. *Extensive* refers to how many sensory inputs the system can accommodate. *Surrounding* refers to how capable the system is to provide the information in relation to physical movement, e.g. if one turns the head in a first person shooter (FPS) game, the sound should also move according to the new direction of the avatar. *Inclusive* refers to how good a system is to block out any real world stimuli. *Vividness* is how good a resolution (graphical) a system has. Lastly *match* refers to how well the system can interpret body movement, if you turn around a highly matching system should make sure that the audiovisual information is aware and adjusts to that. This definition of immersion does not fit our definition of immersion, though this interpretation of immersion does fit well into being a large quality of presence, as presence is related to the capabilities of a technology [63], [72].



Measuring Immersion

The subject of measuring immersion can be difficult since immersion is a subjective feeling, but some attempts to quantify immersion have been made. Jennett et al. [65] have created a questionnaire where they attempt to quantify immersion. The questionnaire consists of 31 questions grouped into five groups of interest which affect immersion: cognitive involvement, real world dissociation, emotional involvement, challenge, and control. *Cognitive involvement* is a description of how much one attends to the game, and deliberately involves cognitive energy to play the game. *Real world dissociation* is the ability of the game being able to make you forget about your immediate environment and real world problems. *Emotional involvement* is referring to how emotionally connected to the game you are, this being both characters and story. *Challenge* is how well the game challenges one's skills, and is based on flow, and finally *control* is rated regarding to how fluid or natural the controls are.

In the study of Jennett et al. [65] they did an experiment comparing levels of immersion in a game with or without audio. They did find significant difference between the nonimmersive (no audio) and immersive (with audio) version of the experiment, though the game play of the conditions was also very different and thus it cannot be said that sound as an isolated condition yielded any differences.

Ermi and Mäyrä [68] mentions briefly that sound has an influence on sensory immersion which might help to:

"[...] overpower the sensory information coming from the real world" - Ermi and Mäyrä [68, p. 7].

Based on this work, Grimshaw and Schott [3] investigated the influence of sound in FPS games. They argue that sound in such games has an immersive effect, in that sound helps to create an auditory ecology around the player. Furthermore, Grimshaw et al. [73] investigated the effect of sound on immersion and physiological states using EMG and EDA where subjects were exposed to music and diegetic sound, music only, diegetic only or no sound. They questioned their participants with the Game Experience Questionnaire (GEQ) [60] where they found that especially flow and immersion was influenced by the different sound conditions. They found no significant difference in either EMG or EDA data but found a correlation between sound being on and the level of immersion as measured by the GEQ.

In Nacke and Lindley [74] the authors did a study into flow and immersion as measured by the GEQ as affected by game play, mechanics, graphical fidelity and sound. They used sound in the immersive version of their experience, while only damped sounds was used for their boredom version. They found that the GEQ was sufficient in identifying flow, though the component of immersion was not found significant. The authors believed it to be the result of gamers not understanding the concept of immersion.



2.4.4 Presence

Presence is a term that is very closely related to immersion, and various definitions also identify presence and immersion as the same thing. Though in our definition presence and immersion are two different things, but closely related to each other. Presence separates itself in that it only is concerned with the feeling of being spatially present within the environment.

Brown and Cairns' definition of total immersion is also referred to as a state of presence [37]. We argue that presence can be achieved in isolation, though total immersion requires you to become transported into the virtual environment.

In a study by Slater et al. [75] a definition of *place illusion* is given, which is similar to the definition of presence in which we use in our study. He defines place illusion as the illusion of being in a place, even though you know you are not there. He defines this is qualia, an unmeasurable subjective sensation.

Measuring Presence

When it comes to measuring presence Witmer and Singer [64] has created a questionnaire to address this. The questionnaire called the Presence Questionnaire (PQ) originally consists of 32 questions with regards to rating presence on a Likert scale. Witmer and Singer originally found four factors which regarded presence: control, sensory, distraction, and realism. These factors were later revised where Witmer et al. [76] conducted a series of analyses on the PQ. This resulted in a new PQ with 29 questions in regards to four factors: Involvement, Sensory Fidelity, Adaptation/Immersion, and Interface Quality. The *Involvement* factor is about the psychological state of being mentally focused on a task or challenge. *Sensory Fidelity* is concerned with how consistent and coherent sensory information is to actions performed by the player. *Adaptation/Immersion* relates to the perceived experience of being enveloped by, included in, and interacting with the VE. Lastly *Interface Quality* refers to the quality and fidelity of the VE.

Usoh et al. [77] does however, question the use of questionnaires as method for measuring presence. Especially they find it problematic for cross-environments such as comparing an experience with HMD with an experience using a monitor. This was shown in an experiment they conducted where they utilized Witmer and Singers questionnaire, as well as their own Slater-Usoh-Steed questionnaire. Slater [78], [79] also questions the use of questionnaires as a method for obtaining quantified data on presence, though he state that:

"[...] at the end of the day, I use questionnaires because, for the time being, I do not know what else to do [...]" - Slater [78, p. 564]

Väljamäe et al. [80] did a study on presence and found that spatial presence was increased when using individualized HRTFs versus non-individualized when moving sounds were presented to the subjects. The subjects were exposed to one or three sound sources moving around them, and they were to report auditory ego-motion when they felt it. The



subjects were sat in a chair which could move, but did not move during the experiment. Around the chair four speakers was placed. This all contributed to an illusion that the chair would turn during the experiment, and this induced auditory ego-motion. They used a questionnaire to quantify presence, however no information about the actual questions were given.

Wiederhold et al. [35] reported a study where they investigated presence in relation to the physiological measures of tonic levels in EDA, HR, respiration rate and skin temperature. The participants were exposed to a simulated flight situation of take off, taxi and landing presented through either a screen or a HMD. They found that both EDA and HR increased as a function of self reported presence. Furthermore, Wiederhold et al. [81] did another study only investigating EDA and HR in the same flight simulation experience. This study did only investigate with a HMD and found a high correlation between percentage change in EDA and HR with self reported presence.

See [82] for a review of methods on measuring presence.



3. Goal

The aim for this study is to investigate how different spatialized audio rendering methods affects the level of immersion, presence and physiological responses. To the best of our knowledge this has not been covered in existing academia and we find it of relevance to both the academic society but also to the game development industry. Here we present the four null hypotheses that defines the goal of this study.

 ${\cal H}_{01} {:}$ Increased spatialization of sound does not increase the level of immersion.

 H_{02} : Increased spatialization of sound does not increase the level of presence.

 H_{03} : Increased spatialization of sound does not affect electrodermal response.

 H_{04} : Increased spatialization of sound does not affect heart rate variability.

We expect that level of immersion and presence will not follow a linear increase as a function of increase in spatial fidelity of sound. We expect levels between mono and stereo to be more apparent than levels between stereo and 3D, see Figure 3.1 for a graphical outline of this hypothesis. This is because we argue that the difference in spatial fidelity between mono and stereo is greater than the difference in stereo and 3D audio. This is because mono do not make use of any spatial components such as ITD, IID or SD, while stereo and 3D audio share the component of IID.





Figure 3.1: A graph depicting our hypothesis of how levels of immersion and presence correlates with spatial fidelity of sound. This graph is entirely hypothetical.



4. Experiment Design

Prior to this experiment design, three iterations in the form of pilot tests have been executed. Information about these pilot tests can be found in Appendix A.

Participants were sat down on a chair facing a table while wearing a HMD. Through the HMD they were exposed to a VE in the form of a dark living room, and the only controls were head rotation. This control scheme was chosen to minimize simulation sickness [83], [84]. The HMD was not calibrated for each participant. The participant's goal was to sequentially locate and collect glowing orbs placed at both different horizontal and elevated positions relative to the participant's avatar. Only one orb was present at any time. An orb could be collected by pointing it with the light from a flashlight, which was controlled by head movement. The flashlight's light cone was fixated in the center of the player's field of view (FOV). Each participant had to find 60 orbs in total. The orbs followed a fixed spawn order which were based on 20 fixed positions, all placed in front of the player. See Figure 4.1 for a top down view of the positions of orb spawns. Each orb had an internal timer, and the player's task was to collect it before the timer ran out. A system was implemented that would make sure that within every 10th orb one failure would occur, causing there to be 6 failures during a game condition. The first orb had a timer of 10 seconds. If the participants successfully collected the orb, the next orb's timer would be 9.1 seconds. For each consecutive successful collected orb, the timer would be reduced by 0.9 seconds. At some point, the timer would be so short, that it was practically impossible for the participants to collect it, as the time to collect an orb was 2 seconds. This means at the 10th orb the timer would be 1.9 seconds. Upon failure, the next orb's timer would be removed, making it impossible for them to fail. The following orbs' would have no timers, until passing the 10th orb, then the timer will be reinstated. This cycle would continue to the end of a game session. This method ensures that all participants would encounter an equal amount of failures. See Figure 4.2 for an example on how the difficulty would function. Participants were not informed of the exact timer pattern, only that the timer would change during the experiment.

By default, an orb would rotate slowly and emit a pink colored light. The closer the orb's internal timer was to zero, the faster it would rotate. When the player pointed the light cone at the orb, it would gradually change its light color towards green. If the player moved the light cone away from the orb, the orb's light color would turn back to pink and the player would have to start over in collecting the orb. In Figure 4.3 one can see how an orb look like both passively and while it is being collected. Upon pointing at





Figure 4.1: A top down view of the environmental setting. Each pink dot represents a location where an orb could spawn. The red shaded area is an approximate view on the FOV for the player. Proper light has been disabled for this image for illustrative purposes.



Figure 4.2: Here is an example on how the timer changed during the experiment. As participants collected orbs, the internal timer would reduce for each successful consecutive orb collected. In this example, the fictive participant has successfully collected so many orbs that he reaches the threshold of which an orb is impossible to collect in time. This situation occurs both at the 10th orb and the 30th orb. But because the next orb (11th and 31th) resets the orb cycle, the internal timer is set to 10 seconds. At the 16th orb, the player also fails to collect the orb in time, but there are still four orbs before a reset in orb cycle, and the internal timer for the 17th to the 20th orb are therefore set to infinity.





Figure 4.3: Screenshots taken from the game. In both images the orb, which the participant had to locate and collect, can be seen. Left image is an orb in its idle state. Right image is an orb, currently being collected, as indicated by the green light color.



Figure 4.4: Screenshots taken from the game. The left screenshot represents the visual effect occurring if the player failed to collect the orb in time. Following this visual effect was a burst of horrific sounds rotating around the player. The image to the right represents the visual effect that was shown when the player succeeded in collecting an orb.

the orb for two consecutive seconds, the orb would disappear followed by a visual effect. After an additional two seconds, a new orb would spawn. The spawn of a new orb was indicated by the flashlight flickering a pink color.

