3D Audio for Virtual Reality Exposure Therapy

A.R.D. Hoekstra
3D Audio for Virtual Reality Exposure Therapy

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Computer Science, track Media & Knowledge Engineering at Delft University of Technology

A.R.D. Hoekstra

August 21, 2013

VERSION 4.0

EXAM COMMITTEE:

DR. IR. W.P. BRINKMAN
PROF. DR. I. HEYNDERICKX
DR. R. EGMOND

Faculty of Electrical Engineering, Mathematics and Computer Science (EWI) · Delft University of Technology
Abstract

Virtual Reality Exposure Therapy (VRET) is an effective method to treat anxiety disorders and comes with many advantages over exposure *in vivo*, the latter being exposure in real-life. VRET is done by exposing patients gradually with their fear by using immersive virtual environments (VEs). Most attention in VRET research goes to visuals and less to audio, haptics, smell and taste. This report focuses on the use of audio, and in particular on 3D audio, in VRET systems. 3D audio is the only audio technique that is able to reproduce sounds as they are heard in real-life. Unlike other audio techniques, 3D audio uses a model of the human hearing in order to replicate accurate horizontal (azimuth), vertical (elevation) and distance information cues. Audio, especially combined with other modalities like vision, provides added value in terms of immersion. Audio is also indicated to have a significant effect on presence and presence is assumed to be a key factor in VRET performance. This report discusses if 3D audio can create more presence, anxiety and spatial perception compared to other commercially available audio techniques. This was done by doing three experiments, using stimuli of a flying wasp in order to generate a global sense of anxiety and/or discomfort. All audio was represented using headphones. Results indicate that people are able to hear differences in a direct comparison between 3D audio and mono, stereo and dolby headphones for the given wasp stimuli. People also report more presence and anxiety for 3D audio without the addition of visual information, but they don’t focus on these differences anymore as soon as visual information is added to the audio. This is probably because the audio is overruled by the visuals when combining the information. Vision is known to be the dominant sense, and when combining audio with visuals, visuals “take over”. As a consequence, different audio techniques may result in a similar experience. Suggestions are done for future research and it is currently advised to use stereo audio in a VRET system until future research shows otherwise. It is also stated that it may be more effective to increase the number of matching modalities instead of optimizing only one in particular.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1-1 Research questions</td>
<td>5</td>
</tr>
<tr>
<td>2 Theoretical background and related research</td>
<td>9</td>
</tr>
<tr>
<td>2-1 What are the possible benefits of 3D audio for VRET?</td>
<td>9</td>
</tr>
<tr>
<td>2-1-1 A typical VRET system</td>
<td>9</td>
</tr>
<tr>
<td>2-1-2 Presence</td>
<td>11</td>
</tr>
<tr>
<td>2-1-3 What is known about the effect of audio on presence and anxiety?</td>
<td>12</td>
</tr>
<tr>
<td>2-1-4 Cross-modality with audio</td>
<td>13</td>
</tr>
<tr>
<td>2-2 Why should 3D audio be superior to other audio techniques?</td>
<td>14</td>
</tr>
<tr>
<td>2-2-1 Human spatial hearing</td>
<td>14</td>
</tr>
<tr>
<td>2-2-2 How to measure spatial perception performance?</td>
<td>21</td>
</tr>
<tr>
<td>2-2-3 3D audio VS mono, stereo and dolby surround</td>
<td>22</td>
</tr>
<tr>
<td>2-3 What is the best 3D audio reproduction technique for VRET</td>
<td>22</td>
</tr>
<tr>
<td>3 ABX perceptual difference listening test</td>
<td>23</td>
</tr>
<tr>
<td>3-1 Method</td>
<td>23</td>
</tr>
<tr>
<td>3-1-1 Experimental design: ABX discrimination test</td>
<td>23</td>
</tr>
<tr>
<td>3-1-2 Participants</td>
<td>24</td>
</tr>
<tr>
<td>3-1-3 Listening room, hardware &amp; software</td>
<td>25</td>
</tr>
<tr>
<td>3-1-4 Stimuli</td>
<td>27</td>
</tr>
<tr>
<td>3-1-5 Procedure</td>
<td>28</td>
</tr>
<tr>
<td>3-2 Results</td>
<td>29</td>
</tr>
<tr>
<td>3-3 Conclusion</td>
<td>30</td>
</tr>
</tbody>
</table>

Master of Science Thesis

A.R.D. Hoekstra
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D  Designing and realizing the auditory virtual environment</td>
<td>115</td>
</tr>
<tr>
<td>D-1  Sound selection</td>
<td>115</td>
</tr>
<tr>
<td>D-1-1  Sound selection</td>
<td>117</td>
</tr>
<tr>
<td>D-1-2  Sound processing</td>
<td>117</td>
</tr>
<tr>
<td>E  Source code 3D audio for VRET with visuals</td>
<td>127</td>
</tr>
<tr>
<td>Bibliography</td>
<td>151</td>
</tr>
</tbody>
</table>
List of Figures

1-1 VRET setup using an HMD, source: [1]. ......................................................... 2
1-2 VRET setup using a CAVE, source: [1]. ......................................................... 3
1-3 The "speaker cage", a very expensive and impractical 3D audio setup. .............. 5

2-1 Communication between therapist, patient, and parts of the VRET system, source [2]. .................................................................................................................. 10
2-2 Spatial hearing .................................................................................................. 15
2-3 The MMA between successive pulses of a tone as a function of the frequency of the tone and the azimuth of the source (black circle = 0°, white circle = 30°, black triangle = 60° and white triangle = 75° (Source: [3])) ................................. 16
2-4 Cone of confusion. Identical ITD and ILD values for two opposite points anywhere on the surface of the cone. E.g. a and b is a front-back ambiguity, and x and y is an elevation ambiguity. (Based on [3]; from [4]) ............................................ 17
2-5 Diagram of the generation of a time delay for a composite signal produced by pinnae reflection (Source: [5]). ................................................................. 18
2-6 HRTFs recorded for a source located at 90° azimuth and 0° elevation from the right ear of 10 listeners (Source: [6]) ......................................................... 18
2-7 Difference in absolute error for azimuth and elevation between reverberant and anechoic stimuli [7] ............................................................................. 19

3-1 Conceptual model of the ABX discrimination listening test in a Within Subject Design ............................................................................................................. 25
3-2 With (left) and without (right) masking the ABX Comparator component of Foo- bar2000 using the Screen Concealer tool ............................................................. 26
3-3 A copyright free stock photo of a wasp. ......................................................... 27
3-4 The RMS of the spectrum peaks of the used flying wasp sound using the Waves PAZ Frequency Analyser ................................................................. 28
3-5 Spectrograms of the wasp stimuli for all audio techniques for each channel (left and right). .......................................................... 31
3-6 Flow of the ABX listening test. The grey box with "3D audio for VRET no visuals" will be discussed in detail in chapter 4.

4-1 Conceptual model of the experiment "3D audio for VRET without visuals": Within-subject design

4-2 The 14 attributes of the IPQ into 4 subscales

4-3 Flow of the "3D audio for VRET without visuals" experiment. The grey box "ABX test" was described in chapter 3.

4-4 Box plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for mono, stereo, dolby surround and 3D audio.

4-5 Box plots of the SPQ and SSPQ data for mono, stereo, dolby surround and 3D audio.

4-6 Box plot of the SUD score data for mono, stereo, dolby surround and 3D audio.

4-7 Fear of wasps score histogram of "3D audio without visuals" experiment.

4-8 Box plot of the heart rate means data for mono, stereo, dolby surround, 3D audio and the baseline measurement.

5-1 Conceptual model of experiment 3D audio for VRET.

5-2 Original 3D wasp model after image mapping of the skin, in standing posture.

5-3 Skeleton adaptation of 3D wasp model for flying posture.

5-4 In-door environment used in the virtual world of experiment "3D audio for VRET with visuals", source: [8].

5-5 Neutral world used for baseline measurement heart rate and galvanic skin response. Left half is screen-shot of the VR world used and right half is a picture of the actual room, source: [9].

5-6 Participant area of experimental setup of the 3D audio for VRET experiment.

5-7 Control area of experimental setup of the 3D audio for VRET experiment.

5-8 Flow and structure of the 3D audio for VRET experiment.

5-9 Box plots of all presence data for audio with visuals for VRET: 14-attributes Overall Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio.

5-10 Box plots of the SPQ and SSPQ data for all audio conditions.

5-11 Box plots of the SUD Score Post and SUD Score Increment data.

5-12 Box plot of the heart rate data for all audio conditions and the baseline measurement.

5-13 Box plot of the heart rate data for all audio conditions and the baseline measurement.

5-14 Video and audio recordings of behaviour participants during VR exposure.

5-15 Clear fright responses followed by laughing in the audio conditions.

5-16 More common subtle behaviour of laughing when wasp was very close to the head for stereo and 3D audio. For no audio the response was mostly neutral.

A-1 Box plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for all audio techniques.

A-2 Histograms of the 14-attributes Total Presence data.

A-3 Histograms of the IPQ sub-scales Spatial Presence and Involvement data.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-4</td>
<td>Histograms of the IPQ sub-scales Experienced Realism and Global Presence data</td>
<td>75</td>
</tr>
<tr>
<td>A-5</td>
<td>Q-Q plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for all audio techniques</td>
<td>76</td>
</tr>
<tr>
<td>A-6</td>
<td>Box plots of the 10-item and 13-item Spatial Perception data for all audio techniques.</td>
<td>77</td>
</tr>
<tr>
<td>A-7</td>
<td>Q-Q plots of the 13-item and 10-item Spatial Perception data for all audio techniques.</td>
<td>77</td>
</tr>
<tr>
<td>A-8</td>
<td>Histograms of the 13-item and 10-item Spatial Perception data.</td>
<td>78</td>
</tr>
<tr>
<td>A-9</td>
<td>Box plot, histograms and Q-Q plots of the SUD score data for all audio techniques.</td>
<td>79</td>
</tr>
<tr>
<td>A-10</td>
<td>Box plot and histograms of the heart rate means data for all audio techniques including the baseline.</td>
<td>80</td>
</tr>
<tr>
<td>A-11</td>
<td>Box plot and histograms of the baseline-corrected heart rate data for all audio techniques.</td>
<td>81</td>
</tr>
<tr>
<td>A-12</td>
<td>QQ plots of the heart rate means data for all audio techniques including the baseline.</td>
<td>82</td>
</tr>
<tr>
<td>A-13</td>
<td>QQ plots of the baseline-corrected heart rate data for all audio techniques.</td>
<td>83</td>
</tr>
<tr>
<td>B-1</td>
<td>Boxplots of all Presence data for audio with visuals for VRET: 14-attributes Overall Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio</td>
<td>86</td>
</tr>
<tr>
<td>B-2</td>
<td>Histograms of the 14-attributes Overall Presence data for audio with visuals for no audio, stereo and 3D audio</td>
<td>87</td>
</tr>
<tr>
<td>B-3</td>
<td>Histograms of the four IPQ sub-scales data for audio with visuals for VRET: Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio</td>
<td>88</td>
</tr>
<tr>
<td>B-4</td>
<td>Q-Q plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio</td>
<td>89</td>
</tr>
<tr>
<td>B-5</td>
<td>Box plots of the 10-item (SPQ) and 13-item (SSPQ) Spatial Perception data for all audio conditions.</td>
<td>90</td>
</tr>
<tr>
<td>B-6</td>
<td>Histograms of the (SPQ) 13-item and 10-item (SSPQ) Spatial Perception data for all audio conditions.</td>
<td>90</td>
</tr>
<tr>
<td>B-7</td>
<td>Q-Q plots of the 13-item (SPQ) and 10-item (SSPQ) Spatial Perception data for all audio conditions.</td>
<td>91</td>
</tr>
<tr>
<td>B-8</td>
<td>Box plot, histograms and Q-Q plots of the SUD score data measured after the exposure for all audio conditions.</td>
<td>92</td>
</tr>
<tr>
<td>B-9</td>
<td>Box plot, histograms and Q-Q plots of the increment in SUD score during the exposure data for all audio conditions.</td>
<td>93</td>
</tr>
<tr>
<td>B-10</td>
<td>Box plot and histograms of the heart rate data for all audio conditions.</td>
<td>94</td>
</tr>
<tr>
<td>B-11</td>
<td>Q-Q plots of the heart rate data for all audio conditions.</td>
<td>95</td>
</tr>
<tr>
<td>B-12</td>
<td>Box plot and histograms of the Galvanic Skin Response data for all audio conditions.</td>
<td>96</td>
</tr>
<tr>
<td>B-13</td>
<td>Q-Q plots of the Galvanic Skin Response data for all audio conditions.</td>
<td>97</td>
</tr>
<tr>
<td>D-1</td>
<td>The Neumann KU 100 binaural dummy head microphone.</td>
<td>118</td>
</tr>
<tr>
<td>D-2</td>
<td>An anechoic room. Source: [10].</td>
<td>119</td>
</tr>
<tr>
<td>D-3</td>
<td>A regular cardioid polar microphone pattern. Source: [11].</td>
<td>120</td>
</tr>
<tr>
<td>D-4</td>
<td>The CLAM Network Editor.</td>
<td>122</td>
</tr>
<tr>
<td>D-5</td>
<td>A glance of the GUI of SoundLocus.</td>
<td>123</td>
</tr>
<tr>
<td>D-6</td>
<td>A glance of the GUI of Longcat AudioStage.</td>
<td>124</td>
</tr>
<tr>
<td>D-7</td>
<td>A glance of the GUI of Longcat H3D plugin.</td>
<td>124</td>
</tr>
</tbody>
</table>
List of Figures
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Audio technique pairs ordered using balanced latin square [12]. (M = Mono, S = Stereo, D = Dolby Surround and 3D = 3D audio)</td>
<td>25</td>
</tr>
<tr>
<td>3-2</td>
<td>Response matrix for ABX Bernoulli trials</td>
<td>29</td>
</tr>
<tr>
<td>3-3</td>
<td>Response matrix for ABX Bernoulli trials</td>
<td>30</td>
</tr>
<tr>
<td>4-1</td>
<td>Audio techniques ordered using balanced latin square [12]</td>
<td>34</td>
</tr>
<tr>
<td>4-2</td>
<td>Spatial Perception questionnaire attributes</td>
<td>38</td>
</tr>
<tr>
<td>4-3</td>
<td>Summary of Friedman test results regarding the effect of all audio techniques on overall presence and its sub-scales for &quot;3D audio for VRET without visuals&quot;. Significant results are marked with an asterisk *</td>
<td>44</td>
</tr>
<tr>
<td>4-4</td>
<td>Significant larger effects for $\alpha = 0.009$ of 3D audio on presence, SUD score, heart rate and spatial perception compared to mono, stereo and dolby headphones</td>
<td>48</td>
</tr>
<tr>
<td>5-1</td>
<td>Audio conditions and stimuli ordered using a balanced Latin square [12]</td>
<td>52</td>
</tr>
<tr>
<td>5-2</td>
<td>Stimuli counter balance ordering using a balanced Latin square [12]</td>
<td>52</td>
</tr>
<tr>
<td>5-3</td>
<td>Audio conditions and stimuli ordered using a balanced Latin square [12]</td>
<td>53</td>
</tr>
<tr>
<td>5-4</td>
<td>Summary of Friedman test results regarding the effect of all audio conditions on overall presence and IPQ sub-scales for &quot;3D audio for VRET with visuals&quot;. Significant results are marked with an asterisk *</td>
<td>60</td>
</tr>
<tr>
<td>5-5</td>
<td>Significantly larger effects for $\alpha = 0.017$ of 3D audio on presence, SUD score, heart rate, galvanic skin response and spatial perception compared to no audio and stereo</td>
<td>67</td>
</tr>
<tr>
<td>5-6</td>
<td>Significantly larger effects of audio on presence, SUD score, heart rate and galvanic skin response compared to no audio</td>
<td>67</td>
</tr>
<tr>
<td>D-1</td>
<td>Sound scenario’s evaluated according to the formulated requirements. All requirements are weighted equally except for R_M1, which is weighted double.</td>
<td>125</td>
</tr>
</tbody>
</table>
Preface

As a child I had a fear of sound. Crying when the curtains were closed, frightened when the doorbell rang and even scared by seeing a balloon that could, but never did explode. I helped my parents stop listening to music and live in silence. And now, after many years of gradual exposure, I am shaking buildings with my music, terrorizing my environment by making loud noises and writing a master thesis about curing people from their fear, by using sound.

I would like to thank my daily supervisor Dr. ir. W.P. Brinkman for his assistance during the whole project. I would also like to thank Prof. dr. I. Heynderickx and Dr. R. van Egmond for their help, feedback and for making all this possible, and Thomas Rens for revising this thesis. Many thanks goes to my parents who always supported and inspired me in both science and creativity. Also thanks to my brother who had to listen to all the loud noises in the room next door, and to my friends for their support. And much love to my girlfriend, for her great love and support.

Delft, University of Technology
August 21, 2013

A.R.D. Hoekstra
Chapter 1

Introduction

Fear is a crucial emotion that serves a function of activating the defensive behavioural system of humans and animals to protect them against potentially dangerous environmental threats [13]. But, for people suffering from an anxiety disorder, their fear is not proportional to the actual potential danger and they are being overwhelmed by an excessive amount of experienced fear. In this case, their fear is not helpful any more and it hinders them in their everyday life, causing avoidance for these fearful situations [14]. Anxiety disorders are not that uncommon. According to the Trimbos Institute, 19.6% of the Dutch population are suffering from at least one anxiety disorder, during their lifetime and 10.1% during the last currently measured 12 months [15]. The Trimbos Institute follows the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) [16]. Other recent surveys have stated, also according to the DSM-IV criteria, that 18% of the American population [17] and 15% of the European population [18] have experienced an anxiety disorder in the last 12 months. According to the DSM-IV diagnostic classification system, anxiety disorders are classified as follows: panic disorder, generalized anxiety disorder, agoraphobia without panic disorder, specific phobia, social phobia, post-traumatic stress disorder, obsessive-compulsive disorder and separation anxiety disorder [17].

Some common anxiety disorders can be successfully treated using psychotherapy, especially Cognitive Behavioural Therapy (CBT) [19]. A key element of CBT is gradual exposure, in which the patient is confronted with their fear in a structured, conscientious manner, until anxiety is reduced [19], hence the name exposure therapy. The aim is that gradual exposure leads to new and more neutral emotional memory structures that overrule the old anxiety ones [20]. Information about the neurological aspects and processes in the brain related to the effect of exposure therapy can be found in [21, 22, 23, 24].

Exposure therapy was originally only done in real life, called exposure in vivo. But, since exposure therapy is sometimes very impractical and costly, for example to treat fear of flying where people have to actually fly in a plane, an alternative solution has been suggested: Virtual Reality Exposure Therapy (VRET). Like the name suggests, exposure therapy by making use of immersive Virtual Environments (VEs). VRET is mostly done using a combination of multiple modalities like visuals, audio, haptics...
Figure 1-1: VRET setup using an HMD, source: [1].
Several studies showed that there were significant reductions in anxiety with VRET [25, 26]. VRET and exposure in vivo were also compared in some studies [26, 27], showing that there were similar reductions of fear with both methods. After 6-month and 1-year follow-up meta-studies it was shown that these improvements were maintained [19]. VRET was found to be as effective as exposure in vivo [26].

For VRET to work, patients need to feel a sense of presence, which means the sense of being in a virtual environment rather than the place in which the participant’s body is actually located [28]. Presence is assumed to be an important moderating variable for VRET. It is still unclear if a higher sense of presence contributes to a higher VRET performance, and if there is a linear relation between anxiety and presence. Some found a linear relationship between presence and anxiety [25, 29, 1] and some did not [30]. But it is assumed that some degree of presence is necessary in order to achieve successful VRET [30, 20, 28]. The relation between presence and anxiety and the relation between presence and VRET performance are still important subjects of research.

Another important phenomenon in VRET is immersion, which refers to the technical capability of the system to deliver a surrounding and convincing environment with which the participant can interact [28]. Important parameters of immersion are the field of view, the number of simulated sensory systems, the quality of rendering in each sensory modality, the extent of tracking, the realism, the frame rate and the latency [28]. Factors of immersion are known to have a high impact on the sense of presence [28] and therefore it can probably play an important role.

Most attention in VRET research, but also in general VR research, is directed at the visuals and less to the other modalities audio, haptics, smell and taste. This is probably because vision is known to be the dominant sense [31]. But, it is also known that other modalities provide valuable added information when properly used in combination with vision [31]. This report focuses on the use of (3D) audio for VRET systems, combined with or without visuals.
Studies indicate that audio has a significant effect on presence. For example, one study showed that spatialized sound was associated with higher reported presence than either no sound or non-spatialized sound [32]. Another study showed that the use of individual head-related transfer functions (HRTFs) was significantly and positively associated with reported presence and illusory self-motion [33]. A HRTF is a filter function based on the shape of the head, pinnae of the ears and shoulders, which results in a highly realistic reconstruction of three-dimensional audio. Furthermore, a known phenomenon is that suddenly deafened individuals have a sense of disconnectedness with their surroundings in the real world, a sense that the world seems "dead", which suggests that they lost a sense of presence in the real world [34]. At very close range, sounds can even cause fight or flight reactions [31].

Studies indicate that audio can have a significant effect on anxiety. Studies have shown that music interventions are able to significantly reduce anxiety [35]. Also, a study related to pre-operation anxiety showed significant reductions in anxiety using audio [36]. In movies and games, audio is used to stimulate emotional responses using music and sound effects [37]. There is a clear relation between audio and anxiety.

In terms of immersion, audio can deliver information of much added value when used together with other modalities. The field of view of sight is around 180 degrees in total [38], while audio in real life delivers information 360 degrees in every location lying on a 3D sphere around us at all times. It can also provide information about visually hidden objects within the field of view or even in the dark [38]. This is possible because the brain, the inner ear and the external ears (pinna) work together to make inferences about location and distance. The ability to localize sound sources may have developed in humans as an evolutionary necessity, since the eyes can only see a fraction of the world around a person, and vision is hampered in darkness. Vision is known to be the dominant sense, but research has shown that the visual modality is dominant for stimuli positions at about 15-25 degrees azimuth (horizontal orientation), but that auditory information provides superior localization information at greater angles within the visual field [38].

Most audio under research or used in VR systems use either two-channel audio (stereo) or multiple-channel surround sound (like dolby surround). Only in a rare occasion 3D audio is used or examined. 3D audio is different from two-channel stereo or multiple channel surround sound, because it is the only audio technique with the possibility to accurately provide information about sound source location outside the head on the horizontal (azimuth) and vertical (elevation) plane and even provide information about distance. Two-channel stereo is either heard inside the head with headphones or from a location on the line between the two speakers. Surround sound delivers more spatial information, but only on the horizontal plane, or in line with the speaker array, but it lacks in correct elevation representation. By doing pinpointing experiments, it is shown that 3D audio is significantly more accurate in sound source localization and distance perception [39, 40, 41]. A lot of commercial audio products are described as having 3D capability, but in fact there is great disparity between the various technologies in use. Unfortunately, many products are marketed with the most exaggerated claims. For example, a number of stereo multimedia speakers are marketed as having real 3D technology. But in fact, these speakers incorporate a simple circuit that has the effect of widening the perceived sound field of a stereo recording. That is, the sound images that would normally extend to the locations of the left and right speakers are widened to extend beyond the speakers. These systems should more properly be called stereo enhancement or "widening" systems instead of 3D audio [31]. They have no ability to position individual sounds around a listener, nor do they have the ability to position sounds...
behind, above, or below the listener. We use the term 3D audio to describe a much more sophisticated system than can ideally position sounds anywhere around a listener. More used terminologies for 3D audio are 3D sound, binaural audio, binaural sound and in some special cases holophonics, which is a patented 3D audio reconstruction technique. With the exception of a few applications, 3D audio is commercially not used as a standard audio technique like two-channel stereo or surround sound. 3D audio is always reproduced using either headphones or multiple speakers, but very different techniques can be used, all having their own weaknesses and strengths [42, 31]. At the moment of writing, none of them is completely flawless and some of them are extremely expensive and impractical to realise, for example the "speaker cage" which is illustrated in figure 1-3. To use 3D audio in a VRET system, or in any VR system at all, requirements and budgetary constraints have to be formulated for the sound component in order to find the most suitable 3D audio technique [31].

![Figure 1-3: The "speaker cage", a very expensive and impractical 3D audio setup.](image)

Currently, stereo and dolby surround are the most commonly used audio techniques in VR [31]. Although there is a lack of literature on which audio technique is most commonly used for VRET, it is plausible that stereo and dolby surround are also most commonly used for VRET applications.

### 1-1 Research questions

Since the main moderating variable of successful VRET is assumed to be presence [1, 20, 28], it would be a relevant question to study the effect of 3D audio on presence. But, since the relation between presence and anxiety is still unclear [1], it would also be a relevant question to study the effect of 3D audio compared to other audio techniques on anxiety.