If the player failed to collect an orb, a red visual effect would be present and a scary event would begin. In Figure 4.4 the two possible visual effects, related to an orb disappearing, can be seen. Each scary event consisted of three audio sources. Each sound would start close to the head, and would rapidly both rotate around the players head while moving away from the player. All the loud sound cues were of horrific nature ranging from screams to ghostly echoes. These events were designed to create emotional arousal in the player. Scary sounds was chosen as stimuli as it has previously been shown that scary sounds induce fear and arousal [36], [48], [54]. For the statistical analysis, the first three scares would be considered training scares, while the remaining three scares were used for further analysis.

It was an intentional design that audio could not be used as an assistance to complete the



given task e.g orbs emitting sound from their position. This was to ensure that skill and localization training did not become a bias, which was seen in one of the pilot studies, see Appendix A.2.1.

The experiment would follow a within-subject design and would consist of two game conditions, see Figure 4.5. This was necessary for eliminating individual differences both in terms of physiological responses and prior gaming experience. Different from each game condition, was the auditory rendering method. All sounds were either rendered with stereo or 3D audio. Each participant would be exposed to both auditory rendering methods during the experiment. The order of audio rendering was counterbalanced. After each game condition participants would have to answer questionnaires on self-evaluated immersion and presence.



Figure 4.5: The order of experimental conditions. All participants were exposed to two game sessions where sound was rendered either using 3D audio or stereo. The first three scares would not be included for statistical tests, and were considered training, while the remaining three scares were used for further analysis. After each game condition, participants had to fill out two questionnaires on immersion and presence. The two groups represents participants with different starting conditions to create a balanced test design.

Before participants could participate in our study, they had to fill out a consent form where they gave consent of having normal or corrected to normal sight and hearing. See Appendix C.3. Before participants began the experiment they were told that: they had to collect as many orbs as possible; how to collect an orb; that orbs have an internal timer; and that "something" would occur if they did not collect an orb in time. After the experiment was done an informal discussion with the participants was done, asking them whether they noticed any difference between the two game conditions they have experienced. These discussions were not recorded.

The goal with the experiment was to induce immersion and presence into our participants through a game were audio was rendered with either 3D audio or stereo. The questionnaires were added to record the participants self-evaluated level of immersion and presence. During the entire experiment, we would measure their EDA and HRV responses. The EDA and HRV responses caused by the scary events are values of interest, and were later extracted and analyzed along with the questionnaire responses.



4.1 Materials

This section describes the different materials used for implementing and setting up the game experiment.

4.1.1 Virtual Environment

The primary tool that was used for implementing the game was Unity3D [34] (henceforth referred to just as Unity). We used Unity 5.0.1 for the implementation which provided physically based shaders and real time global illumination. We used this to achieve light and reflections of materials which provided a pleasant viewing experience for participants. All scripting for the experiment was done using C# and this was used for controlling the system in terms of both gameplay, environmental events, and game condition control.

The HMD we used was an Oculus Rift DK2, which was interfaced with the Oculus SDK version 0.4.4-beta for Unity.

The virtual environment that we created for this project was a virtual living room. The interior was created with 3D models from Whodat's Home Interior Pack [85] which is a collection of 3D models of typical western interior.

The VE contained four background sounds that played continuously throughout each condition. These background sounds were placed all around the listener: A radio playing a Thai broadcast played on a table to the right of the player; A light bulb flickering periodically played in a loft lamp in front-above the player accompanied by a visual light flicker; A tick-tock sound of a clock was played at a wall clock left-above to the player; A rain sound was present at the window behind the player. All sounds were chosen to represent a broadband spectrum, see Appendix C.4 for spectrum analysis of the sounds. In order to avoid breaking the illusion of a living room the background sounds were chosen based on their relation with objects one can find in a living room; hence all sounds were of diegetic nature.

The virtual environment was rendered using a Windows 8.1 computer with a 3.4 GHz Intel Core i5 3570 processor with a Gigabyte GeForce GTX 680 OC graphics card. Sound was output through the on-board Realtek ALC892 sound card on an Asus P8Z77-V PRO motherboard, through a pair of Beyerdynamics DT 990 PRO headset.

The experiment was conducted in a quite laboratory environment with the conductors sitting approximately three meters from the participant, observing the experiment through a monitor rendering the same image as the Oculus Rift. This allowed for the conductors to observe the behavior of the participants.

4.1.2 Sound Rendering

The audio rendering engine used for this project was a plug-in for Unity called 3Dception created by Two Big Ears [86]. This plugin both supports stereo and 3D audio render-


ing. It is to us not known which HRTFs is used and at what resolution, or method of interpolation. We performed a test confirming that the 3D audio implementation in 3Dception utilizes IID, ITD and SD. The stereo rendering is to our knowledge done with a simple panning method. An important note of this implementation is, whenever the sound source is either directly to the left or right of the player, the amplitude of one channel would have full intensity, while the other would have zero. We performed a test confirming that the stereo implementation does not utilize ITD.

4.1.3 Data Collection

The metrics recorded from the game were written to comma separated values (CSV) files at a sample rate of 50Hz.

The physiological data was collected using the Shimmer 3 device by Shimmer which has the possibility of connectivity via Bluetooth. The Shimmer device was used to collect EDA and HR. The EDA was collected with two electrodes placed on the medial phalanx of the ring and middle finger and HR data was collected using PPG where the sensor was placed on medial phalanx of the index finger. Data collected with the Shimmer device was written to a CSV file with a sample rate of 51.2Hz.

The data that was logged for this experiment were:

- Camera movement (continuously)
- Angle between direction and target (continuously)
- Target spawn time and position relative to player
- Target acquisition time
- Audio events
- Failed or successful acquisitions
- EDA (continuously)
- PPG (continuously)

The data stream from the Shimmer device and the experimental software can become slightly unsynchronized, due to initiating each independent data stream manually. A more throughout description can be found in Appendix B.1.

The ShimmerCapture 0.2.0 software was run on a Macbook Pro 8.2 using Parallels emulating Windows 8.1 as the ShimmerCapture software only works for Windows. The Bluetooth connection to the Shimmer device was established using a Belkin F8 Bluetooth dongle.



4.1.4 Questionnaires

Self-evaluated immersion was measured using the Immersion Questionnaire developed by Jennet et al. [65]. The questionnaire measures for total immersion (31 questions) and the sub-categories: cognitive involvement (10 factors); real world dissociation (6 factors); emotional involvement (12 factors); challenge (5 factors); and control (8 factors). Self-evaluated presence was measured using a questionnaire created by UQO Cyberpsychology Lab [87], a revised and shortened version of Witmer and Singer's presence questionnaire [64]. This questionnaire is measuring for total presence (22 questions) with the subcategories of: realism (7 factors), possibility to act (4 factors); quality of interface (3 factors); possibility to examine (3 factors); self-evaluation of performance (2 factors); and sounds (3 factors). The presence questionnaire also contains two questions on haptic, though this was found irrelevant for this study. Both the immersion and presence questionnaires.



5. Data Extraction

37 participants participated in our study whereof 19 were males and 18 were females with an average age of 24.03 years. All reported normal or corrected to normal sight and hearing. All 37 participants had their EDA and HR measured and filled out a presence questionnaire. 20 out of the 37 participants filled out an immersion questionnaire. 16 of 37 participants EDA and HR were discarded from further analysis: 12 of 37 participants were discarded due to technical complications; 2 of 37 participants were discarded for not having any EDA during the two game conditions; 2 of 37 participants questionnaires, EDA and HR data were discarded due to excessive talking during the experiment and bad participation. Because of data exclusion, participants starting with stereo were represented with one more sample, than participants starting with the binaural condition. In order to obtain balanced results, a random participant with a stereo starting condition were excluded from the data analysis.

For participants who did the immersion questionnaire, the total experiment took approximately 45 minutes, while for those participants who did not, the total experiment took approximately 30 minutes. The participants' answers to questionnaires can be found in Appendix C.5 and visual representation of all participant's physiological data can be found on the Appendix CD in the folder 'Participant Data'.

5.0.5 Deriving Sound-Related Events

For this study, the EDA events which are caused by the scary events will be referred to as Sound-Related (SR) events, in order to distinguish them from other EDA events. In order to obtain SR events from the participants the EDA signal had to be further analyzed. In order to analyze the EDA signal, we made use of EDA Toolbox, a MATLAB framework, which automatically detects and classify EDA events. An EDA event is defined as a pair of values consisting of a valley and a peak. An example of how a participant's EDA signal appears, can be seen in Figure 5.1.

Each participant's EDA signal went through the following process: Before being able to use EDA Toolbox, the participant's EDA signal required a unit conversion from kOhms to microSiemens. The following conversion took place: mSI(kOHM) = 1/(kOHM) * 1000 see Figure 5.2

The EDA signal was passed into a 5th order low-pass Buttersworth filter with a cutoff



Time(s) Figure 5.1: The EDA signal from a participant measured in kOhm.



Figure 5.2: A participant's EDA signal after being converted to microSiemens from kOhm.

frequency at 1Hz to eliminate quantization artifacts, see Figure 5.3. A function implemented in the EDA Toolbox was used for this.



Figure 5.3: The left image is an example of the EDA signal before the 1Hz low-pass filter. The steps are caused by a quantization artifact in the raw EDA signal. The right image is after the filter was applied.



Next, automatic detection of EDA events was performed on the signal, using another tool of the EDA toolbox framework. The algorithm identifies peaks and valleys from a first time-derivative of the signal, by identifying peaks and valleys when the sign in the first derivative of the signal changes. A change in sign either indicates a valley or peak depending on the previous stage of the change. The output of the algorithm is a range of EDA responses. However, the algorithm can take an optional input parameter, making it possible to filter out responses not fulfilling some input criteria such as slope, amplitude size and rise time. For this study a response amplitude of minimum 0.02 microSiemens was required before an EDA event would be registered. See Figure 5.4.



Figure 5.4: An illustration of the different valleys and peaks identified. Red dots represent valleys while blue dots represents peaks. A pair of a valley and a peak is considered an EDA event.

The EDA Toolbox framework is capable of classifying EDA events based on applying a window of interest (WOI). The position of the window was given by an onset, which in our case was the onset of a scary event. The EDA Toolbox simply looks up which responses lie within the WOI and groups them with the event. This classification was used to identify which EDA events could be be identified as SR events. The size of the WOI was 2.5 seconds before onset and 5 seconds after onset. The 2.5 second before the onset were to capture valleys for participants who felt arousal due to anticipation, and the 5 second is based on the response delay for an EDA event [56]. An EDA event could still be considered as a SR event if its valley lied within the WOI even though its peak did not. By plotting out the mean response for each corresponding condition, and zeroing the signal at the lower bound of WOI value, we got the following Figure 5.5. Here we see very low slope right before the onset, while after the onset a high slope occurs. Around 4 seconds after onset, the peak is reached. Based on this we argue that the size of the WOI has been properly selected, as it captures both the valley and the peak of a general SR event.