Since vision is known to be the dominant sense [31], VRET is nowadays largely used in combination with visuals, presented by either monitor screens or HMDs and CAVE environments, either in 2D or 3D. So a relevant question is to study the effect of 3D audio compared to the other audio techniques combined with visuals. But, in order to see the isolated effect of 3D audio, it is also important to study the effect without visuals.
The main research question of this master thesis, motivated by the factors mentioned above, is:

*Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET?*

In order to answer this question, three experiments were done, each answering a separate research question that leads to a final conclusion of the main research question. The first banal question that needs to be answered before any useful interpretation can be done about comparison experiments is:

*Can people perceive a difference between 3D audio, dolby surround, stereo and mono?*

If the answer to this question is "no", then the comparison between the different audio techniques in the main research question will be useless since there is no perceptual noticeable difference and the variance of the outcome of the comparison, if there is any, must be caused by another effect than switching the dependent variable audio technique. In that case, only the global effect of audio on presence, spatial perception and anxiety can be researched, regardless of the audio technique. If the answer is "yes", the difference in effects on presence, anxiety and spatial perception is probably caused by the change in audio technique.

After answering this question, the following question can be asked:

*Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET without using visuals?*

The reason for the "without visuals" is to isolate the effect of 3D audio on presence, anxiety and spatial perception from any other modality. If 3D audio doesn’t have a significant effect on one of these dependent variables, then it is likely that 3D audio also has no significant effects with other modalities present. First the audio was isolated in order to see the effect of the audio techniques itself on presence, anxiety and spatial perception. It is suggested by Begault [31] that if visuals are absent, audio plays a more important role in our perception, since vision is known to be the dominant sense, suppressing the focus on audio. If one closes his/her eyes he or she can almost "see" what is happening.

But, since VRET is mostly done with visuals, the final and most important question for practical VRET usage is:

*Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET with using visuals?*

The questions were answered in this order since the outcome of each question is related to the next. Even stronger: if the results were negative in one question, it directly influences the results of the next.

The first research question "Can people perceive a difference between 3D audio, dolby surround, stereo and mono?" was answered by doing a difference listening test, which is described in chapter 3. Since its answer was "yes" for all audio techniques compared to 3D audio, differences in presence, anxiety and spatial perception obtained for different audio techniques gathered by follow-up experiments can be associated with the audio technique. The experiment that answers the research question
"Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET without using visuals?" is described in chapter 4. To answer the final research question "Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET with using visuals?", an experiment was designed based on the experiment without visuals, but now done with a typical VRET setup including visuals. This final experiment is covered in chapter 5.
Chapter 2

Theoretical background and related research

This chapter gives an overview of the most important related theoretical background and related research. Section 2-1 starts with the question why 3D audio can be of possible benefit for VRET. Section 2-2 explains why 3D audio should be superior to other audio techniques in terms of spatial perception. To conclude the theoretical background, a suggestion is given in section 2-3 for the best and currently most suitable 3D audio technique for VRET usage.

2-1 What are the possible benefits of 3D audio for VRET?

To answer this question, we first have to know how a typical VRET system works, described in subsection 2-1-1. Since presence is seen as the most important moderating variable to VRET performance, this topic will be described in subsection 2-1-2. Also, an overview is given about some related studies of the effects of audio on presence and anxiety in subsection 2-1-3. The use of multiple modalities together with audio and the effect on VRET related topics is presented in subsection 2-1-4.

2-1-1 A typical VRET system

VRET is effective only if the following three conditions are met [1]:

1. Participants need to feel present in the virtual environment to be able to experience the environment fully as a place that was actually visited.

2. The virtual environment should be able to elicit emotions (e.g. anxiety), otherwise, extinction of anxiety will not occur.

3. Extinction and co-occurring cognitive changes have to generalize to real simulations so that real-life situations will not be avoided any longer or will be endured with less anxiety.
A typical VRET system is a system in which one (or multiple) patients are gradually exposed to their fear, controlled and supervised by a therapist. In the future, the therapist can possibly also be a virtual therapist using artificial intelligence and reasoning techniques. This requires ethical and technological problems to be solved and a very good understanding in all aspects of VRET. Communication and parts of a VRET system designed by the Delft University of Technology [2] is illustrated in figure 2-1. The patient is typically being exposed to a VR world using an HMD and speakers or headphones. The computer system contains a tracker, following the movement of the patient for interaction possibilities. The view of the patient and the sound of the VR world are also sent to the therapist. The therapist can control certain variables and properties of the VRET system and the VR world by using a Keyboard, joystick and/or a mouse. The posture of the patient is also perceived by the therapist. The patient mainly perceives the sound of the real world and his or her actions in it. There are also other VRET systems with different a structure, components and modalities.

Figure 2-1: Communication between therapist, patient, and parts of the VRET system, source [2].

In a broad sense, VRET treatment starts by putting a patient in an immersive virtual environment where the person is exposed to his or her fear using one or a combination of different modalities (visual, audio, touch, taste and/or smell). The patient can possibly interact with the virtual environment through different sources of input like joysticks, haptic sensors, movement sensors, head rotation sensors, etc. If the presence is high enough and the virtual environment elicits an experience of anxiety, the exposure results in a psychological and physiological response on the patient. The patient will probably feel less comfortable and certain physiological factors like heart rate, galvanic skin response, body temperature, but also behaviour like eye blinking, will change. The psychological and physiological state can be measured by asking the patients to rate their anxiety in Subjective Units of Discomfort (SUDs, either 0 - 10 or 0 - 100) [1], by objective measurements through sensors like
galvanic skin conductors and heart rate sensors, by observing the patient's behaviour, by interviewing the patient and/or using questionnaires. Typically, anxiety is measured using SUD ratings, objective measurement sensors and observation [1]. The therapist will evaluate all measurements and choose a new suitable anxiety exposure (by changing some parameters or choosing a complete new exposure). This is an iterative process until the patient has been exposed enough and the therapist ends the session.

**Moderator variables to successful VRET performance**

The main goal of VRET is to maximize the performance of the treatment. Or in other words, to reduce the anxiety level of the patient for the real-life scenario as good as possible in an efficient and effective way. The moderator variables that determine the success of the VRET treatment and how they are related are still subject to research, but some important known moderator variables are the level of immersion [43], the sense of presence [1], personality characteristics of the patient (like hypnotizability and absorption, but also characteristics like age, gender, and ethnicity) and VR system characteristics [1]. The chosen exposure and how it is represented is important, including whether the different modalities match or mismatch [38]. Also factors like quality or fidelity probably play a role in VRET performance. Good use of sound can elicit emotions like fear. This is done in cinema and gaming, but it can also be done in VRET.

### 2-1-2 Presence

Many definitions of presence exist, but it is commonly used as "the sense of being in a virtual environment rather than the place in which the participant's body is actually located" [44, 45, 46, 47, 48, 49]. Multiple suggestions are done to further define presence into more special subtypes of presence [50, 51], where all subtypes together encapsulate the global sense of presence.

It is known that a sufficient amount of presence is needed in order to achieve effective VRET performance [52, 53], but it is still not known if a higher degree of presence contributes to a higher degree of anxiety [1]. Some say that it does [54, 55, 52] and others say that it doesn’t [56]. This is an important subject of research. If we can understand the effect of presence on anxiety and VRET performance, and if we know the effects of audio and possibly (combinations of) other modalities and factors on presence, we get a better total understanding of how to increase VRET performance.

There are many factors that can increase the sense of presence. Surprisingly, it is known that realism is not the most important factor for increasing presence [57]. It seems to be far less important than other immersion parameters such as head tracking, latency [58, 59] and interaction methods [57].

There are different methods for measuring presence, but it is still an important challenge because most methods have their shortcomings. The most commonly used methods for measuring presence are discussed below.

**Questionnaires**

A common approach is using questionnaires [59, 60, 61, 62]. Questions mostly have an ordinal scale that correspond to responses between two extremes, for example, 1 meaning "no presence" and 7 meaning "complete presence". But, questionnaires have been shown to be unstable [63, 57] and typical questionnaires cannot discriminate between presence in a VE and physical reality [62]. There are multiple known questionnaires to measure presence. For example, the Presence Questionnaire (PQ) by Witmer and Singer [64], The ITC sense of presence inventory [65], the Slater-Usoh-Steed...
2-1 What are the possible benefits of 3D audio for VRET?

The Questionnaire (SUS) [66] or the IGroup Presence Questionnaire (IPQ) [67]. The IPQ is a 14-item questionnaire to measure overall presence, but it can also be decomposed into four different subscales of presence: Spatial Presence, Involvement, Experienced Realism and Global Presence. Spatial Presence is the sense of being physically present in the VE, involvement is the attention devoted to the VE and the involvement experienced, Experienced Realism is the subjective experience of realism in the VE and Global Presence is the general "sense of being there". This can give more insight into multiple detailed aspects of the global concept "presence".

**Behaviour**

Behaviour can also be used to measure presence [57]. If participants behave as if they are in an equivalent real environment, there is a notion of presence. This behaviour can be for example eye blinking, moving the head to avoid an incoming virtual object, and many more scenario based behaviour. Behaviour can be recorded by video, audio or observation.

**Physiological measurements**

Physiological measures can be used to measure presence, which is a specialization of the behaviour methodology. Some common physiological responses that are used to measure presence are heart rate, galvanic skin response and body temperature. If the physiological response of a person in a real life situation is equivalent to the physiological response in the simulated virtual environment, this is a sign of presence. But, these measurements are only useful for situations with obvious physiological responses, like for example in recreating a feared situation, which is the case in VRET.

**Breaks in presence BIP**

Another method to measure presence is by looking at the breaks in presence (BIP) that the participant experiences [57]. For example, if the participant collides with the mounted equipment or if s/he perceives a virtual object as a pixel map instead of a real object that it is supposed to represent, the person experiences a break in presence in the virtual environment and feels to be back into the real world again. This method is not scenario dependent and therefore can be used in every VR application to measure the presence.

2-1-3 What is known about the effect of audio on presence and anxiety?

The effects of audio on presence are frequently underrated [68]. But, some studies indicate that sound has a significant impact on presence. Studies about "inverse presence", presence in physical reality when people experience a sudden and lasting loss of hearing [57], state that they experience a sense of disconnectedness with their surroundings and a sense that the world seems ‘dead’ [69].

According to Ramsdell [70], there are three different levels that the auditory modality provides:

*The social level* which contains communication signals which are responsible for all symbolic information like music, language, etc.

*The warning level* which contains signals that carry information about signalling or warning significance

*The primitive level* which contains signals that function as the auditory background to everyday life. Background sounds with neither symbol nor warning, but incidental sounds made by objects in the environment and by ourselves as we interact with objects in the environment.
Especially the incidental background sounds give people the feeling of being part of a living world and contribute to the sense of being live, increasing the sense of presence [70, 69, 71, 32, 72]. Studies about the absence or addition of auditory cues in virtual environments have shown a positive effect on presence with added (matching) audio [73, 74]. Another study showed that spatialized sound was associated with higher reported presence than either no sound or non-spatialized sound [32]. According to [75, 76], the lower frequencies of sound, approximately from 20 to 100 Hz that are both felt as vibration and heard as sound, increase the naturalness, enjoyment and interest in the virtual environment. A study about the use of individual head-related transfer functions (HRTFs) (more information about HRTF can be found in section 2-2-1) states that the use of these individual HRTFs was significantly and positively associated with reported presence and illusory self-motion [77]. Findings about the effect of quality or fidelity of audio on presence are mixed [38]. It is still unclear if a higher fidelity audio system will provoke a higher sense of presence. There is of course a limit on where sound distortions are becoming distracting and possibly decrease presence. It is also pointed out that audio is an important contributor for social presence, especially when communication is done orally [78].

Audio can also have an effect on anxiety. For example, music interventions are able to significantly reduce anxiety [35], or pre-operation anxiety showed significant reductions in anxiety using audio [36]. Also, in games and/or movies sound effects and music are frequently used to elicit emotions, among which, anxiety [37]. There is still not much known about the relation between anxiety and audio in the field of cognitive behavioural therapy.

All findings mentioned above are global effects of sound on presence and anxiety, and do not say anything about the 3D aspect of audio. There is still a gap in knowledge on whether the additional spatial information cues in 3D audio have a higher effect on presence and/or anxiety.