A SR event can consist of multiple EDA events, the SR event is based on the closest valley to the onset, while the peak as the highest peak after the onset and the SR valley. The value of the SR event is the magnitude from the selected valley to the selected peak. This is the value which is used for further analysis.



Figure 5.5: The mean SR event response, summarized from responses across the two conditions. The red line indicates the onset of a SR event, and the triangle indicates a zero slope after onset.

See Figure 5.6 for a visual representation on the determination of SR events. An additional observation from Figure 5.6 is that participants varied significantly in responses. Some participants showed clear and steep changes in EDA upon exposure of scary sound effects, while other participants did not have any visible responses.

For the statistical analysis we considered the first three SR event values for each group to be training events and discarded them from the analysis. When all SR event values were extracted, the values were grouped with their corresponding independent variable stereo or 3D audio. A mean was computed for each group.

5.0.6 Deriving Heart Rate Variability

In order to extract the HRV from the participants' HR data, see Figure 5.7, the HR data had to be converted to R-R intervals. R-R intervals are defined as the time between two R peaks (two heart beats), and is the reciprocal of HR. In most cases R-R intervals are measured in milliseconds. The following unit conversion took place RR(HR) = (60/HR) * 1000.

However, we were only interested in obtaining the HRV at two specific periods during the experiment: from the beginning of the 4th scary event to the end of the 6th scary event, in both corresponding groups, stereo and 3D audio.

Next, the signal would have to be decomposed into its corresponding frequencies components. This was done using a fast Fourier transform (FFT). The output of the FFT is a range of values, where each value consist of a real and an imaginary number, describing the amplitude of R-R in certain frequency bands. By taking the absolute value for these real and imaginary numbers you get the magnitude of the R-R intervals, and if those values are then squared you get the power. HRV were retrieved by taking the integral of the power of low frequencies (LF, 0.04Hz-0.15Hz) over the integral of the high frequencies (HF, 0.15Hz-0.40Hz). The expression for HRV is given in Equation 5.1.





Figure 5.6: A few EDA signal examples from three different participants. The green line indicates the onset on where a scary sound were played. The WOI formed around the onset is marked with two red lines, and is the area in which EDA events are observed. The larger red circle indicates the valley closest to the onset, while a larger blue circle indicates the largest peak that lie after the onset, and after the selected valley. The difference in height between these two points is the SR event value that was used for further analysis.

$$LF/HF = \frac{\int_{0.04Hz}^{0.15Hz} |FFT(RR)|^2 df}{\int_{0.15Hz}^{0.40Hz} |FFT(RR)|^2 df}$$
(5.1)

 $36 \, \text{ of } 85$



Figure 5.7: A participant's heart rate in beats per minute during the entire experiment.

In Figure 5.8 an illustration of the R-R interval power spectrum from a participant can be seen. The red line indicates the split between LF and HF. This value is the HRV, and was the value used for further analysis.



Figure 5.8: The R-R interval power spectrum from a participant. The red line indicates the split between LF and HF, LF to the left and HF to the right.

5.1 Observations and Participant Discussions

After the participant had completed both game conditions and answered all questionnaires, an informal discussion were initiated with the participant. The participant was asked whether they perceived any difference between the two game sessions. Most participant stated that they did not perceive any difference. However, a few participants stated that they perceived a change in volume. Post-experiment, we looked into this statement and we saw a difference in signal intensity in the output of the rendering systems, see Figure 5.9.

We believe this is caused by an increase in intensity when applying a HRTFS filter to



Figure 5.9: A difference in signal intensity can be seen between the two audio rendering methods. Both signals is the outcome of a 100 milliseconds burst of white noise emitted from a sound source three meters in front of the listener. The left signal is rendered with 3D audio and the right is rendered with stereo.

a signal. When observing the frequency spectrum from a stereo signal and a 3D sound signal, see Figure 5.10, we see that their spectrum differences are different, however, this phenomenon is already explained in Section 2.1. However, in the 3D sound's frequency spectrum we see that a larger intensity is observed in the higher frequency bands than in the stereo sound's frequency spectrum. The physical properties of wave forms is that higher frequencies produces more energy than lower frequencies. Because 3D audio contains more high frequency components than stereo this explains the intensity difference between the two, which is audible.





Figure 5.10: Frequency spectrums for a white noise signal of 1 second duration rendered either with stereo (Top image) or 3D audio (Bottom image), recorded at a three meter distance in front of the audio listener. Here we see that the frequency spectrum is much different, which is caused by applying a HRTFS filter. However, what is also of interest is that the 3D sound signal contains more high frequency components than the stereo sound signal. This means that 3D audio contains more energy than stereo which is audible.



6. Results

Because the experiment followed a within-subject design, a paired test can be applied. It would be preferred if a parametric test such as Student's t-test was applicable, due to its strong power, but the data had to fulfill the assumptions of normality.

In a paired design, each group does not require normality, but the difference between them does. In order to address data normality one can use normality tests such as the Kolmogorov–Smirnov test or the Shapiro–Wilk test. Visual inspection of data is another alternative such as a quantile-quantile plot (Q-Q plot) or an empirical distribution function (ECDF). In Figure 6.1 are illustrations on how normality distribution should appear in a Q-Q plot and a ECDF plot. These plots were generated for illustrative purposes $[\bar{x} = 0, s = 1, N = 10000]$



Figure 6.1: Illustrations of how a desired normality distribution should look like for a Q-Q plot and an ECDF plot. The left is a Q-Q plot and the right is an ECDF plot.

As an example, the normality plot for the participants' SR event values can be seen in Figure 6.2. From this we can see that the data appears to be close to normal, but not yet acceptable from our point of view. Additionally, when running a Kolmogorov–Smirnov test, the results indicate that the data is non-normal (Kolmogorov–Smirnov test, p < 0.05).

For testing the data a Wilcoxon signed-rank test was used. For a test result, mean (\bar{x}) , standard deviation (s), median (\tilde{x}) , and interquartile range (IQR) is reported. When using any paired test, it is the differences between two groups that are being tested. Therefore, the reported mean or median's direction is of importance. For our analysis,



Figure 6.2: Visual inspection of the SR event values' normality. Left plot is a Q-Q plot. Right plot is an ECDF plot.

a positive mean or median indicates a larger effect towards 3D audio while negative indicates a larger effect towards stereo.

For this study a critical value of $\alpha = 0.05$ was chosen.

A significant difference was found for the differences in SR event values (Wilcoxon signed-rank test, p < 0.05, N = 22) $\bar{x} = 0.15 \ \mu\text{S}$ (s = 0.25), $\tilde{x} = 0.129 \ \mu\text{S}$ (IQR = 0.34). The significant difference was maintained even for minor changes to the WOI.

No significant difference was found for differences in HRV. (Wilcoxon signed-rank test, p = 0.97, N = 22) $\bar{x} = -0.01 \text{ LF/HF}$ (s = 2.64) $\tilde{x} = 0.11 \text{ LF/HF}$ (IQR = 1.17).

For testing the questionnaire responses a non-parametric test was used regardless of normality. That is because questionnaire responses are of ordinal data type.

No significant differences was found in the immersion questionnaire (Wilcoxon signed-rank test, p > 0.05, N = 20). See Table 6.1 for a summary of the immersion questionnaire scores. In Figure 6.2 a summary of the immersion questionnaire score differences between the two groups is presented.

	3D audio					Stereo				
	\bar{x}	s	\tilde{x}	IQR	\bar{x}	s	\tilde{x}	IQR		
Total Immersion	67.56	10.08	69.50	10.00	65.28	17.31	69.00	14.00		
Challenge	14.72	2.24	15.00	3.00	13.83	1.86	14.00	1.00		
Cognitive Involvement	36.50	4.71	36.00	6.00	34.94	7.88	37.00	8.00		
Dissociation	1.83	3.87	1.50	6.00	2.44	3.97	2.50	5.00		
Involvement	37.22	5.14	37.50	9.00	35.78	8.20	37.00	8.00		
Control	25.00	2.70	25.00	3.00	24.39	3.68	24.00	4.00		
Self-Evaluted Immersion	6.33	1.28	6.00	1.00	6.44	1.38	7.00	3.00		

Table 6.1: Depicts the mean, standard deviation, median and interquartile range separately for the two groups, 3D audio and stereo for the immersion questionnaire responses.

No significant differences were found in the presence questionnaire (Wilcoxon signed-rank test, p > 0.05, N = 37). See Table 6.3 for a summary of the presence questionnaire scores.



Table 6.2: Depicts the mean, standard deviation, median and interquartile range from the two groups, 3D audio and stereo, subtracted from each other for the immersion questionnaire responses.

	P-Value	\bar{x}	\mathbf{S}	\tilde{x}	IQR
Total Immersion	0.90	2.28	14.12	3.00	13.00
Challenge	0.41	0.89	2.89	0.00	2.00
Cognitive Involvement	0.42	1.56	5.60	0.00	6.00
Dissociation	0.66	-0.61	3.71	0.00	4.00
Involvement	0.71	1.44	6.66	1.00	4.00
Control	0.38	0.61	3.38	1.50	5.00
Self-Evaluated Immersion	0.65	-0.11	1.18	0.00	2.00

In Figure 6.4 a summary of the presence questionnaire score differences between the two groups is presented.

Table 6.3: Depicts the mean, standard deviation, median and interquartile range separately for the two groups, 3D audio and stereo for the presence questionnaire responses.

		- D				a.			
	3D audio				Stereo				
	\bar{x}	s	\tilde{x}	IQR	\bar{x}	s	\tilde{x}	IQR	
Total Presence Score	79.65	10.26	80.00	14.00	79.53	10.82	80.50	13.00	
Realism	38.94	5.89	40.00	9.00	38.32	6.57	39.50	7.00	
Possibility to Act	21.91	3.32	22.00	4.00	22.00	2.89	23.00	5.00	
Quality of Interface	-7.76	3.25	-6.50	6.00	-7.24	3.29	-6.50	4.00	
Possibility to Examine	14.50	2.65	15.00	3.00	14.32	2.92	14.50	4.00	
Self Evaluation of Performance	12.06	1.58	12.00	2.00	12.12	1.51	12.50	2.00	
Sound	14.56	2.80	14.00	4.00	15.24	3.23	16.00	4.00	



Table 6.4: Depicts the mean, standard deviation, median and interquartile range from the two groups, 3D audio and stereo, subtracted from each other for the presence questionnaire responses.

	P-Value	\bar{x}	\mathbf{s}	\tilde{x}	\mathbf{IQR}
Total Presence Score	0.96	0.12	7.93	0.00	12.00
Realism	0.42	0.62	4.76	1.00	4.00
Possibility to Act	0.84	-0.09	2.70	0.00	4.00
Quality of Interface	0.23	-0.53	2.54	0.00	3.00
Possibility to Examine	0.64	0.18	1.82	0.00	2.00
Self Evaluation of Performance	0.87	-0.06	1.61	0.00	2.00
Sound	0.08	-0.68	2.61	0.00	4.00



7. Discussion

From this study there were no findings that directly indicated players became either more immersed or achieved a higher level of presence when exposed to 3D sound compared to stereo. Based on discussions with the participants and results from the questionnaire responses, it appears that there have been no noticeable difference between the two conditions. However, it appears the spatialization have had a subconscious effect, based on the results from the SR events.

The first explanation for the difference in SR events could be due to the intensity difference between the two rendering systems, which supports the findings of Toprac and Meguid [36]. This also fits with 3D audio having a larger response as a product of larger intensity. However, only a few participants stated they noted a difference, so it is up for discussion whether this difference in intensity manifested itself into the participants who did not notice any intensity difference.

Another possibility could be that the SR event difference could be due to a subconscious difference in level of immersion or presence. If a participant have felt more immersed, he may have felt a stronger emotionally response upon missing an orb. If the participant have felt more presence in the VE, horrific sounds may have appeared more realistic, and thus increasing the physiological response. If this is the case, with either more subconscious immersion or presence, one could argue that self-evaluation of immersion and presence is not sufficient when subtle changes are made, supporting the claims of Slater [78], [79]. We argue that the difference in audio rendering system is a subtle change as most participants did not notice the difference.

We argue that the difference in intensity or a change in immersion or presence is the primary cause for the difference in SR events, however we do propose alternative suggestions:

One could assume that the difference in SR events were triggered by the body's autonomic fight-or-flight systems, preparing the body to flee from the perceived danger. This fits well, if the difference in SR events is caused by a difference in intensity, as the danger would be perceived as either closer of further away from the listener, as suggested by Garner and Grimshaw [48]. If this is not the case, then this hypothesis would be contradicted by the lesser spatial fidelity elicited by stereo rendering makes it more problematic to localize the position of the danger. In this case, we believe that the stereo condition should have induced a larger response than 3D audio, because the body should prepare



incoming danger from any position, which would support the findings of Ekman and Kajastila [47].

Lastly, the difference in SR events could be caused due to level of frustration caused by an inability to localize origin of audio. The feeling of frustration should induce a lower level of arousal based on the circumplex model of emotions by Russell [43]. This contradicts with our findings because stereo compared to 3D audio have lesser spatial fidelity, and therefore stereo should have elicited a higher response.

We found no significant difference in HRV even though a significant difference for SR events were found. We postulate that this is because of an insufficient data collection. The length of the signals used to obtain the HRV was of approximately two minutes length. When performing a Fourier transform on this signal only a few frequencies band within the range of 0.04Hz to 0.40Hz had any noticeable energy, which is apparent when observing a the power spectrum graph, see Figure 5.8. Another explanation of the lack of significant difference could be because HRV is not suitable to detect the subtle change between the two game conditions. However, this would contradict with the significant difference found in SR events, if this difference is caused by a higher level of immersion or presence. If the body is capable of responding with an increased EDA upon a subtle change why would HRV be any different? Further investigation has to be performed before anything specific can be said of HRV's behavior

If a higher level of immersion where achieved, one could argue that in relation to Brown and Cairns'[37] notion of three levels of immersion our participants did not reach the highest level of immersion. This we base on how the scores for presence turned out as well as immersion scores. It can though be discussed whether they reached the first or second level of immersion or even neither of them. However, we do argue that often players reached at least the first level, because of the natural interaction with the HMD and how they answered the immersion questionnaires.

An important observation to add to this discussion is the presence questionnaires responses related to sound. Even though the result was insignificant, the p-value were close to the set critical value, and we believe therefore it is worth for a discussion. These results indicated that stereo was easier to localize with compared to 3D audio. This brings up some valid questions: Studies have shown that localizing with 3D sound is easier than with stereo. Does this disprove the other studies? If stereo is easier to localize with, compared to 3D audio, is the difference in physiological response caused by a frustration? We argue that 3D audio is still better for localization based on previous findings [10], [13]. In the experiment, participants had no task which involved auditory localization, which means there were no incentive to pay attention to audio positions. Listening with stereo audio is distinct to how we perceive sound in real-life and may gain attention as it appears out of place. With more focus on localization during the stereo condition, participants might have responded accordingly. As an extension to previous argument, stereo may be better at localizing audio near 90° to -90° azimuths. At these azimuths during exposure of stereo audio IID will be zero, which is not the case for 3D audio. Zero or near zero IID is uncommon and rare in real-life and may stand out to real-life audio localization. When the horrific sounds were rotating around the player,



zero IID would occur multiple times during a single event, shifting zero IID between ears. When participants are asked, how easy it was to localize sounds, it is first of all ambiguous whether it is referring to the static or horror sounds, but how can one confirm one's personal localization abilities if there are no visuals? We therefore believe, that the responses from the questionnaire is heavily weighted by attention and lack of visuals to confirm one's ability to localize sounds, and can therefore be difficult to interpret due to the nature of the experiment.

We believe a reason for not seeing any differences in the questionnaire responses is due to the experimental design. The participants were told that their primary goal was to get as many orbs as possible. The player took action right as the game began, and had no option of pausing. While always being active in a search task, the participant were induced high perceptual load and was required to put all their attention into the task. Ignoring task-irrelevant distractors is seen for tasks which requires high perceptual load [88]. We believe that no attention were put into the auditory environment, hence no distinguishable changes were observed from the participants between the two game conditions. An extension to this argument: the game's design may not utilize the spatialization of audio to such a degree that a noticeable difference were possible. The player had no option of moving around in the VE leaving out the spatial audio cue of changes in distance. This cue may play an important role for immersion and presence for auditory spatialization, and should be further investigated.

Another factor for the lack of noticeable difference between the questionnaire responses may be the usage of non-individualized HRTFs. Using non-individualized HRTFs have a reduced effect when comparing it to individualized HRTFs, which is supported by the study of Valjamae et al. [80]. However, more research is required to investigate this.

An approach to validate this study's findings, one could perform a similar study, but instead compare mono and 3D audio. Because mono audio consist of fewer spatial cues than stereo, the level difference in spatial fidelity will be larger between mono and 3D audio than between stereo and 3D audio, see Figure 3.1. Using the same experimental design as this study, we hypothesize that a significant difference could be found in the questionnaires of such study.

One could also argue, that in order to measure differences in immersion, a certain threshold of immersion has to be achieved. As an example, if a study attempts to investigate the differences in immersion of the effect of small visual distractors against no visual distractors. Players who play a boring and uninteresting game, may feel little to no difference in immersion when adding visual distractors. Compared to players who play a highly immersive first person shooter, where the visual distractors interfere with the experience. It is therefore also of question, whether the game implemented in our study induces enough immersion for participants to feel any measurable difference.

With 20 data samples for immersion questionnaire scores and SR event values and 36 presence questionnaire scores, it is also worth discussing whether the amount of gathered samples is sufficient to state anything conclusive. Because of the varying nature of physiological responses, more samples would benefit this study, but we believe sufficient



information can be derived from our data, primarily because we made use of a withinsubject design. If a between-subject design were used instead, we believe 20 samples would have been too few.



8. Conclusion

In this study the effects of spatial audio using EDA, HRV, and questionnaires for immersion and presence for players exposed to either stereo or 3D audio was addressed. No significant difference in immersion or presence was found, however a significant difference was found in phasic EDA events (SR events). The results suggest that even though players did not perceive an audible difference, the difference in spatialization have a subconscious effect. Authors believe this effect is either caused by a subconscious change in level of immersion or presence or a difference caused by the different intensity levels between the two audio systems, but further investigation is required before anything conclusive can be given. No significant difference were seen for HRV.



9. Future Work

Based on the significant increase in SR events in the condition of 3D sound this will have to be investigated further. We did not find any conclusive evidence in this study that supported this tendency.

One of the first approaches to investigate the difference in SR events is to do the same experiment, but comparing two different rendering methods whose output is of equal intensity. This would help in order to determine whether the significant difference were either caused by a change in intensity or a change in spatialization of audio.

Another approach to investigate this is to observe whether the same tendency will occur when arousal is raised by other emotional reactions than fear. One could attempt to induce excitement or happiness as these are also placed high in arousal in Russell's [43] circumplex model of emotions, but opposite of fear they are of positive valence. As an additional emotion one could investigate whether the effect could be caused by frustration. It would also be interesting to investigate this phenomenon in environments inducing low level of arousal, such as environments with meditational purposes, where we believe 3D audio would induce lower arousal levels than stereo.

One of the main concerns of this study has been the experiment design. In future works it would be of interest to investigate at least two different types of design. One would be to see if immersion and presence is affected in an experiment where the game element is left out of the question, so that it simply becomes an interactive experience in a VE. This would help to introduce moments where the player have time to investigate the environment more, and therefore be consciously aware of the environmental sounds. One could in such a scenario use different tools to guide the player's attention around the environment, or simply let it be a self-exploratory experience. Secondly it would be interesting to use the sound active in the context of a game. Make the objects that the participant has to locate audible, and therefore use the spatial fidelity of audio as a tool. This would possibly affect the success rate for the player, and therefore the feeling of success might have an effect on the level of immersion or presence, so one would have to consider this in the game design.

As an addition to the physiological measurements, one could attempt to make use of an EEG or EMG. Through the usage of these methods, we believe it may be easier to distinguish what type of emotion a player is experiencing. For this study, the usage of EEG or EMG could clarify whether the difference in EDA were caused by a fight-or-flight



response or a response as a product of frustration. We hypothesize that the usage of these methods could yield a better understanding on the underlying effects of spatial audio on player experience.

Another interesting area where one could investigate the effect of 3D audio on immersion and presence is to utilize the technology in already existing games where immersion and presence is reported high. At first, this should in our opinion be limited to games which utilize a first person view, though such an investigation could include both games played with a regular screen or HMD. This would allow for a comparison of games which already have an established gameplay, and therefore has proven themselves as interesting games. This could help to eliminate the feeling of experimental/laboratory games, that in our opinion, often has a negative influence on the experience. Such findings would be applicable to the industry of games.

As an alternative method to using questionnaires, one could perform a study with a qualitative approach. Such an approach could include both individual and focus group interviews, video analysis of user behavior or vocal transcription. We believe such methods could be better for investigating the effect of subtle changes. A possible scenario could be that participants can not recall any differences between the two game conditions, but by putting emphasis on some of the participant's behaviors, the participant may recall a more precise evaluation of the participant's experience, than what can be found in an investigation using questionnaires.