2-1-4 Cross-modality with audio

In terms of immersion, sight and audio complement each other quite well. The field of view of sight is only ±90 degrees [38] and audio delivers information at any location 360 degrees around us, but with less accuracy [79]. Sound can trigger the head to move to determine the exact location with sight. Also, a known problem with audio are front-back reversals and up-down reversals [79], which refers to the phenomenon that the auditory system cannot distinguish between sounds coming from the front or the back or from above or below. More information about this phenomenon can be found in section 2-2-1. These problems can be solved with sight (if the objects are not hidden or in the dark) [79]. On the other hand, audio can still provide information in the dark or about visually hidden objects within the field of view [38].

It is known that using 3D audio techniques can create localization errors [79]. These errors can be compensated with visuals if the error is not larger than approximately 10 degrees [80] known as the ventriloquism effect [81]. A ventriloquist is a person that manipulates his or her voice so that it appears that the voice is coming from elsewhere, usually a puppet that he or she controls with his or her hands [81]. Sound that is highly correlated with the motion of a visual object can be coupled in a way that the auditory information seems to come from the perceived visual object, even if the auditory object is localized slightly different [79]. This ventriloquism effect is a very important phenomenon in 3D audio reproduction when combining it with visuals. But, if the error is too large (more than
approx. 10 degrees in space [80]), the audio and visuals will result in a mismatch.

It is also shown that visual quality can be increased just by adding high quality audio [38]. For example, it has been shown that medium quality rated visuals can be significantly improved to high quality rated visuals just by adding medium or high quality audio [82]. The opposite is less clear, but a common statement made in the gaming industry is that high quality visuals do not make audio sound better; in fact, they make audio sound worse [83].

A time latency between audio and visuals can be annoying, but it is not clear what the effect is [38]. But it is shown that temporal asynchrony affects speech intelligibility because speech is more intelligible with the presence of correct lip movement [83].

2-2 Why should 3D audio be superior to other audio techniques?

In order to answer the question why 3D audio should be superior to other audio techniques, it is important to understand some concepts of human spatial hearing. Human spatial hearing is discussed in section 2-2-1, current measurement instruments in order to measure 3D audio spatial attributes are discussed in section 2-2-2 and finally a comparison is made of 3D audio with mono, stereo and dolby surround in section 2-2-3.

3D audio reproduction, either through loudspeakers or headphones, is based on the principle that people have two ears. The differences in the signal perceived at the two ears are called interaural differences and provides cues for localization perception.

2-2-1 Human spatial hearing

Human spatial hearing includes all aspects that determine why a listener can observe a sound source including its location, both horizontally (azimuth) and vertically (elevation), its distance, the image size of the sound source, also called sound source width, and the environmental context. This is illustrated in figure 2-2. Most localization information cues are based on the fact that humans listen with two ears, both receiving different information about the sound around them.

The human localization capability is not flawless. The threshold angle over which the listener cannot distinguish the location of two separate sound sources anymore is called the minimum audible angle (MMA). The MMA varies for different degrees of azimuth and elevation, but also per frequency range. An illustration of the MMA at certain degrees of azimuth for different frequencies is given in figure 2-3. The MMA gets larger for larger azimuths, especially in the frequency range of 1 kHz to 2.5 kHz, and above 4 kHz. It is also known that humans are better in localizing sound sources horizontally than vertically [31]. Sound source localization also varies between listeners, since every person’s outer ear’s shape, head and torso differ.

This section provides insight into the most relevant aspects that are needed in order to understand the basics of human spatial hearing, which is the fundamental theory for 3D audio reconstruction.

Azimuth and elevation

Interaural differences

The two most important cues for azimuth sound localization are the Interaural Time Differences (ITD)
Theoretical background and related research

**Figure 2-2: Spatial hearing**

and *Interaural Level Differences (ILD)* (together called *binaural cues*) [31]. ITD are the time differences of the arrival of a sound at both ears. ILD are the differences in intensity level of a sound at both ears. The ITD cue works for frequencies from 125 Hz to 1.5 kHz. Frequencies lower than about 125 Hz cannot be localized accurately. Frequencies higher than 1.5 kHz have a wavelength that is smaller than the diameter of a human head and the ITD are larger than the duration of a few signal periods. For these frequencies, the auditory system is no longer able to determine the time differences. But, if the amplitude of these frequencies fluctuates, an "envelope cue" [79] is created with a certain repetition pattern and this information can be used for localization [79].

For frequencies higher than 1.5 kHz, the head functions as an acoustic barrier, meaning that when a sound comes from aside, it will reach the contralateral ear (the ear that is on the other side of the head in relation to the sound source) with a lower intensity than the ipsilateral ear (the ear that is closest to the sound source). These interaural intensity differences are used for localization. Both ITD and ILD are only specifying the azimuth, not the elevation.

There is a region where both ITD and ILD are similar, called the Cone of confusion [84], which is illustrated in figure 2-4. For sound sources in this region, the ITD and ILD do not provide enough information for accurate localization, resulting in front-back reversals and up-down confusions. Front-back and up-down confusions can be disambiguated by the so-called pinna cues, the resonances and anti-resonances of the outer ear that specify the direction where a sound comes from, which is illustrated in figure 2-5.

*Head-related transfer functions*

Before a sound reaches the eardrum, it is filtered by the shape and size of the outer ear (pinnae), the head and the torso. All cues together that are characteristic for each individual person to localize sound sources, is called the *head-related transfer function* (HRTF) [79]. Each person has its own
Why should 3D audio be superior to other audio techniques?

Figure 2-3: The MMA between successive pulses of a tone as a function of the frequency of the tone and the azimuth of the source (black circle = 0°, white circle = 30°, black triangle = 60° and white triangle = 75° (Source: [3])

individual HRTF, almost like a fingerprint [85]. An illustration of some individual differences can be found in figure 2-6. HRTF is a very important topic in 3D audio reproduction because much information is present in the HRTF. If the individual HRTF of the listener is used in an audio system, the performance is very accurate for this person (almost as accurate as in real life if all other cues are optimized). But, gathering and using individual HRTFs is a very complex, expensive and impractical task [79]. Therefore, most of the time global HRTFs are used. The use of global HRTFs results in a degrading localization accuracy, externalization errors and reversal errors [86, 87, 88, 89]. Externalization is the degree of hearing sound sources outside one’s head together with a perception of distance. Externalization errors occur when sound sources are being perceived as being too close or even inside the head, called inside the head location (IHL) [79]. Certain people have a higher accuracy on average in correctly localizing sounds in real-life. These people are called “good localizers” [79]. To improve results with global HRTFs, HRTFs of good localizers can be used [31]. More information about good localizers and localization performance can be found in [90].

Head Movement

Another cue of information comes from Head Movement of the listener. It is known that allowing a listener to move his/her head improves sound localization [91, 92, 93]. If a listener wants to localize a sound source, he or she moves his or her head in order to minimize the interaural differences as a sort of “pointer”. Ambiguities such as front-back reversals can also be solved with head movement. For example, if there is an ambiguity whether the sound comes from 30° or from 150° azimuth to the right, the listener moves his or her head to the right. If the interaural differences are minimized, it must be at the front, but, if the interaural differences are increased, it must be to the rear. Listeners apparently integrate some combination of the changes in ITD, ILD, and movement of spectral notches and peaks that occur with head movement over time [79].
Sound Source Movement
A moving sound source causes interaural changes which can be used for localization. But still, localizing a moving sound source is known to be hard and when the head of the listener is also moving (without moving the head along with the sound source) it is almost impossible [79]. Whether or not source movement will improve or degrade localization ability is dependent on the movement velocity, type of sound source, and location of movement. But the cue is not as strong as localizing static sound sources by moving the head [79].

Reverberation
Reverberation in virtual and real worlds can decrease the localization performance, especially in situations where the strength and timing of early reflections is such that the sound is "smeared" in time and space [79]. Figure 2-7 illustrates the results of an experiment by Begault (1992) [7] where the difference in error for azimuth and elevation between reverberant and anechoic stimuli is visible. But, reverberations can help to externalize sound sources, which is needed for headphone 3D audio reproduction, because sounds are often perceived as being too close. It is also known that reverberation is an important cue for distance perception.

Expectation and Familiarity
The familiarity and expectation of a listener with a sound source is a strong cue for localization [79]. If a listener is familiar with a sound source, certain schemas are activated in the brain, expecting certain behaviour of the sound source. For example, if a person is at home and hears a car driving, s/he knows almost exactly the location of a passing car because the position of the sound source in the environment around is known. Also, there is the previously mentioned ventriloquism effect, which
Why should 3D audio be superior to other audio techniques?

Figure 2-5: Diagram of the generation of a time delay for a composite signal produced by pinnae reflection (Source: [5]).

Figure 2-6: HRTFs recorded for a source located at 90° azimuth and 0° elevation from the right ear of 10 listeners (Source: [6])

couples sound with visual objects if they are highly correlated in a way that the auditory information seems to be coming from the perceived visual object, even if the auditory object is localized slightly different [79]. Expectation and familiarity is also an important cue for distance perception.

**Distance perception**

In distance perception, there is *absolute distance* and *relative distance*. Absolute distance perception refers to a listener’s ability to estimate the distance (in for example centimetres or meters) of a sound source upon initial exposure, without benefit of the cognitive familiarity. Relative distance perception includes the benefits gained from listening to the source at different distances over time, perhaps within an environmental context or in comparison with other objects, like for example "This object is twice as far as that object".

**Intensity (or perceived loudness) and intensity changes**

The intensity or intensity changes of a sound source is a well-known distance cue [79]. If one is
familiar with the sound source, its intensity can give an idea about how far away the sound is from the observer. Also, if the intensity of a sound source is becoming louder, the sound source is probably coming closer and when getting less loud it probably moves away from the listener. Auditory distance is learned from a lifetime of visual-aural observations, correlating the physical displacement of sound sources with corresponding increases or reductions in intensity [79]. Theoretically, under anechoic conditions, the \textit{inverse square law} can be used to predict distance, which means that a decrease of 6 dB in sound intensity corresponds with a doubling of distance. The Sound Intensity Level (SIL) in dB is the ratio of a sound source’s intensity to a reference level:

\[ L = 10 \log_{10} \left( \frac{I}{I_{ref}} \right) dB \]

where \( I \) is the intensity of the sound source and \( I_{ref} \) is the intensity of the reference in W/m\(^2\).

When sound sources are getting close to the head, a stronger effect of intensity changes will occur. The closer to the head, the more the head functions as an acoustic barrier for high frequencies. Very close to the head, for high frequencies, the intensity in the ipsilateral ear increases with smaller distance, but decreases in the contralateral ear. These changes in ILD appear to be a very strong source of information for the perception of a sound source that approaches the listener very closely from aside.

\textbf{Reverberation and R/D Ratio}

In a reverberant context, when a sound source moves away from the listener, the level of the sound source will decrease, while the reverberation level will remain constant [79]. This \textit{R/D Ratio}, or Reverberant-to-direct sound ratio tells something about the distance of a sound source. But, there is still a discussion about this phenomenon and how it relates to other (distance) cues. According to [79], in a reverberant context, the R/D Ratio functions as a stronger cue for distance than intensity scaling. Sheeline [94] states that reverberation provides the "spatiality" that allows listeners to move from the domain of loudness inferences to the domain of distance inferences. Von Bekesy [95] states
that a changing R/D ratio causes a perception of sound source movement, but that the R/D ratio itself is not the basis for auditory distance. Despite the different interpretations of the role of the R/D ratio, it is still clear that it provides information for distance perception.

**Familiarity and expectation**

Familiarity and expectation are very strong distance cues. If a listener has knowledge about a perceived sound source and about the environmental context, the listener can expect the sound source to be at a certain distance. Also, when a listener hears multiple familiar sound sources, one can predict the relative distance between all objects in combination with the perceived loudness and environmental context. If the environmental context and the sound sources are known, certain schema’s in the brain are activated and the distance is mapped to these schema’s in order to get a good prediction. Auditory distance is learned from a lifetime of correlating visual-aural observations [79].

**Spectral changes**

When a sound source moves into the distance, there is the effect of *diminished high-frequency content* [96], because the higher frequencies of a complex sound are increasingly affected by air humidity and temperature [79]. How the higher frequencies are affected by air humidity and temperature is not exactly known, since the results in literature vary [79]. This change in high-frequency content is known not to be a very strong distance cue [79].

If a sound source is closer than approximately 3 meters, there is a curved wavefront instead of a planar wavefront [97]. This results in an increased loudness perception for lower frequencies, called *tone darkening* [79]. This tone darkening should indicate the approach of a sound source [95]. When the sound source moves at a distance of beyond 3 meters, the wavefront becomes more planar and the tone darkening effect can be neglected [97, 95]. At large distances, environmental effects like wind, ground cover, buildings etc. can influence the spectrum of a sound source, but it is shown that these phenomena are not a very strong cue for distance perception [79].

**Doppler shift**

The *Doppler shift*, which is a change in pitch of a moving sound source, provides information about whether the sound source is moving away (decreasing pitch over time) or coming closer (increase of pitch over time) to the listener [79]. For a good prediction of the distance, other cues are also needed.

**Environmental Context**

The environmental context is mostly determined by reverberation properties [31]. In a reverberant context, the sound that reach the ear is a mixture of the *direct sound*, which is the wavefront that reaches the ear first by a linear path, and the *reverberation*, which are all the wavefronts that reaches the ear indirectly reflected one or multiple times by surfaces in the environment. This perceived idea about the environment is called *environmental context* and it is characterized by its reverberations. Only in anechoic chambers or typical environments like large, open fields of snow-covered ground [79], there is no reverberation.

There are a couple of relevant physical parameters of reverberation that affect the perception of an environmental context [79]. The *volume or size*, which is cued by the reverberation time and level. The *absorptiveness* of the reflective surfaces, which are frequency dependent and vary for each envi-
ronmental context, makes it possible to categorize or compare environmental contexts. The *complexity* of the shape of the enclosure, which will shape the spatial distribution of reflections to the listener, particularly the early reflections.

All these properties can be used in order to model reverberations and include them by implementing their algorithms into a 3D audio system. Also, it is good to understand the effect of reverberations, either negative for sound localization performance, but positive for externalization, distance perception and to have an experience of the environmental context.

**Sound Source Width**

Familiarity is a very strong cue for the sound source width, because knowledge about the sound source implicates knowledge about the size of the sound source [79]. For unknown sound sources, reverberation plays an important role [79]. Sound source width is increased with increasing reverberation [79]. Direct sounds (with no reverberation) are perceived as being smaller with greater density and reverberant sounds are found to be larger, but also more diffuse [79].

### 2-2-2 How to measure spatial perception performance?

Sound quality measurements are often done using standards like ITU-R BS.1116 [98]. Such a standard prescribes how to ask persons to rate the overall sound quality and certain specific factors like fidelity, distortion, dynamic range and many more. The overall quality and underlying factors are usually assessed via a Mean Opinion Score (MOS) ranging from 1 meaning "very annoying" to 5 meaning "imperceptible". These measurements are commonly used to find impairments like sound distortions or perceived preferences for different audio codecs that compress or process an original sound fragment. Persons are then asked to compare the processed sound with the original one. Some standards state that they also support spatial properties for multi-channel sound systems, but there is still a lack of good measurements of the spatial 3D attributes for sound systems [99, 100]. A suggestion is done in [100] to cover different groups of spatial attributes that will be evaluated separately on top of other measurements already covered by the standards:

- **Width**
  - covering individual source width, ensemble width and environment width.

- **Depth and distance**
  - covering Individual source distance, ensemble distance, individual source depth, ensemble depth, environment depth and scene depth.

- **Envelopment, spaciousness, and spatial impression**
  - covering individual source envelopment, ensemble source envelopment, environmental envelopment, presence.

It is surprising to see that localization is not covered by these measurements, but [100] already mentions that a number of further attributes or characteristics may be important in the evaluation of spatial audio. Our suggestion would be to add localization to this scene-based paradigm. More studies and future research have to form good standards to measure real spatial 3D audio quality,
since the need is becoming increasingly higher with the further development of 3D virtual reality and 3D television. If we are able to measure the quality of 3D audio, we can also study the relevance and effect of the spatial 3D factor on phenomena like presence or VRET performance.

2-2-3 3D audio vs mono, stereo and dolby surround

Since 3D audio is the only technique that is able to recreate a listening experience that is “just as in real life”, other audio techniques like mono, stereo and dolby surround are always inferior to some degree. Mono for example is a one-channel audio technique that can only be heard “in the middle” using headphones, or in the centre of the loudspeaker array. The amount of externalization is none, reverberant properties can only be added in the centre on top of the sound source, so no real stereoscopic image can be created. Distance perception is on the other hand easily realized by adjusting the loudness level of the sound source, recreating a perception of a moving sound source in distance. Stereo is a dual-channel audio technique that can position a sound source either on the left, on the right, or everywhere in between either on the line of the two ears (with headphones), or on the line of a speaker array. Stereoscopic images can be created by adding reverberant properties and by “widening” the sound field. Distance information can also be added by adjusting the loudness level of a sound source. So in a sense, already some perception of movement, distance and location can be created, but the actual sound is still heard “inside the head” using headphones or on the line of the loudspeaker array. Dolby surround can create the perception of sound sources moving on the horizontal plane, outside the head, since original dolby surround is done through loudspeakers. But still, it lacks in elevation information. Dolby surround also has a dolby headphone variant which simulates a dolby surround sensation through headphones. 3D audio really contains the model of the human hearing and therefore actual positioning on the horizontal and vertical plane is possible, while also perceiving a real sense of distance because the sound source can be heard “outside the head”. But it is still unclear if these more accurate and realistic audio cues will contribute to a higher sense of presence and/or anxiety.

2-3 What is the best 3D audio reproduction technique for VRET

3D audio can be reproduced in many ways using different techniques and setups, but always using either headphones or speakers [31]. A literature study prior to this master thesis was done in order to see which 3D audio representation technique is most suitable for using in VRET systems [42]. The conclusion was: using headphones, because of its potential for commercial use regarding costs, practicability, flexibility and environmental independence. Speaker setups are much more impractical because of the complex speaker setups and an inevitable problem called “crosstalk”, causing a sweet spot, which limits the listener’s freedom to move and limits the amount of potential listeners, since the sweet spot is rather small. Also speaker setups are more expensive. Headphones have their own problems, mainly the problem of head movement interaction, which requires more complex techniques to allow the auditory environment to move interactively with the movement of the listeners head. And, the problem of non-individualized Head Related Transfer Functions (HRTFs), causing a decrease in accuracy of localization. But, the advantages are much more clear: headphones are very cheap, they are environmentally independent, they have no sweet spot and multiple listeners can listen at the same time using the right reproduction techniques [31]. So, the most obvious choice for 3D audio is using headphones.
Chapter 3

ABX perceptual difference listening test

The first banal question that needs to be answered before any useful interpretation can be done about the results of the preceding VRET experiments is:

Can people perceive a difference between 3D audio, dolby surround, stereo and mono with headphones?

If the answer to this question is "no", then the comparison between the different audio techniques in the main research question will be of little use since there is no perceptual difference and the variance of the outcome of the comparison, if there is any, must be caused by another effect than switching the dependent variable audio technique. Therefore, only the global effect of audio on presence, spatial perception, SUD score and heart rate can be researched, regardless of the audio technique.

3-1 Method

The experimental design is explained in subsection 3-1-1. The participants are described in subsection 3-1-2. An overview and explanation of all used equipment, software and the experimental room is given in subsection 3-1-3. Details about the stimuli are given in subsection 3-1-4 and finally the whole procedure is presented step by step in subsection 3-1-5.

3-1-1 Experimental design: ABX discrimination test

This experiment is an ABX discrimination test. An ABX test is a double blind method to compare two stimuli in order to find if they are perceptually different [101]. The procedure is as follows: Participants are presented three audio fragments, A (the reference), B (the sample) and X. A is never equal to B and X is always either A or B randomly chosen. The participants can listen and switch between
all three stimuli, as much and as long as they want, and have to determine if stimulus X is either A or B. This can be repeated for several trials.

The assumption is that if the choice is made as a gamble, the correct answer is given 50% of the times and therefore the difference between A and B cannot be heard. Increasing the number of trials increases the statistical power of this assumption. According to the protocol, a hypothesis $H$ and a null hypothesis $H_0$ have to be formulated and the null hypothesis can be rejected if a confidence interval of 95% has been reached [102]. Formulations of the hypotheses are:

$$H = "Stimulus \text{ A is perceptually different from B}"$$

$$H_0 = "The choice is made as a 50\% \text{ gamble, stimulus A is not perceptually different from B}"$$

If the $H_0$ can be rejected with a 95% confidence interval, it can be said that A is perceptually different from B with statistical significance. This would answer our research question if we do a separate ABX test for all audio technique combinations and selecting one audio technique as A and the other as B.

According to the QSC guidelines, formulated by the company that originally designed the ABX test, it is recommended that no more than 25 trials should be done per participant to avoid fatigue effects [103]. For ABX tests it is more common to use around 25 trials for each participant, and present the analysis of the result for each participant separately [102]. Since our research question is more focused on if "the average" listener as a group can hear perceptual differences between all audio technique combinations, a method is chosen to present every audio technique combination to each participant. Six audio technique combinations have to be compared, and so, each participant had in total $6 \times 4 = 24$ trials, 4 for each audio technique combination. In order to reduce carry-over effects, like ordering, fatigue, learning and confidence, a balanced Latin square was used. In a balanced Latin square, the serial effects are balanced out by arranging the standard Latin square [12] in such a way that each audio technique is equally preceded by every other audio technique equally often [104]. Since the amount of audio techniques is even, the procedure to arrange the balanced Latin square is:

The first row is obtained by:

$$0,1,n-1,2,n-2,3,n-3,\ldots,n/2$$

where $n$ is the amount of cases, functioning as an index for each case in the algorithm. The successive rows can be obtained by adding 1 to the immediately preceding row and filling in the corresponding case of that number. If the addition results in $n$, then take the case belonging to 0. The result can be seen in table 3-1.

The conceptual model is visualized in figure 3-1, which is a graphical representation of the dependent variable perceptual difference, which is either "yes" or "no", for changing the independent variable audio technique combination in a within-subject design.

The head movement is excluded from this experiment.

### 3-1-2 Participants

22 Participants, not suffering from total deafness in one or both ears, age ($M = 27.7$, $SD = 8.4$), mostly PhD, student or staff of the Delft University of Technology, but also general public participated in this experiment. Before the experiment started, the participants were asked if they had a hearing...
Participant nr. | Order of audio technique combinations
---|---
1 | M-S M-D D-3D M-3D S-3D S-D
2 | M-D M-3D M-S S-D D-3D S-3D
3 | M-3D S-D M-D S-3D M-S D-3D
4 | S-D S-3D M-3D M-D D-3D M-S
5 | S-3D D-3D S-D M-S M-3D M-D
6 | D-3D M-S S-3D M-D S-D M-3D
7 | M-S ... ... ... ... ...
22 | S-D S-3D M-3D D-3D M-D M-S

Table 3-1: Audio technique pairs ordered using balanced latin square [12]. (M = Mono, S = Stereo, D = Dolby Surround and 3D = 3D audio)

Audio technique combination

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual difference</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-1: Conceptual model of the ABX discrimination listening test in a Within Subject Design

impairment. One responded that he had a hearing capacity of 5% at the left ear. All 21 others answered "No", or "Not that I am aware of". For this experiment it was important that people could hear with both ears, since the 3D audio was realized by making use of "binaural" hearing techniques, requiring both ears in order to hear 3D audio. Deafness at one ear should result in a largely decreased effect of the 3D element, because in this case the interaural differences are absent and the HRTFs are affected. Background information about this matter is given in section 2-2-1. No preference was made for gender, age (although older people have a higher probability for suffering from hearing loss), educational or cultural background. Due to ethical concerns, no actual phobia patients were asked for this study.

3-1-3 Listening room, hardware & software

Since the ABX listening test was focussed solely on audio, a professional listening room was chosen for most optimal conditions. The experiment was done at a professional listening room at the Delft University of Technology at the SIPLAB [105]. The SIPLAB comes with high-quality audio equipment, among which Bowers & Wilkins Nautilus 800 loudspeakers and Mark Levinson equipment. The listening room was acoustically isolated from the outside world in order to reduce external background noises. Also the light conditions could be controlled. The experimental setup in the room was divided into two main areas: The experimenter control area and the participant area or experimental

Master of Science Thesis A.R.D. Hoekstra
area. Two desks were located in front of each other, with a large black isolating screen in between that separated the experimental room into two parts. On the experimenter control section, there was a Windows 7 computer with two 23 inch widescreen LCD monitors, a keyboard, a mouse and a chair. On the participant exposure part, there was a Beyerdynamic DT 770 headphone, a mouse, a 17 inch Iiyama LCD monitor that was initially turned off and two chairs. The Beyerdynamic DT 770 is a professional closed dynamic diffuse-field studio headphone with a frequency response of 5 - 35,000 Hz, an impedance of 250 Ohms with a cushion system that allows it to be used comfortably for long periods of time. The ambient noise reduction of the headphone is approximately 18 dB(A), so the ambient and external noises inside the experimental room are reduced by 18dB(A). It is assumable that crosstalk is negligible because of the acoustical properties of the room and the damping behaviour of the cushion system. The 17 inch LCD monitor and the mouse were used for the ABX test. The screen contained a clone of the second screen of the windows computer in the experimenter’s area. This was done to control the ABX test from the experimenter area. For the ABX test the software Foobar2000 together with the ABX Comparator component [106] were used in order to let the participant do the test autonomously. Initially, the ABX Comparator Component contained too much Graphical User Interface (GUI) elements, making the test less of the "standard" mentioned in literature [107, 103, 102, 101]. All unnecessary GUI parts were hidden and disabled for the participants using the software tool Screen Concealer for Windows [108], which is a tool that allows hiding and disabling certain parts of the computer screen using masks. So, a mask was created and placed on the Foobar2000 ABX Comparator component. The result before and after GUI masking is illustrated in figure 3-2. The ABX test doesn’t include a button "Y". So, the choice for Y was hidden, including other related parts. Also, there was no need for the control playing bar at the bottom, making the test unnecessary more complex. The statistics were also hidden from the GUI and instead of "Score: x/y" there was shown "Trial" in order to keep track of the amount of trials. The results of each ABX session were saved to a log file for later analysis.