Bibliography

- M. Lalwani, Surrounded by sound: how 3d audio hacks your brain, Feb. 2015. URL: http://www.theverge.com/2015/2/12/8021733/3d-audio-3dio-binauralimmersive-vr-sound-times-square-new-york (visited on 05/21/2015).
- D. Murphy and F. Neff, "Spatial Sound for Computer Games and Virtual Reality", in *Game Sound Technology and Player Interaction: Concepts and Developments*, M. Grimshaw, Ed., Sep. 2010, pp. 287–312.
- [3] M. Grimshaw and G. Schott, "Situating Gaming as a Sonic Experience: The acoustic ecology of First Person Shooters", en, 2007.
- [4] K. Mcmullen, "The potentials for spatial audio to convey information in virtual environments", in 2014 IEEE VR Workshop: Sonic Interaction in Virtual Environments (SIVE), Mar. 2014, pp. 31–34.
- [5] D. A. Mauro, R. Mekuria, and M. Sanna, "Binaural Spatialization for 3d Immersive Audio Communication in a Virtual World", in *Proceedings of the 8th Audio Mostly Conference*, ser. AM '13, New York, NY, USA: ACM, 2013, 8:1–8:8,
- [6] H. Møller, "Fundamentals of binaural technology", Applied Acoustics, vol. 36, no. 3–4, pp. 171–218, 1992,
- [7] D. R. Begault and L. J. Trejo, "3-D Sound for Virtual Reality and Multimedia", Tech. Rep. NASA/TM-2000-209606, Aug. 2000.
- [8] M. Gröhn, "Application of spatial sound reproduction in virtual environments experiments in localization, navigation, and orientation", PhD thesis, Citeseer, 2006.
- [9] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound", in 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Sep. 2014, pp. 53–61.
- [10] C. H. Larsen, D. S. Lauritsen, J. J. Larsen, M. Pilgaard, and J. B. Madsen, "Differences in Human Audio Localization Performance Between a HRTF- and a non-



HRTF Audio System", in *Proceedings of the 8th Audio Mostly Conference*, ser. AM '13, Piteå, Sweden: ACM, 2013, 5:1–5:8,

- [11] R. D. Shilling and B. Shinn-Cunningham, "Virtual auditory displays", Handbook of virtual environment technology, pp. 65–92, 2002.
- [12] A. W. Mills, "On the Minimum Audible Angle", The Journal of the Acoustical Society of America, vol. 30, no. 4, pp. 237–246, Apr. 1958,
- [13] C. H. Larsen, D. S. Lauritsen, J. J. Larsen, M. Pilgaard, J. B. Madsen, and R. Stenholt, "Aurally Aided Visual Search Performance Comparing Virtual Audio Systems", English, Audio Engineering Society, Oct. 2014.
- [14] G. D. Romigh and B. D. Simpson, "Do you hear where I hear?: isolating the individualized sound localization cues", *Frontiers in Neuroscience*, vol. 8, Dec. 2014,
- [15] J. A. Veltman, A. B. Oving, and A. W. Bronkhorst, "3-D Audio in the Fighter Cockpit Improves Task Performance", *The International Journal of Aviation Psychology*, vol. 14, no. 3, pp. 239–256, Jun. 2004,
- [16] P. Minnaar, S. K. Olesen, F. Christensen, and H. Moller, "The importance of head movements for binaural room synthesis", 2001.
- [17] Y. Iwaya, Y. Suzuki, and S. Takane, "Effects of listener's head movement on the accuracy of sound localization in virtual environment", 2004, pp. 26–29.
- [18] T. V. Wilson, Sound Waves and Cues. URL: http://electronics.howstuffworks. com/virtual-surround-sound.htm (visited on 05/25/2015).
- [19] G. H. de Sousa and M. Queiroz, "Two approaches for HRTF interpolation", 2009.
- [20] D. Hammershøi and H. Møller, "Methods for Binaural Recording and Reproduction", Acta Acustica united with Acustica, vol. 88, no. 3, pp. 303–311, May 2002.
- [21] H. Ziegelwanger, A. Reichinger, and P. Majdak, "Calculation of listener-specific head-related transfer functions: Effect of mesh quality", *Proceedings of Meetings* on Acoustics, vol. 19, no. 1, p. 050017, Jun. 2013,
- [22] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Using a Typical Human Subject for Binaural Recording", English, Audio Engineering Society, May 1996.
- [23] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Efficient HRTF Interpolation in 3d Moving Sound", English, Audio Engineering Society, Jun. 2002.
- [24] G. Enzner, C. Antweiler, and S. Spors, "Trends in Acquisition of Individual Head-Related Transfer Functions", en, in *The Technology of Binaural Listening*, ser.



Modern Acoustics and Signal Processing, J. Blauert, Ed., Springer Berlin Heidelberg, 2013, pp. 57–92,

- [25] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity", *The Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2202–2215, Oct. 2006,
- [26] M. Queiroz and G. H. M. d. Sousa, "Efficient Binaural Rendering of Moving Sound Sources Using HRTF Interpolation", *Journal of New Music Research*, vol. 40, no. 3, pp. 239–252, Sep. 2011,
- [27] E Choueiri, "Optimal crosstalk cancellation for binaural audio with two loudspeakers", *Princeton University*, p. 28, 2008.
- [28] R. Elen, "Ambisonics: The surround alternative", 2001.
- [29] D. G. Malham and A. Myatt, "3-D Sound Spatialization using Ambisonic Techniques", *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, Dec. 1995,
- [30] C. Mendonça, J. A. Santos, G. Campos, P. Dias, and J. Vieira, "On the Adaptation to Non-Individualized HRTF Auralizations: A Longitudinal Study", English, Audio Engineering Society, Mar. 2012.
- [31] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. A. Santos, "On the Improvement of Localization Accuracy with Non-Individualized HRTF-Based Sounds", English, *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 821–830, Nov. 2012.
- [32] C. Mendonça, G. Campos, P. Dias, and J. A. Santos, "Learning Auditory Space: Generalization and Long-Term Effects", *PLoS ONE*, vol. 8, no. 10, Oct. 2013,
- [33] AM3D, Zirene 3d. URL: http://www.am3d.dk/home-english/products/zirene% C2%AE3d.aspx (visited on 05/26/2015).
- [34] Unity Technologies, Unity Game Engine. URL: http://unity3d.com/ (visited on 05/20/2015).
- [35] B. K. Wiederhold, R. Davis, and M. D. Wiederhold, "The effects of immersiveness on physiology", Studies in health technology and informatics, pp. 52–62, 1998,
- [36] P. Toprac and A. Abdel-Meguid, "Causing Fear, Suspense, and Anxiety Using Sound Design in Computer Games", en, in *Game Sound Technology and Player Interaction: Concepts and Developments*, M. Grimshaw, Ed., IGI Global, Sep. 2010, pp. 176-191,
- [37] E. Brown and P. Cairns, "A Grounded Investigation of Game Immersion", in CHI '04 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA '04, New York, NY, USA: ACM, 2004, pp. 1297–1300,



- [38] B. Perron, "A Cognitive Psychological Approach to Gameplay Emotions", May 2005.
- [39] I. Ekman, "Psychologically Motivated Techniques for Emotional Sound in Computer Games", Proc. AudioMostly 2008, pp. 20–26, Oct. 2008.
- [40] C. Bateman, *Imaginary Games*, en. John Hunt Publishing, Nov. 2011,
- [41] S. Schneider, Paradox of Fiction. URL: http://www.iep.utm.edu/fict-par/ (visited on 05/22/2015).
- [42] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen, "Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation", *Media Psychology*, vol. 8, no. 4, pp. 343–367, Nov. 2006,
- [43] J. A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980,
- [44] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies", *International Journal* of Human-Computer Studies, vol. 65, no. 4, pp. 329–347, 2007.
- [45] R. L. Hazlett, "Measuring Emotional Valence During Interactive Experiences: Boys at Video Game Play", in *Proceedings of the SIGCHI Conference on Human Factors* in Computing Systems, ser. CHI '06, New York, NY, USA: ACM, 2006, pp. 1023– 1026,
- [46] J. Ellerm, Clickbait, Virality And Why It Matters To Fintech, Mar. 2015. URL: http://jessicaellerm.com/2015/03/31/clickbait-virality-and-why-itmatters-to-fintech/ (visited on 05/19/2015).
- [47] I. Ekman and R. Kajastila, "Localization Cues Affect Emotional Judgments Results from a User Study on Scary Sound", English, Audio Engineering Society, Feb. 2009.
- [48] T. Garner and M. Grimshaw, "A Climate of Fear: Considerations for Designing a Virtual Acoustic Ecology of Fear", in *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, ser. AM '11, New York, NY, USA: ACM, 2011, pp. 31–38,
- [49] R. L. Mandryk and K. M. Inkpen, "Physiological Indicators for the Evaluation of Co-located Collaborative Play", in *Proceedings of the 2004 ACM Conference* on Computer Supported Cooperative Work, ser. CSCW '04, New York, NY, USA: ACM, 2004, pp. 102–111,
- [50] R. L. Mandryk, "Physiological Measures for Game Evaluation", en, in *Game Us-ability: Advice from the Experts for Advancing the Player Experience*, K. Isbister and N. Schaffer, Eds., CRC Press, Aug. 2008, pp. 207–235,



- [51] S. Raaijmakers, F. Steel, M. de Goede, N. van Wouwe, J. Van Erp, and A.-M. Brouwer, "Heart Rate Variability and Skin Conductance Biofeedback: A Triple-Blind Randomized Controlled Study", in 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), Sep. 2013, pp. 289–293.
- [52] T. Garner, "Identifying Habitual Statistical Features of EEG in Response to Fearrelated Stimuli in an Audio-only Computer Video Game", in *Proceedings of the 8th Audio Mostly Conference*, ser. AM '13, New York, NY, USA: ACM, 2013, 14:1– 14:6,
- [53] L. E. Nacke, S. Stellmach, and C. A. Lindley, "Electroencephalographic Assessment of Player Experience A Pilot Study in Affective Ludology", en, Simulation & Gaming, vol. 42, no. 5, pp. 632–655, Oct. 2011,
- [54] T. A. Garner and M. Grimshaw, "Psychophysiological Assessment Of Fear Experience In Response To Sound During Computer Video Gameplay", English, Jul. 2013.
- [55] I. M. Marks, *Fears and Phobias*, en. Academic Press, 1969.
- [56] J. M. Kivikangas, G. Chanel, B. Cowley, I. Ekman, M. Salminen, S. Järvelä, and N. Ravaja, "A review of the use of psychophysiological methods in game research", *Journal of Gaming & Virtual Worlds*, vol. 3, no. 3, pp. 181–199, Sep. 2011.
- [57] G. Park, M. W. Vasey, J. J. Van Bavel, and J. F. Thayer, "When tonic cardiac vagal tone predicts changes in phasic vagal tone: the role of fear and perceptual load", eng, *Psychophysiology*, vol. 51, no. 5, pp. 419–426, May 2014,
- [58] K. C. Bilchick and R. D. Berger, "Heart Rate Variability", en, Journal of Cardiovascular Electrophysiology, vol. 17, no. 6, pp. 691–694, Jun. 2006,
- [59] W. Waterink and A. van Boxtel, "Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task", *Biological Psychology*, vol. 37, no. 3, pp. 183–198, Jul. 1994,
- [60] W. Ijsselsteijn, K Poels, and Y. de Kort, "Measuring player experiences in digital games. Development of the Game Experience Questionnaire (GEQ)", Manuscript in preparation, 2008.
- [61] M. Salminen and N. Ravaja, "Oscillatory brain responses evoked by video game events: the case of super monkey ball 2", eng, Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, vol. 10, no. 3, pp. 330–338, Jun. 2007,
- [62] G. Calleja, "Immersion in Virtual Worlds", in *The Oxford Handbook of Virtuality*, M. Grimshaw, Ed., Oxford University Press, 2013, pp. 222–236.