![Figure 3-2: With (left) and without (right) masking the ABX Comparator component of Foobar2000 using the Screen Concealer tool.](image-url)
3-1-4 Stimuli

The function of the ABX test was to verify if the chosen stimuli for both VRET experiments "3D audio for VRET without visuals" and "3D audio for VRET with visuals" were perceptually different between the audio techniques. The stimuli used for the two preceding VRET experiments were chosen in a way that the stimuli should induce a universal sense of discomfort and/or fear, since all participants were known to not suffer from an anxiety disorder. In appendix D the method for choosing the appropriate stimuli is described. The analysis resulted in choosing for the sound of a flying wasp. A picture of a real-life wasp is given in figure 3-3. The wasp stimulus was found to be most suitable for both VRET experiments since it is known that a flying wasp near a person’s head creates a universal kind of irritation and/or discomfort and maybe even fear. Also, the 3D audio effect should be present since the wasp has both high and lower frequency sounds, which is known to be more easily located than sounds with only low or only certain higher frequencies. More information about human spatial hearing is given in section 2-2-1. The Root Mean Square (RMS) of the spectrum of the complete duration of the used flying wasp sound is given in figure 3-4 for a verification of the spectral sound properties. The wasp is also known to be able to fly around the listener’s head on a 3D sphere in space. Therefore, the 3D audio effect could probably be tested properly.

Figure 3-3: A copyright free stock photo of a wasp.

Section 2-3 suggested that the most ideal choice for 3D audio production for VRET is through headphones. Mono and Stereo can also be represented by headphones. For dolby surround, algorithms have been developed by Dolby, called dolby headphones, creating a dolby surround simulation through headphones. In order to keep the experimental setup as consistent as possible, headphones were used for all audio techniques. This way, the amount of comfort was the same, and the stimuli sound colour was more consistent because of the same reproduction device. It should be mentioned that the sound colour can be influenced by the 3D audio and dolby surround algorithms. Also, the current 3D audio reproduction techniques through headphones are still not able to realize real-time interactive head-tracking 3D audio in a convincing way. Either the spatial quality is partially lost, or the audio has a disturbing "lag", which is a delay between head movement input and the perceived change in the auditory virtual world. Hence, a trade-off between reducing the audio quality or excluding interactive head movements has to be made. The latter also means a decrease in spatial perception, since it is known that front-back reversal errors in spatial perception are mainly solved by moving the
head [31]. So, the stimuli have to be realized in the audio techniques mono, stereo, dolby surround and 3D audio, with headphones, without head movement.

![Image](image_url)

**Figure 3-4:** The RMS of the spectrum peaks of the used flying wasp sound using the Waves PAZ Frequency Analyser.

In order to realize these requirements, different audio recording techniques, audio tools, processing techniques and playback techniques were evaluated and the wasp stimuli were finally realized using a mono-recorded real-life flying wasp sound which was downloaded from an audio library website [109] and by placing it in the 3D world using the 3D audio tool SoundLocus [110]. SoundLocus recreates 3D audio effects using HRTFs, human hearing modelling and a small Doppler effect. Mono sounds can be loaded into the software and a flying / movement path can be created. It was also possible to export the flying wasp stimulus including its movement path to all the different audio techniques mono, stereo, dolby surround and 3D audio. More information about the sound selection, recording techniques, pre-processing, post-processing and realization of the stimuli can be found in the appendix in chapter D. The result was a 0.57 seconds flying wasp sound with a certain constant movement path for all different audio techniques. No other environmental sounds were added. Spectrograms of all audio techniques for the left and right channels are given in figure 3-5 to show how the intensity of all frequencies vary over time. From this figure it is clear that stereo varies the intensity of the whole frequency range of the sound quite equally, which is also the case in mono. Also sometimes the intensity becomes zero for all frequencies in one channel when the sound source is completely "panned" to the other. Dolby surround does some selective frequency manipulation, i.e. there is a small selective decrement around 800 Hz. Also, the whole signal is more smeared out, probably due to the added room simulation. 3D audio also does some selective frequency manipulation, i.e. especially the higher frequencies are decreased more compared to stereo when the sound source is moving further away from the observer. Of course for 3D audio there are interaural difference cues that cannot be seen in this figure.

### 3-1-5 Procedure

The ABX listening test was combined with the "3D audio for VRET without visuals" experiment described in chapter 4 for practical reasons. The flow of the ABX test is presented in figure 3-6.

The GUI of the ABX Comparator of the Foobar2000 audio software, which is illustrated in figure 3-2 right half, was presented to the participant on a black background. All Windows tool-bars or other possibly distracting elements were hidden from the screen. The concept of the ABX test was...
ABX perceptual difference listening test

<table>
<thead>
<tr>
<th>Stimulus / Response</th>
<th>Response &quot;A&quot;</th>
<th>Response &quot;B&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = A</td>
<td>Correct</td>
<td>Incorrect</td>
</tr>
<tr>
<td>X = B</td>
<td>Incorrect</td>
<td>Correct</td>
</tr>
</tbody>
</table>

Table 3-2: Response matrix for ABX Bernoulli trials

explained as follows: There are three stimuli A, B and X, where A is never equal to B and X is either equal to A or to B. All GUI control elements were explained. The participant could directly switch between all three stimuli A, B and X by pressing the play-buttons. When a play-button was pressed, the sound immediately switched to the other stimulus and it played from the exact same moment in time of where the other stimulus ended. So, it looked like the wasp sound kept on playing, while it switched between audio techniques. The participant could switch between the three stimuli A, B and X as often and as long as he or she wanted. When the stimulus ended after 0:57 seconds, the participant only had to press a play button again in order to restart the stimulus. After the instructions, a small training session was done prior to the real experiment. During this training the participants performed four comparisons of the mono versus stereo technique. These results were not saved. During the real experiment the participant did 4 trials of all 6 different audio technique combinations, resulting in a total of 24 trials. Each time a participant reached 4 trials of comparing two audio techniques, he or she gave the experimenter a sign for loading the next combination, until all 6 combinations were done.

3-2 Results

This section answers the research question "Can people perceive a difference between 3D audio, dolby surround, stereo and mono?".

Each trial of the ABX test can be seen as a Bernoulli trial, because each trial is a test in which the outcome can be classified in two mutually exclusive and exhaustive ways, either "Correct" or "Incorrect" [102]. The response matrix is visualized in table 3-2.

A common method is to use the Binomial Distribution [107] to determine the statistical significance of the test. Two assumptions are necessary to use the Binomial Distribution. The ABX software has to randomly distribute correct answers of X = A and X = B throughout the test and when a subject is unable to identify the correct answer, the response is random and uncorrelated to the audio tests [102]. The two hypotheses were formulated and mentioned in subsection 3-1-1:

\[ H = "Stimulus A is perceptually different from B" \]
\[ H_0 = "The choice is made as a 50% gamble, stimulus A is not perceptually different from B" \]

The \( p \)-value can be calculated using equation 3-1.

\[ P_{k \text{ out of } N} = \frac{N!}{k!(N-k)!} \left( p^k \right) \left( q^{N-k} \right) \]  

(3-1)

Where:
N = the number of trials;

Master of Science Thesis  
A.R.D. Hoekstra
Table 3-3: Response matrix for ABX Bernoulli trials

<table>
<thead>
<tr>
<th>Combination</th>
<th>Trials</th>
<th>Correct answers</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>mono - stereo</td>
<td>88</td>
<td>84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mono - dolby surround</td>
<td>88</td>
<td>86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mono - 3D audio</td>
<td>88</td>
<td>88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>stereo - dolby surround</td>
<td>88</td>
<td>88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>stereo - 3D audio</td>
<td>88</td>
<td>88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>dolby surround - 3D audio</td>
<td>88</td>
<td>88</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 3-3 presents the results of each audio technique combination. For all audio technique combinations, the $H_0$ can be rejected with a significant $p$ value of $< 0.001$. For almost all combinations the participants were able to give flawless answers, indicating an almost 100% recognition.

3-3 Conclusion

From the results presented in section 3-2, it is clear that for the given wasp stimuli, a significant perceptual difference was found for all audio technique combinations: mono - stereo, mono - dolby surround, mono - 3D audio, stereo - dolby surround, stereo - 3D audio and dolby surround - 3D audio.

k = the number of "Correct" answers;
p = the probability that a "Correct" answer is given;
q = the probability that an "Incorrect" answer is given.
Figure 3-5: Spectrograms of the wasp stimuli for all audio techniques for each channel (left and right).
Figure 3-6: Flow of the ABX listening test. The grey box with "3D audio for VRET no visuals" will be discussed in detail in chapter 4.
The main research question to be answered in this experiment is:

*Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET without using visuals?*

In this experiment, the focus lies solely on audio. The reason for this is that it is suggested by Begault [31] that if visuals are absent, audio plays a more important role in our perception, since vision is known to be the dominant sense, suppressing focus on audio.

Since a perceptual difference was heard for the flying wasp stimuli for all audio technique combinations, it can now be assumed that a change in effect on presence, anxiety and spatial perception may be because of the difference in audio techniques. The experiment described in this chapter was the first exposure of people to the feared stimuli. Mono was included as a "fear for the stimulus" reference baseline. If a person already feels fear by listening to a mono sound, the fear is probably caused by the sound itself, rather than the spatial properties which are more present in stereo and even more in dolby surround and most in 3D audio.

The method of the experiment is explained in section 4-1. The results are presented in section 4-2 and the conclusion, discussion and future research recommendations are presented in section 4-3.

### 4-1 Method

This section contains the description of all methodological aspects for the experiment "3D audio for VRET without visuals". The Experimental design is explained in subsection 4-1-1. The participants are described in subsection 4-1-2. The used stimuli are discussed in subsection 4-1-3. An overview and explanation of all used equipment, software and the experimental room is given in subsection 4-1-4. The used measurement instruments are discussed in subsection 4-1-5 and finally the whole procedure is presented step by step in subsection 4-1-6.
4-1-1 Experimental design: Within-subject design

A Within-subject design, also called a Repeated Measure design, was chosen for this experiment. The conceptual model is visualized in figure 4-1.

![Conceptual model of the experiment "3D audio for VRET without visuals": Within-subject design](image)

The effect of one independent variable audio technique, with the four audio techniques being mono, stereo, dolby surround and 3D audio, on four separate dependent variables Presence, Spatial perception, SUD Score and Heart rate was studied.

A within-subject design was chosen because it has more statistical power compared to a between-subject design, making it less resource heavy. The main disadvantage for using a within-subject design is that there are certain carry-over effects, like ordering, fatigue, learning, confidence, but also the previously mentioned decreasing effect of fear and discomfort [104]. To reduce these effects, a balanced Latin square was used, so each audio technique was equally preceded by every other audio technique equally often [104]. The amount of audio techniques is even, so the standard algorithm could be used, which was explained in section 3-1-1. The result is presented in table 4-1.

<table>
<thead>
<tr>
<th>Participant nr.</th>
<th>Order of audio techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mono Stereo 3D audio Dolby Surround</td>
</tr>
<tr>
<td>2</td>
<td>Stereo Dolby Surround Mono 3D audio</td>
</tr>
<tr>
<td>3</td>
<td>Dolby Surround 3D audio Stereo Mono</td>
</tr>
<tr>
<td>4</td>
<td>3D audio Mono Dolby Surround Stereo</td>
</tr>
<tr>
<td>5</td>
<td>Mono Stereo 3D audio Dolby Surround</td>
</tr>
<tr>
<td>6</td>
<td>Stereo ... ... ...</td>
</tr>
<tr>
<td>22</td>
<td>Stereo Dolby Surround Mono 3D audio</td>
</tr>
</tbody>
</table>

*Table 4-1: Audio techniques ordered using balanced latin square [12]*

Master of Science Thesis
A.R.D. Hoekstra
4-1-2 Participants

The exact same group of participants was used as for the ABX listening test, described in section 3-1-2.