- [63] P. Cairns, A. Cox, and A. I. Nordin, "Immersion in Digital Games: Review of Gaming Experience Research", en, in *Handbook of Digital Games*, r. C. Angelides and H. Agius, Eds., John Wiley & Sons, Inc., 2014, pp. 337–361,
- [64] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire", *Presence: Teleoperators & Virtual Environments*, vol. 7, no. 3, pp. 225–240, Jun. 1998,
- [65] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games", *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, Sep. 2008,
- [66] M. Csikszentmihalyi, Flow: The Psychology of Optimal Experience, en. New York, NY, USA: Harper & Row, 1990,
- [67] G. Calleja, In-Game: From Immersion to Incorporation, English, 1 edition. Cambridge, Mass: The MIT Press, May 2011,
- [68] L. Ermi and F. Mäyrä, "Fundamental components of the gameplay experience: Analysing immersion", Worlds in play: International perspectives on digital games research, vol. 37, 2005.
- [69] C. I. Jennett, "Is game immersion just another form of selective attention? An empirical investigation of real world dissociation in computer game immersion", eng, Doctoral, UCL (University College London), Jul. 2010.
- [70] K. Poels, Y. de Kort, and W. IJsselsteijn, "Identification and categorization of digital game experiences: a qualitative study integrating theoretical insights and player perspectives", Westminster Papers in Communication and Culture, vol. 9, no. 1, pp. 107–129, 2012.
- [71] M. Slater, V. Linakis, M. Usoh, R. Kooper, and G Street, "Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess", 1996, pp. 163–172.
- [72] M. Grimshaw, The Oxford Handbook of Virtuality, en. Oxford University Press, Dec. 2013,
- [73] M. Grimshaw, C. Lindley, and L. Nacke, "Sound and immersion in the first-person shooter: mixed measurement of the player's sonic experience", in *Proceedings of Audio Mostly Conference*, 2008.
- [74] L. Nacke and C. A. Lindley, "Flow and Immersion in First-person Shooters: Measuring the Player's Gameplay Experience", in *Proceedings of the 2008 Conference* on Future Play: Research, Play, Share, ser. Future Play '08, New York, NY, USA: ACM, 2008, pp. 81–88,



- [75] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments", en, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, Dec. 2009,
- [76] B Witmer, C Jerome, and M Singer, "The Factor Structure of the Presence Questionnaire", *Presence*, vol. 14, no. 3, pp. 298–312, Jun. 2005,
- [77] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using Presence Questionnaires in Reality", Presence: Teleoperators & Virtual Environments, vol. 9, no. 5, pp. 497– 503, Oct. 2000,
- [78] M. Slater, "Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire", Presence: Teleoperators & Virtual Environments, vol. 8, no. 5, pp. 560–565, Oct. 1999,
- [79] M. Slater, B. Spanlang, and D. Corominas, "Simulating Virtual Environments Within Virtual Environments As the Basis for a Psychophysics of Presence", in ACM SIGGRAPH 2010 Papers, ser. SIGGRAPH '10, New York, NY, USA: ACM, 2010, 92:1–92:9,
- [80] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments", English, in *Chalmers Publication Library (CPL)*, 2004.
- [81] B. K. Wiederhold, D. Jang, M. Kaneda, I. Cabral, Y. Lurie, T May, I. Kim, M. D. Wiederhold, and S. Kim, "An investigation into physiological responses in virtual environments: an objective measurement of presence", *Towards cyber psychology: mind, cognitions and society in the Internet age. IOS Press, Amsterdam*, pp. 175–183, 2001.
- [82] B. Insko, "Measuring Presence: Subjective, Behavioral and Physiological Methods", English, in *Being There: Concepts, effects and measurement of user presence in synthetic environments*, ser. Studies in new technologies and practices in communication. G Riva, F Davide, and W. IJsselsteijn, Eds., Amsterdam, Netherlands: IOS Press, 2003, pp. 109–119,
- [83] F. Steinicke and G. Bruder, "A Self-experimentation Report About Long-term Use of Fully-immersive Technology", in *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction*, ser. SUI '14, New York, NY, USA: ACM, 2014, pp. 66–69,
- [84] Oculus, Best Practice Guide. URL: http://static.oculus.com/sdk-downloads/ documents/Oculus_Best_Practices_Guide.pdf (visited on 05/25/2015).
- [85] Whodat, Home Interior Pack. URL: https://www.assetstore.unity3d.com/en/ #!/content/12382 (visited on 05/20/2015).
- [86] Two Big Ears, 3dception. URL: http://twobigears.com/ (visited on 05/20/2015).



- [87] UQO Cyberpsychology Lab, Presence Questionaire, 2004.
- [88] N. Lavie, "Distracted and confused?: selective attention under load", eng, Trends in Cognitive Sciences, vol. 9, no. 2, pp. 75–82, Feb. 2005,
- [89] Doc-Ok.org, Fighting black smear, Blog, Oct. 2014. URL: http://doc-ok.org/?p= 1082 (visited on 05/26/2015).
- [90] Epic Games, Unreal Engine. URL: https://www.unrealengine.com/ (visited on 05/20/2015).
- [91] Crytek, CryEngine. URL: http://cryengine.com/ (visited on 05/20/2015).
- [92] Audiokinetic, Wwise. URL: https://www.audiokinetic.com/products/wwise/ (visited on 05/20/2015).
- [93] Firelight Technologies, FMOD. URL: http://www.fmod.org/ (visited on 05/20/2015).
- [94] Interactive Audio Special Interest Group, Interactive 3d Audio Rendering Guidelines, Sep. 1999.



Appendices



A. Pilot Testing

In order to optimize the experimental design, we performed additional pilot tests with the goal to eliminate design flaws and incorrect data collection. Because we had only limited experience with measuring physiological data, we found it difficult to foresee how the resulting data would present itself. It was therefore important for us to perform these pilot tests, which included fewer participants, allowing us to make more iterations on our experiment design. Because the pilot tests resulted in a low number of data samples, nothing conclusive could be determined. Therefore the pilot tests were primarily used for getting a better understanding on the behavior of questionnaire responses and the nature of EDA and HR. This means, even though we did not see any apparent difference in a pilot test it was not argument to leave it out for the final experiment. The following pilot tests appears similar to the final experiment, so the reader should assume that the pilot test is similar to the final experiment unless stated otherwise.

During the development and testing of the environment we encountered a number of problems of significance, which are further describer in Appendix B.

A.1 First Pilot Test

In the early stages of the experimental design, the experiment would include a horror game. The participant's goal was still to sequentially collect orbs, similar to the experiment described in Chapter 4. The participant were standing while wearing a HMD. At certain fixed events, a disfigured head would appear close to an orb, forcing the player to encounter it at some point. When the head entered the player's FOV the head would rapidly charge towards the player. If the head reached the player a jump scare effect would be executed and a loud scream would be played. See Figure A.1 for a screenshot that exemplifies this.

When the head charged towards the player, he could look away from the head until the head left the player's FOV which caused it to disappear. The goal for the player was to collect as many orbs without one of the disfigured faces reaching the player. Before the pilot test was conducted, we believed that the horror factor would only complicate the pilot test by making it difficult to gather participants due to the horrific nature of the game. Additionally, we believed that the EDA signal would be difficult to interpret,





Figure A.1: This figure illustrates the pop-up scare in the first pilot test design. These visuals were followed by a horrific scream.

because elements such as anticipation and fear caused by non-auditory events would be mixed into the signal.

Therefore, a time element was implemented into the game instead. Each orb would have an internal timer. If the timer went out, the orb would no longer be available. The player's task was now to collect as many orbs as possible. For each 6th orb, the internal timer would be reduced with half a second, causing the game to be more difficult as the player progressed. Whenever an orb spawned, a distinct short auditory cue was played at its position. The orbs could spawn all around the player.

In each game session, three blue bonus orbs would spawn, which emitted an auditory cue that lasted the orbs entire life-span. The bonus points would have a fixed life-span of 5 seconds. The participants were told that a bonus orb would score points corresponding to five ordinary orbs. The bonus orbs could co-exist with a regular orb, causing the player to choose between the bonus orb and the regular orb. It was possible for the participant to collect both orbs, if fast enough.

Whenever an orb disappeared, eight particles would spawn and bounce away from the orb's position. Upon impacting with anything in the environment, the particles would emit a sound and continue bouncing. After 2.5 seconds, the particles would disappear. Each particle would have a colored trail following it. The trail's color was green, if the player successfully collected the orb, otherwise red. See Figure A.2 for an example of this.

For diminishing the effect of training, a sequence of 10 orbs and a bonus orb, was presented before each game session. This training condition would include audio rendered using mono. The implementation of mono is explained in Appendix C.2.





Figure A.2: A screenshot of an orb disappearing. Eight particles spawned from the position of the orb. The color of the trail depends on whether the orb was successfully collected (green) or not (red).

A.1.1 Observations

Six participants were used for this pilot test. The experiment took around to 45 minutes. The questionnaires had a tendency to take longer time to fill out than completing the two game conditions. Most participants voiced that motion sickness was not an issue. When asked if they noticed any difference between the two game conditions, they voiced a noticeable difference in auditory playback. A single participant voiced that he clearly found it easier to locate orbs in the first (for him 3D audio) condition.

Based on our own personal observations during the pilot test, it appeared that EDA increased when the participant had to look behind themselves. We therefore performed a couple of internal tests to investigate this and it appeared to have an effect.

Additionally, we observed that anticipation played a role on EDA. Because each orb had an internal timer, the participants knew the longer they took in finding a given target, the closer they were to failing.

During the experiment the participants were exposed to 140 auditory cues that an orb had spawned. We discussed whether the repeated exposure to the same stimuli could diminish the phasic EDA.

The questionnaires indicated no significant difference. The presence questionnaires regarding sound were neither significant. We started asking the questions: Were the spatial properties not apparent enough? Were the spatial properties inaudible due to the sounds being non-continuous? Or were the sample size simply too small to see a difference?

At this state it was difficult for us to state anything definitive about our results, and we



Figure A.3: An illustration of the EDA over time. The blue lines indicate onsets of audio feedback from the orbs.

decided to continue with another pilot test.

A.2 Second Pilot Test

We changed the sound rendering method for the second pilot test to be mono and stereo for the game conditions, and no sound for the training condition. If we were not able to show a difference between the mono and stereo rendering method, then we assumed that it would not be possible to find a difference between stereo and 3D audio either.

The orb's internal timer was removed. This was to eliminate the participant's arousal based on anticipation. When removing this feature, it became questionable whether our pilot test was still considered a game. This pilot test was conducted using screen and mouse to eliminate any noise which were related to body movement using the HMD.

A.2.1 Observations

Six participants participated in the second pilot test. The pilot test took around 45 minutes for each participant. EDA looked similar to what we saw in the first pilot test. When observing the EDA event responses, there was no significant difference, and no clear pattern were observed. There seemed to be no correlation between an auditory event and corresponding EDA events, see Figure A.3. We believe this was due to diminished EDA events upon continuous exposure of the same stimuli.

When asked whether they noticed any difference between the game conditions, they voiced that it was something related to audio. They stated there was a difference in the localization difficulty of the orbs. This observation was confirmed through investigation of the acquisition time, which showed a tendency for stereo having lower acquisition times than mono. When analyzing the questionnaire responses a significant difference was not found in the total immersion score or in the presence score. However, in the presence questions related to sound a significant difference was found favoring stereo.



We realized that the amount of auditory cues created too much noise and therefore we decided to conduct a third pilot test.

A.3 Third Pilot Test

In the third pilot test we continued comparing mono audio with stereo. Instead of investigating a change in EDA upon orb spawns as auditory events, we decided to only expose them to a few unexpected auditory events.

In the third pilot test, the participants were told that they had to sequentially located orbs around the room. They were also told that each orb had different internal timers, and if that timer ran out, "something" would happen. They were further told that their goal was to collect as many orbs as possible. However, the information they received about the orb's internal timer was deliberately false. In reality, the orbs had no timers and the participants could actually use as much time as they wanted. However, during specific events in the game, certain orbs would not be visible to the player. After 10 seconds a scary sound would be played as an indication that the player had failed. The sound consisted of three audio sources orbiting the player's avatar.

With this design, we attempted to create the illusion that this failure was caused by the player's own inability to localize the orb. This design was to ensure that each participant encountered the same amount of events, at the same times. If we implemented actual timers into each orb it would vary between participants how many failures they would experience, as the results would then become skill dependent.

The particle effect which occurred upon successfully collecting an orb was removed. In order to save time, we decided to skip the questionnaires for this pilot.

A.3.1 Observations

Twelve participants participated in this third pilot test. The pilot test took between 15-20 minutes for each participant. All participants were informed that the pilot test would include scary sounds. The EDA events from this pilot test, was different and we could observe a pattern different from the previous pilot test observations. We could see a correlation between a scary event and phasic EDA events which differed in amplitude from other EDA events. The nature of EDA responses between participants were found different. Some participants had a clear response while other participants were more ambiguous. See Figure A.4 for examples on the differences in EDA.

We observed that the first scary event had the largest response. We believe this was caused by participants being unaware of what would happen if they failed, as they were not introduced to the event prior to pilot test.

After the pilot test we asked the participants whether they noticed any difference between the two game conditions. Most did not notice any difference between the two sessions. We





Figure A.4: An illustration of the EDA over time for three participants. The blue lines indicate onsets of auditory events.


believed this were due to too few auditory cues, as only the scary events and background noise in the form of rain were present. We also asked the participants if they realized that we deceived them. More than half of the participants figured out that they were deceived. They voiced that if they could not find the target within a short period of 3-4 seconds, they knew that a scare would come regardless of what they did.



B. Problems Encountered

During the implementation and execution of the experiment we encountered a few problems which are presented here.

B.1 Shimmer3 and Unity

The device that we used for measuring physiological data, Shimmer3 connecting with Bluetooth, comes with a number of tools for using the device. One of these tools is a program, ShimmerConnect 0.2.0, which is used to connect and stream the data from the Shimmer device into a computer. This program happens to be written in C# which is also the primary language that we used for developing the experiment in Unity, and source files for this program has been made available by Shimmer. We therefore thought that we could bring this implementation into Unity, and directly log the data from there, which would bring some benefits for later use of the data. First being that synchronization of time between all the different data streams and secondly we only had to run a single program instead of having multiple programs running simultaneously.

Our first concern about the ShimmerConnect program was that the target framework is .NET 4.0 and Unity only runs a Mono (Not be confused with mono audio) implementation of .NET 3.5. We found that by trimming away the UI of the application, we could just switch target framework to 3.5 and everything was still working. With this we began building the application as a plug-in for Unity compiled into a DLL library file. Unity was then used to execute the plug-in and signed up for a callback which should provide the necessary data from the device. This however did not seem to work as expected, because the connection to the Shimmer device was unstable. We believed that it might have to do with it was a plug-in and not "written in Unity". The reason why we had it running as a plug-in was because a number of libraries that the Shimmer program used did not appear to be available through Unity, where one was the System.IO.Ports. Later we found that by default Unity only has activated a subset of the .NET functionality and by changing from '.NET subset' to '.NET' these missing libraries became available.

Our second approach was therefore to directly have the source files compiled by Unity, by simple having the Shimmer program's source code directly integrated into Unity. We soon realized that this caused the same kind of problems, where we could connect to the Shimmer device periodically. By analyzing the different responses from the original



program and the Unity implementation we found that the device did not respond correctly to the different requests made by the Shimmer program. Often in Unity the response from the device was *null*, whereas the original program did not receive any null values. We found a couple of values that the device responded with and hard coded them into the Unity implementation. This did at a point create a stabilization in connecting with the device, though streaming from the device was never achieved properly.

The best guess to why this implementation of the connection with the Shimmer 3 device did not work from within Unity, is that there are some differences between pure .NET and the Mono implementation of .NET that Unity utilizes. Whether this is related to speed or stability of the Mono implementation we do not know, but we decided to stop trying to make it work due to time consumption. Instead, we ran the softwares on two different computers, synchronizing the streams by initiating each program simultaneously with two button pushes. This procedure may induce a minor offset between streams, but we believe this effect would be minuscule and insignificant.

B.2 Black Smearing

One observation that we did in our experiment, was that when using the Oculus Rift we saw a visual artifact occurring. This artifact established itself when looking at a dark object with a light background. Moving the head led to a trail of black pixels following the dark object. We found that this is a phenomenon called *black smearing*. This phenomenon can occur because of two factors. First is the vestibulo-ocular reflex (VOR). The VOR is a reflex that causes our eyes to move in the opposite direction of the head movement which helps our visual image to stay stabilized, this reflex is caused by the vestibular system. Secondly is the delay that occurs for LED screens to turn on and off pixels which, unfortunately, is not done in constant time. When we have the screen just in front of our eyes, as is the case with a HMD, this effect is visible because we do not have our eyes fixed on the middle of the screen, but rather on a visual object that we can see in the scene. Therefore, when we move our head our eyes move in opposite direction, caused by VOR, focusing on a new set of pixel, which start to turn off or on. Because of this, we see the time it takes for the pixels to turn on or off, hence it creates a black smear. This phenomenon is only occurring because the screen is so close to our eyes. When looking at a normal television the same effect is not perceivable. The effect can also occur for bright objects on dark background leaving a white trail. [89]

The effect of black smearing can be reduced by sacrificing some contrast in the image. Because the effect is mostly occurring around total black or total white pixels, these pixel values can be reduced to the extent that black pixels are not totally black but rather a very dark gray. This causes the display to never turn off the pixels minimizing the transition time. We did not however reduce the effect, as we only discovered the problem late in the process, and at that time we had gathered almost all data for the final experiment, so to use this method to compensate seemed irrelevant at that point. However, we do not believe this had any significant influence on our results.



C. Additional Information

This chapter contains additional information, which had little value in the discussion of the study.

C.1 Tools for Developers

As a game developer it is often necessary to rely on third party software for your main tool (game engine). When it comes to 3D audio rendering, most developers rely on a third party tool for this, because it is not a simple task to recreate this technology. In the past 3D audio rendering was achieved using specialized hardware, and this complicated things for developers. It was unknown whether their users would have access to the required hardware. Therefore, the standard OpenAL was introduced as an audio API and was adopted by the different audio card manufactures. If a game used OpenAL, the developer could pass the audio data through the OpenAL API and if a user had obtained a sound card, such as a Turtle Beach Santa Cruz or a Sound Blaster X-Fi MB series, which supported this API, the user could then make use of 3D audio. As an alternative to Open AL, Windows also introduced a 3D audio API in DirectSound3D, however this API were discontinued since the release of Windows Vista.

The reason for 3D audio previously had to be rendered on specialized hardware was due to its heavy demand off resources. A rapid increase in available resources for computers has occurred since the 3D sound technology was introduced. While the hardware capabilities have increased, the 3D audio technology has not grown in terms of computational requirements. This has led to the possibility of rendering the audio directly on the computer's CPU and therefore be independent of any hardware solutions.

Recently a number of 3D audio software solutions have become available for game developers. A number of them support some of the most used game engines such as Unity [34], Unreal Engine[90] and CryEngine [91]. Most game engines do also use an audio engine middleware such as Wwise [92] or FMOD [93]. We have compiled a list of current available solutions for 3D audio in games to the best of our knowledge which can be seen in Table C.1.



Table C.1: A list of currently known 3D audio software providers and which platforms or commonly used engines they support.

Engine name	Company	Supported engines	Platforms
3Dception	Two Big Ears	Unity, Wwise	Win, OS X, Linux, Android, iOS
AstoundSound	GenAudio	Unity, Unreal, FMOD, Wwise,	Win, OS X, Android, iOS, Xbox One
Auro-3D Headphones	Auro Technologies		
PapaEngine	Somethin' Else		iOS
Phonon 3D	Impulsonic	Unity, Unreal, FMOD, Wwise	Win, OS X, Android, iOS, Xbox One, PS 4
Premium Sound 3D	SRS		Win
QSurround	QSoundLabs		Android, iOS
Real Space 3D Audio	VisiSonics	Unity, FMOD, Wwise	Win, OS X, Android, iOS

C.2 Rendering of Mono Audio

In some of our pilot experiments we decided to test mono sound against stereo sound. One thing about mono sound is that it is independent of position relative to the listener, though in our situation it was necessary to have distance attenuation included. This was for the sake of consistency between comparing the different systems.