4-1-3 Stimuli

The exact same stimuli of the flying wasp were used as for the ABX listening test, described in section 3-1-4.

4-1-4 Listening room, hardware & software

This experiment was done in the same listening room of the ABX listening test, so all properties of the acoustically isolated SIPLAB listening room were already mentioned in section 3-1-3. The used software and equipment was different. The experimental room was divided into two sections: The experimenter control area and the participant area or experimental area. Two desks were located in front of each other, with a large black isolating screen in between that separated the experimental room into two parts. On the experimenter control part, there was a Windows 7 computer with two 23 inch widescreen LCD monitors, a keyboard, a mouse and a chair. On the participant exposure part, there was a Beyerdynamic DT 770 headphone, a Mobi8 hardware device with heart rate and galvanic skin conductance sensors, a normal black eye-mask, a pile of books, a pencil, two chairs. The light conditions of the listening room could be controlled. The stimuli were played using a Windows 7 computer having the Foobar2000 audio player installed. The heart rate and galvanic skin response were measured using a Mobi8, although the galvanic skin conductance sensors were not working properly due to a defective cable. The measurements were logged by the VRET software tool DRVRET, developed by the Delft University of Technology [111]. Also, "flags" were placed when an exposure started or ended for later analysis. DRVRET is capable of creating a VR world in order to treat people with certain anxiety disorders. It also has a GUI for the therapist where he/she can control certain exposure properties, measure physiological properties of the patient, enter SUD scores, monitor the patient over distance through an internet connection and using a web-cam and place certain "flags" in time with corresponding labels such as "tense" or "relaxed". A standard black eye-mask was used to blindfold the participants and a chin-rest made of books with a cushion was used in order to keep their head on a fixed position and orientation. During a pilot test, different chin-rests were evaluated for head fixation, controllability and comfort. A normal chin-rest that is usually being used for visual experiments demanding a fixed head and eye orientation was evaluated together with a chin rest made of a pile of books. Participants found the pile of books much more comfortable and less distracting. The height of the pile of books could be easily adjusted by adding or removing small books. Also, the head remained fixed during the exposures of the pilot test, so no distractions were present of moving the head while the VR world remained fixed.

4-1-5 Measurement instruments

During this experiment, several measurement instruments were used. Some were physiological properties measured by hardware devices, but most were done using questionnaires. The questionnaires were filled in using pen-and-paper. All questionnaires were presented in the English language, since most participants had difficulties with the Dutch language.
Descriptive questionnaire

A descriptive questionnaire was used in order to get some global information about the participant and to measure some properties that are important for this experiment. The first question, which is crucial for determining the value of the results, was "Do you suffer from any kind of known hearing impairment? If yes, please describe". If people suffer from an hearing impairment affecting one or both ears, like partial deafness, the 3D audio perception is also affected, since 3D audio requires both ears. Also two questions were asked that indicate a global fear towards wasps: "Do you have a fear of wasps?" and "How scared are you to do this experiment?". Finally some descriptive questions were asked about the age, educational level and gender.

Fear of Wasps questionnaire

Prior to the experiment participants were asked to fill in the "Fear of Wasps questionnaire". This questionnaire was meant to give a measure of how fearful people are towards wasps and if a person has a real phobia towards wasps. The Fear of Wasps Questionnaire was created by adapting and combining two questionnaires: Fear of Spiders Questionnaire (FSQ) [112] and the Spider Phobia Questionnaire (SPQ) [113]. These two questionnaires are commonly used in order to check if a person has a phobia towards spiders. Both questionnaires have their own weaknesses and strengths, so for the Fear of Wasps questionnaire, both questionnaires were used as a basis. An adaptation of these questionnaires is more commonly used in research regarding phobia for small animals, for example in treating people with a phobia for cockroaches [114]. All questions were revised by changing the subject, wasps instead of spiders, and also refining the sentence if needed. For example the question "If I saw a spider now, I think it will try to jump on me." was adapted to "If I saw a wasp now, I think it will try to land on me.". Most questions only needed the substitution of "spider" by "wasp". The result was a questionnaire of 61 questions; the first 17 questions had to be scored on a 7-point Likert scale where 1 = "Totally disagree" and 7 = "Totally agree" and the remaining had to be answered with "yes" or "no". The complete questionnaire can be found in the appendix C.

IGroup Presence Questionnaire

There are multiple known questionnaires to measure presence. For example, the Presence Questionnaire (PQ) by Witmer and Singer [64], the ITC sense of presence inventory [65], the Slater-Usoh-Steed Questionnaire (SUS) [66] or the IGroup Presence Questionnaire (IPQ) [67]. In this experiment, the IPQ questionnaire was used. IPQ is a 14-item questionnaire with questions about presence that are easy to understand. In order to get an overall sense of presence, all 14 items could be used. The questionnaire can also be decomposed into four different sub-scales of presence: Spatial Presence, Involvement, Experienced Realism and Global Presence. Spatial Presence is the sense of being physically present in the VE, involvement is the attention devoted to the VE and the involvement experienced. Experienced Realism is the subjective experience of realism in the VE and Global Presence is the general "sense of being there". This decomposition can give more insight into multiple detailed aspects of the global concept "presence". The subdivision of the questions according to the four sub-groups is illustrated in figure 4-2. The 14 questions are rated on a 7-point Likert scale from -3 to 3. The scale is also sometimes reversed. The IPQ has a good consistency and is used in many presence related research [115, 116, 9, 117]. Therefore, it was found suitable for this experiment.

Master of Science Thesis A.R.D. Hoekstra
The commonly used audio quality evaluation methods mostly concern non-spatial attributes like timbre, sound colouration or distortion. However, the increasing use of multi-channel and 3D audio has recently motivated the study of spatial attributes as a part of audio quality evaluation [118, 119, 120, 121]. Based on literature defining measures for 3D audio spatial attributes [118, 119, 120, 121], the following eight common spatial attributes were selected for the Spatial Perception Questionnaire: localization, distance/depth, externalization, movement, immersion, sense of space, realism and quality. Some spatial attributes had commonly used synonyms. These synonyms were used as separate questions, all covering the main topic, in order to get more consistency and insight about an attribute. For example, localization was covered by the two attributes direction and localization, immersion by the attributes envelopment and immersion, realism by the attributes representation of the environment, naturalness, and finally, realism and distance / depth by the attributes distance and depth. In total, 13 questions were formulated, one for each (sub) attribute with a 7-point Likert rating scale. The questionnaire, including the attributes, formulated questions and scales are presented in table 4-2.

The Spatial Perception Questionnaire is only trying to get an overall measure of the perceptual strength of a spatial attribute in the perceived stimuli. It is not a precision-test where participants have to tell or point out the exact location or movement of the sound source.

**Immersive Tendency Questionnaire**

The Immersive Tendency Questionnaire (ITQ) is a questionnaire that measures one’s sensibility of being captivated by an immersive VE. A person that has a high score in ITQ, probably will experience a higher sense of presence more easily. This questionnaire was used in order to screen the participants later on how sensitive they were for experiencing presence. The complete ITQ can be seen in the appendix C.
<table>
<thead>
<tr>
<th>Nr.</th>
<th>Attribute</th>
<th>Question</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>envelopment</td>
<td>The sonic environment sounds ... enveloping.</td>
<td>not - very</td>
</tr>
<tr>
<td>2</td>
<td>immersion</td>
<td>I feel ... immersed in the sonic environment</td>
<td>not - very</td>
</tr>
<tr>
<td>3</td>
<td>representation of the environment</td>
<td>The representation of the sonic environment was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>4</td>
<td>externalization</td>
<td>The sense of externalization of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>5</td>
<td>Localization</td>
<td>The sense of localization was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>6</td>
<td>distance</td>
<td>The sense of distance of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>7</td>
<td>movement</td>
<td>The sense of movement of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>8</td>
<td>directions</td>
<td>The sense of directions of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>9</td>
<td>depth</td>
<td>The sense of depth of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>10</td>
<td>sense of space</td>
<td>The sense of space of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>11</td>
<td>realism</td>
<td>The realism of the scene was ...</td>
<td>poor - good</td>
</tr>
<tr>
<td>12</td>
<td>naturalness</td>
<td>The naturalness, true to life, was ...</td>
<td>Not truthful - Truthful</td>
</tr>
<tr>
<td>13</td>
<td>quality</td>
<td>The quality of the reproduction was ...</td>
<td>poor - good</td>
</tr>
</tbody>
</table>

Table 4-2: Spatial Perception questionnaire attributes.
SUD Score

The Subjective Units of Discomfort (SUD) score measures an overall sense of anxiety and/or discomfort. Participants were asked to rate the SUD score on a scale from 0 "completely relaxed" to 10 "panic attack". The SUD score was asked directly after each stimulus.

Heart rate

The physiological measurement heart rate of each participant was measured during the whole phase using the Mobi8 system from TMSi with an Xpod Oximeter. The participant inserted a finger into an adult articulated finger clip sensor. The Mobi8 sent the average heart rate to the VRET system. Updates of the heart rate were irregular, about once every 2 seconds. A log file was created by the VRET system and flags were placed in order to get a chronological flow which was logical for analysis of the resulting log file.

General comments

After having filled in all questionnaires belonging to an exposure, the participant was asked to "speak up freely" if he or she had some remarkable points to mention about anything that popped up in their mind after the exposure. These comments were noted. Also, at the end of the phase the participants were asked two additional open questions: "Did you have difficulties answering all questions due to fatigue effects or other external effects like discomfort due to the experimental setup?" and "Do you have any additional remarks after this phase?". Responses to these questions were noted. This was done since this experiment was also functioning as a pilot-test for the experiment "3D audio for VRET with visuals".

4-1-6 Procedure

The flow and structure of the experiment is visualized in figure 4-3. During a pilot-test with two participants - these results were not included in the dataset - it was clear that the experiment took about 1 hour and 10 minutes in total. Prior to the experiment, participants were provided with the experimental pack, containing a description of the experiment, which had to be signed when the experiment took place. The experimental pack contained a consent form that was approved by the University ethics committee. Also, each participant was asked to fill in the Fear of Wasps Questionnaire online. If the participant could speak Dutch fluently, the verbal communication between the experimenter and the participant was in Dutch.

Intro

The participant was welcomed and asked to sign the experimental pack when all content was read, clear and accepted. An instruction was given about the experiment and if all was clear, the participant was asked to fill in the descriptive questionnaire.
Figure 4-3: Flow of the "3D audio for VRET without visuals" experiment. The grey box "ABX test" was described in chapter 3.

Baseline

Before the baseline measurement, the chin rest was adjusted to its optimal height and the Mobi8 sensors, the headphones and the blind fold were mounted. If all apparatus were mounted correctly and the participant was feeling comfortable, the baseline measurement of "total silence" was started. After 5:00 minutes, the baseline measurement was done and the eye-mask and headphones were unmounted. For the first time the SUD score was asked. The experimenter placed flags at the beginning and ending of the baseline for later analysis.

Exposure 1, 2, 3 and 4

The real experiment started. The eye-mask and headphones were again mounted on the participant’s head and the participant was asked if he or she felt comfortably. If so, the experimenter told the participant that he or she was about to hear a sound of a wasp in 5 seconds if the experimenter says "in 5 seconds". This was because flags could be appropriately placed on accurate time intervals in order to be able to do the data analysis more easily. The sound was played for 0:57 seconds and when the sound ended, the experimenter placed another flag that the exposure was ended. The headphones and eye-mask were again unmounted and the chin rest was moved to the side. The participant was asked for the SUD score. After the SUD score was noted, both the IPQ presence questionnaire and the Spatial Perception Questionnaire were given to the participant. The participant was instructed to start with the IPQ presence questionnaire. This was done because presence was of higher importance and...
presence was measured by doing a "memory task". The experimenter pointed out that the participant should keep good notice on the scale of the IPQ, since it is sometimes reversed. Also, if the participant had any questions, he or she could ask them at any time. The experimenter left the participant alone filling in the questionnaires. When the participant completed both questionnaires, the next exposure followed, until all four stimuli of the wasp in different audio techniques were presented.