To utilize monaural rendering in Unity we had two approaches. Because 3Dception did not include an option for mono rendering, our first approach was to exchange all 3Dception components with regular Unity sound source components, which offers monaural rendering. This approach was later changed, and instead we changed the settings of Unity to render sound monaurally through the audio settings, which resulted in all sounds being played with mono rendering. Because none of the described methods of mono rendering includes distance attenuation, we implemented our own. This implementation followed the IASIG I3DL2 attenuation model [94] (Also at times called Inverse Distance Clamped Model), which to our knowledge is a commonly implemented distance attenuation model and is close to how sound is attenuated in the real world. For mono audio this was simply achieved by attenuating the volume of the sound source, see Equation C.1.

 $\frac{minimumDistance}{minimumDistance + rollOffFactor * (distance - minimumDistance)}$ (C.1)

C.3 Consent Form Description

Before each participant participated in our experiment, they all had to fill a consent form. The consent form's primary function was to make sure we were allowed to collect and use their data. The consent form also included questions such as age, sex, sight and hearing.

Often in auditory localization studies, it is considered acceptable to ask the participants whether they have normal hearing however the optimal solution would be to actually perform a preliminary test for each participant which includes a standardized hearing



test. But due to the already long experiment of 40-50 minutes we presented them this consent form. On the next page, the consent form can be seen.



Consent form

When signing the document, the signee gives group 151031 the rights to both use and publish the data collected by the signee during the experiment. The data can be of the form of digital metrics, physiological measurements, questionnaires, video and images.

day month year

signature

please fill the information below

age



_E FEMALE

To my best believe, I report to have normal sight	YES	NO
To my best believe, I report to have normal hearing	YES	NO
Have you ever tried a head mounted display before?	YES	NO
Do you easily experience motion sickness?	YES	NO



C.4 Sounds

In the following section a collection of figures presents the frequency spectrums for the sounds used in the experiment. There were three different event sounds (Figure C.1, C.2 and C.3) and four different environmental sounds (Figure C.4, C.5 and C.6), though one of the environmental sounds is not presented below, due its short duration could not be frequency analyzed. The frequency spectrums are generated with Audacity 2.0.5 with a resolution of 16384 bins. The sounds were selected for making use of a broadband spectrum.



Figure C.1: Scary event 1.





Figure C.2: Scary event 2.



Figure C.3: Scary event 3.



Figure C.4: Radio environmental sound.



Figure C.5: Rain environmental sound.

AALBORG UNIVERSITY





Figure C.6: Clock environmental sound.



C.5 Participant Data

Here are all the summarized SR event values for each participant and for their corresponding game condition.



Participant ID	$\mathbf{SoundSystem}$	SR event (microSiemens)
11	3D audio	1.04
11	Stereo	0.76
12	3D audio	0.09
12	Stereo	0.15
13	3D audio	0.9
13	Stereo	0.43
14	3D audio	0.11
14	Stereo	0.17
16	3D audio	0.32
16	Stereo	0.37
17	3D audio	0.83
17	Stereo	0.35
20	3D audio	0.1
20	Stereo	0.09
21	3D audio	0.04
21	Stereo	0.14
22	3D audio	0.79
22	Stereo	0.35
23	3D audio	0.39
23	Stereo	0.19
25	3D audio	1.18
25	Stereo	0.88
27	3D audio	1.04
27	Stereo	0.23
28	3D audio	0.53
28	Stereo	0.71
29	3D audio	0.47
29	Stereo	0.34
30	3D audio	0.15
30	Stereo	0.06
32	3D audio	0.22
32	Stereo	0.42
33	3D audio	0.25
33	Stereo	0.12
34	3D audio	0.34
34	Stereo	0.24
35	3D audio	0.36
35	Stereo	0.19
36	3D audio	0.25
36	Stereo	0.12



C.6 Participants' Immersion Questionnaire Scores

20 participants immersion questionnaire scores. Score and sub-score names are abbreviated as the following:

- PID: Participant ID
- TI: Total Immersion
- C: Challenge
- CI: Cognitive Involvement
- D: Dissociation
- I: Involvment
- CT: Control
- SEI: Self-evaluated Immersion



PID	Condition	TI	\mathbf{C}	CI	D	Ι	\mathbf{CT}	SEI
0	3D audio	67	16	33	7	35	21	6
0	Stereo	62	16	31	5	34	19	6
1	Stereo	77	16	44	1	39	27	8
1	3D audio	73	16	43	1	37	25	9
2	3D audio	79	20	44	4	45	29	6
2	Stereo	74	15	40	1	43	26	6
3	Stereo	75	14	40	6	41	26	4
3	3D audio	79	15	40	8	40	30	3
4	3D audio	69	16	36	6	33	29	6
4	Stereo	56	11	28	6	30	23	5
5	Stereo	75	15	39	1	43	24	7
5	3D audio	71	15	37	0	44	26	8
6	3D audio	64	13	34	0	38	22	7
6	Stereo	71	14	35	6	37	24	8
7	Stereo	61	14	32	0	41	27	8
7	3D audio	44	13	29	-6	35	25	7
8	3D audio	81	16	40	2	44	27	7
8	Stereo	97	16	44	6	50	33	8
9	Stereo	70	15	42	2	37	28	6
9	3D audio	64	13	36	1	33	24	5
11	Stereo	80	14	38	6	41	24	8
11	3D audio	70	15	37	-2	43	26	6
12	3D audio	74	14	41	3	42	26	6
12	Stereo	65	14	37	1	40	23	7
13	Stereo	54	14	33	1	27	22	7
13	3D audio	57	12	34	1	26	23	6



PID	Condition	\mathbf{TI}	\mathbf{C}	CI	D	Ι	\mathbf{CT}	SEI
14	3D audio	77	14	43	6	38	23	7
14	Stereo	71	14	37	6	35	25	7
16	3D audio	52	18	27	-3	35	25	6
16	Stereo	17	9	12	-10	17	22	5
17	Stereo	65	14	38	-1	36	23	5
17	3D audio	55	12	33	-3	32	21	5
18	3D audio	69	16	34	2	38	22	7
18	Stereo	37	13	23	4	20	16	4
19	Stereo	68	11	36	3	33	27	7
19	3D audio	71	11	36	6	32	26	7



C.7 Participants' Presence Questionnaire Scores

34 participants presence questionnaire scores. Score and sub-score names are abbreviated as the following:

- PID: Participant ID
- TPS: Total Presence Score
- R: Realism
- PoA: Possibility to Act
- QoI: Quality of Interface
- PoE: Possibility to Examine
- SEoP: Self Evaluation of Performance
- S: Sound



PID	Condition	TPS	R	PtA	QoI	PoE	SEoP	S
0	3D audio	80	43	17	-5	13	12	12
0	Stereo	87	43	20	-5	16	13	12
1	Stereo	69	28	19	-5	15	12	15
1	3D audio	78	34	22	-6	15	13	11
2	3D audio	64	28	20	-8	14	10	14
2	Stereo	53	26	19	-12	11	9	18
3	Stereo	74	37	21	-9	13	12	16
3	3D audio	84	37	24	-5	16	12	18
4	3D audio	57	29	15	-13	15	11	13
4	Stereo	69	36	17	-11	15	12	16
5	Stereo	76	39	23	-15	16	13	16
5	3D audio	73	40	21	-15	16	11	18
6	3D audio	90	42	25	-6	16	13	17
6	Stereo	90	41	23	-3	16	13	17
7	Stereo	92	40	28	-5	15	14	21
7	3D audio	87	45	25	-11	14	14	19
8	3D audio	88	44	27	-12	19	10	20
8	Stereo	101	46	24	-4	21	14	20
9	Stereo	75	37	16	-9	20	11	18
9	3D audio	73	36	21	-11	16	11	14
11	Stereo	91	45	24	-6	15	13	9
11	3D audio	78	41	20	-9	14	12	10
12	3D audio	68	34	18	-6	12	10	14
12	Stereo	65	32	19	-10	11	13	14
13	Stereo	84	42	24	-5	12	11	18
13	3D audio	85	41	24	-6	12	14	15
14	3D audio	88	43	27	-9	13	14	15
14	Stereo	82	44	21	-9	12	14	13



PID	Condition	TPS	R	PtA	QoI	PoE	SEoP	S
16	3D audio	78	34	22	-6	16	12	13
16	Stereo	58	17	25	-5	12	9	10
17	Stereo	76	36	23	-7	14	10	12
17	3D audio	84	40	23	-6	14	13	14
18	3D audio	80	34	25	-5	15	11	16
18	Stereo	71	26	25	-6	15	11	10
19	Stereo	77	39	24	-9	12	11	9
19	3D audio	64	31	23	-11	10	11	11
20	3D audio	67	29	23	-5	10	10	13
20	Stereo	74	31	24	-6	12	13	16
21	Stereo	85	39	21	-5	19	11	19
21	3D audio	91	42	22	-5	19	13	19
22	3D audio	80	39	21	-8	16	12	15
22	Stereo	86	40	22	-6	17	13	17
23	Stereo	76	39	19	-8	14	12	15
23	3D audio	80	41	23	-10	16	10	17
24	3D audio	88	45	27	-10	12	14	14
24	Stereo	88	44	27	-9	12	14	16
25	Stereo	81	44	18	-8	15	12	17
25	3D audio	85	49	13	-7	16	14	11
27	Stereo	85	46	19	-3	10	13	13
27	3D audio	85	47	19	-3	9	13	12
28	3D audio	74	31	21	-4	13	13	14
28	Stereo	80	37	23	-4	11	13	17
29	Stereo	71	40	22	-9	10	8	8
29	3D audio	61	32	19	-9	11	8	8
30	3D audio	87	45	25	-6	12	11	16
30	Stereo	84	42	25	-7	12	12	16



PID	Condition	TPS	R	PtA	QoI	PoE	SEoP	\mathbf{S}
31	Stereo	64	35	18	-15	15	11	17
31	3D audio	70	36	20	-12	15	11	13
32	Stereo	84	43	25	-13	17	12	18
32	3D audio	82	42	22	-12	18	12	16
33	3D audio	101	47	24	-5	21	14	13
33	Stereo	98	45	24	-4	20	13	17
34	Stereo	77	40	20	-8	12	13	15
34	3D audio	74	36	20	-12	16	14	15
35	3D audio	100	48	27	-3	14	14	18
35	Stereo	92	45	23	-3	13	14	17
36	3D audio	84	39	20	-3	15	13	17
36	Stereo	89	39	23	-3	17	13	16