4-2 Results

There was one participant where all questionnaires were not labelled and therefore it was impossible to discover which results belonged to which audio technique. Therefore, this participant was excluded from the dataset, resulting in a complete dataset of 21 cases. In this section an answer is given for the research question Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET without using visuals?

Box plots, histograms and Q-Q plots of the data are given in the appendix A. Non-parametric tests were conducted for data that significantly differed from a normal distribution. Parametric tests were conducted for data with a resemblance of a normal distribution by looking at the histograms and Q-Q plots and by doing Shapiro-Wilk normality tests.

4-2-1 The effect of 3D audio on presence without visuals

Data preparation & reliability

The IPQ, explained in section 4-1-5, consists of 14 different items, grouped into 4 sub-scales. The sub-scales spatial presence, involvement, experienced realism and global presence were acquired by taking the mean of questions 6, 9, 10, 13; 1, 7, 11, 14; 2, 4, 5; and 12 respectively [67]. The IPQ questions 2, 11 and 13 had to be reversed in scale [67]. The overall sense of presence was acquired by calculating the mean of all 14 questions [122].

A reliability test of the overall sense of presence for each audio technique separately resulted in Cronbach’s alpha levels between 0.89 and 0.96, which is higher than 0.7 and therefore an acceptable level of reliability [123]. A reliability test for the four sub-scales of spatial presence, involvement, experienced realism and global presence for each audio technique separately resulted in Cronbach’s alpha levels between 0.77 to 0.93, except for the experienced realism of dolby surround, having a Cronbach’s alpha level of 0.59. Since all but one factors had a Cronbach’s alpha level higher than 0.7, the sub-scales were found to be of an acceptable reliability level [123].

Statistical analysis

By looking at the box plots in figure 4-4, there seems to be a trend that the scored values of 3D audio on presence and presence sub-scales are in all cases higher than for all other audio techniques.

There was a statistically significant difference in the overall sense of presence depending on audio
technique whilst running a Friedman test, $\chi^2(3) = 12.26$, $p = 0.007$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.009. Median ($Q_1$ to $Q_3$) experienced presence for the audio techniques mono, stereo, dolby surround and 3D audio were -0.86 (-1.61 to 0.32), 0.29 (-0.46 to 1.11), 0.71 (-0.57 to 1.96) and 1.29 (0.18 to 1.71), respectively. There was a significant difference between 3D audio and dolby surround ($Z = -2.90$, $p = 0.004$), and between 3D audio and mono ($Z = -3.51$, $p < 0.001$). No significant difference was found between 3D audio and stereo ($Z = -1.61$, $p = 0.107$).

There was also a significant difference between stereo and mono ($Z = -2.71$, $p = 0.007$) and between dolby surround and stereo ($Z = -2.67$, $p = 0.008$). No significant difference was found between stereo and dolby surround ($Z = -0.070$, $p = 0.948$).

A Friedman test with post-hoc Wilcoxon Signed-Rank tests and a Dunn-Sidak corrected alpha level of 0.009 was also done for all four IPQ sub-scales separately, see table 4-3. A significant difference was found between 3D audio and stereo on involvement ($Z = -2.595$, $p = 0.009$). Significant differences were found on all subscales between 3D audio and mono. Surprisingly, no significant difference was found between 3D audio and dolby surround on the sub-scales, while the overall presence was significantly different.

### 4-2-2 Effect of audio technique on spatial perception

#### Data preparation & reliability

The Spatial Perception Questionnaire (SPQ) had a total of 13 items, described in section 4-1-5. The SPQ had some items overlapping with the IPQ, like "immersion", "realism" and "naturalness". These items were significantly correlated near 0.8. An alternative measure was gathered by excluding these overlapping items to get a SPQ that is more focussed on the remaining distinctive spatial attributes. This "stripped" version of the Spatial Perception Questionnaire will be called the Stripped Spatial Perception Questionnaire (SSPQ).

A reliability test of the SPQ and the SSPQ for all audio techniques separately resulted in Cronbach’s alpha levels between 0.93 and 0.97, which is higher than 0.7 and therefore an acceptable level of reliability [123].

#### Statistical analysis

By looking at the box plots in figure 4-5, there seems to be a trend that the scored values of 3D audio on spatial perception is higher than for all the other audio techniques, although stereo follows closely.

For the SPQ data, there was a statistically significant difference in the spatial perception depending on audio technique whilst running a Friedman test, $\chi^2(3) = 23.26$, $p < 0.001$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in an alpha level of 0.009. Median ($Q_1$ to $Q_3$) spatial perception for the audio techniques mono, stereo, dolby surround and 3D audio were -1.23 (-1.77 to 0.46), 1.31 (-0.15 to 1.73), 1.23 (0.42 to 1.50) and 1.39 (0.65 to 1.92), respectively. There was a significant difference between 3D audio and mono ($Z = -3.81$, $p < 0.001$), but there was no significant difference between 3D audio and stereo ($Z = -1.22$, $p = 0.223$) and between 3D audio and dolby surround ($Z = -2.26$, $p = 0.024$) for $\alpha = 0.009$. 

Master of Science Thesis

A.R.D. Hoekstra
**Figure 4-4:** Box plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for mono, stereo, dolby surround and 3D audio.
### Sense of presence (14 IPQ attributes together)

<table>
<thead>
<tr>
<th>Overall differences</th>
<th>$\chi^2(3) = 12.26$</th>
<th>$p = 0.007$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combination</strong></td>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
</tr>
<tr>
<td>3D audio vs. mono</td>
<td>-3.51</td>
<td>$&lt; 0.001^*$</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-1.61</td>
<td>0.107</td>
</tr>
<tr>
<td>3D audio vs. dolby surround</td>
<td>-2.90</td>
<td>0.004*</td>
</tr>
<tr>
<td>stereo vs. mono</td>
<td>-2.71</td>
<td>0.007*</td>
</tr>
<tr>
<td>stereo vs. dolby surround</td>
<td>-0.070</td>
<td>0.948</td>
</tr>
<tr>
<td>dolby surround vs. mono</td>
<td>-2.67</td>
<td>0.008*</td>
</tr>
</tbody>
</table>

### Spatial Presence

<table>
<thead>
<tr>
<th>Overall differences</th>
<th>$\chi^2(3) = 16.47$</th>
<th>$p = 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combination</strong></td>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
</tr>
<tr>
<td>3D audio vs. mono</td>
<td>-3.53</td>
<td>$&lt; 0.001^*$</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-1.29</td>
<td>0.197</td>
</tr>
<tr>
<td>3D audio vs. dolby surround</td>
<td>-2.56</td>
<td>0.010</td>
</tr>
</tbody>
</table>

### Involvement

<table>
<thead>
<tr>
<th>Overall differences</th>
<th>$\chi^2(3) = 10.79$</th>
<th>$p = 0.013$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combination</strong></td>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
</tr>
<tr>
<td>3D audio vs. mono</td>
<td>-2.99</td>
<td>0.003*</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-2.60</td>
<td>0.009*</td>
</tr>
<tr>
<td>3D audio vs. dolby surround</td>
<td>-2.340</td>
<td>0.019</td>
</tr>
</tbody>
</table>

### Experienced realism

<table>
<thead>
<tr>
<th>Overall differences</th>
<th>$\chi^2(3) = 19.08$</th>
<th>$p &lt; 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combination</strong></td>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
</tr>
<tr>
<td>3D audio vs. mono</td>
<td>-3.76</td>
<td>$&lt; 0.001^*$</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-1.43</td>
<td>0.152</td>
</tr>
<tr>
<td>3D audio vs. dolby surround</td>
<td>-2.32</td>
<td>0.020</td>
</tr>
</tbody>
</table>

### Global presence

<table>
<thead>
<tr>
<th>Overall differences</th>
<th>$\chi^2(3) = 14.95$</th>
<th>$p = 0.002$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combination</strong></td>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
</tr>
<tr>
<td>3D audio vs. mono</td>
<td>-3.07</td>
<td>0.002*</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-1.27</td>
<td>0.205</td>
</tr>
<tr>
<td>3D audio vs. dolby surround</td>
<td>-1.62</td>
<td>0.105</td>
</tr>
</tbody>
</table>

*Table 4-3:* Summary of Friedman test results regarding the effect of all audio techniques on overall presence and its sub-scales for "3D audio for VRET without visuals". Significant results are marked with an asterisk *. 

---

Master of Science Thesis

A.R.D. Hoekstra
There was also a significant difference between mono and stereo (Z = -3.25, p = 0.001) and between mono and dolby surround (Z = -2.84, p = 0.005). No significant difference was found between stereo and dolby surround (Z = -0.26, p = 0.794).

For the SSPQ data, there was a statistically significant difference in the spatial perception depending on audio technique whilst running a Friedman test, $\chi^2(3) = 19.75, p < 0.001$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in an alpha level of 0.009. Median ($Q_1$ to $Q_3$) spatial perception for the audio techniques mono, stereo, dolby surround and 3D audio were -0.9 (-1.7 to 0.7), 1.3 (0.1 to 2), 1 (0.6 to 1.55) and 1.6 (0.8 to 1.9), respectively. There was a significant difference between 3D audio and mono (Z = -3.74, p < 0.001), but no significant difference was found between 3D audio and stereo (Z = -1.29, p = 0.198) and between 3D audio and dolby surround (Z = -2.11, p = 0.035) for $\alpha = 0.009$.

There was also a significant difference between mono and stereo (Z = -3.27, p = 0.001) and between mono and dolby surround (Z = -2.67, p = 0.007). No significant difference was found between stereo and dolby surround (Z = -0.4, p = 0.689).

When we compare the results of the SPQ and the SSPQ questionnaire, we find the same conclusions.

### 4-2-3 Effect of audio technique on SUD score

**Data preparation & reliability**

The SUD score was scored from 0 to 10 and did not have to be processed before the analysis.

**Statistical Analysis**

By looking at the box plot in figure 4-6, there seems to be a trend that the scores of 3D audio on SUD are higher than for all the other audio techniques, although the differences among the audio techniques are not very large.

A Friedman test was done by comparing the SUD score of all audio techniques with the baseline.
measurement in order to verify if the exposure itself caused any significant increment in SUD score.

There was a statistically significant difference in SUD score depending on audio technique whilst running a Friedman test, $\chi^2(3) = 31.44, p < 0.001$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.005. Median ($Q_1$ to $Q_3$) SUD score baseline, mono, stereo, dolby surround and 3D audio were 1 (0.5 to 2.5), 2 (1 to 4.5), 3 (1.5 to 4), 3 (1.5 to 3) and 4 (2 to 5), respectively. There was a significant difference between baseline and stereo ($Z = -3.18, p = 0.001$), between baseline and dolby surround ($Z = -3.06, p = 0.002$), between baseline and 3D audio ($Z = -3.75, p < 0.001$) and between baseline and mono ($Z = -2.91, p = 0.004$).

So sound could elicit anxiety.

In order to validate the stimuli as fearful stimuli, the participants were divided into two groups for which "fear for wasps" was "yes" or "no". By looking at the histogram presented in figure 4-7, participants with rated SUD score from 0 to 2 were categorized as "no", and participants with a rated SUD score of 3 or higher were categorized as "yes". There was a statistically significant difference between the median of the SUD Score of the two groups for mono ($Z = -2.25, p = 0.025$), stereo ($Z = -2.06, p = 0.039$), dolby surround ($Z = -2.22, p = 0.027$) and 3D audio ($Z = -2.00, p = 0.046$) by running a Mann-Whitney test. So, the chosen wasp stimulus was able to elicit more fear with participants who were more afraid of wasps than with participants who were less afraid.

In order to check if there was a significant difference between the audio techniques, another Friedman test was done.

There was a statistically significant difference in the SUD score depending on audio technique whilst running a Friedman test, $\chi^2(3) = 12.06, p = 0.007$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.009. Median ($Q_1$ to $Q_3$) SUD score for the audio techniques mono, stereo, dolby surround and 3D audio were 2 (1 to 4.5), 3 (1.5 to 4), 3 (1.5 to 3) and 4 (2 to 5), respectively. There was a significant difference between 3D audio and dolby surround ($Z = -3.09, p = 0.002$). No significant difference was found between 3D audio and mono ($Z = -2.29, p = 0.022$) for $\alpha = 0.009$. Also, no significant difference was found between 3D audio and stereo ($Z = -1.89, p = 0.058$). Furthermore, no significant difference was found between mono and stereo ($Z = -0.82, p = 0.410$), between mono and dolby surround ($Z = -0.48, p = 0.634$) and between stereo and dolby surround ($Z = -0.62, p = 0.537$).

### 4-2-4 Effect of audio technique on heart rate

#### Data preparation & reliability

The heart rate was measured on beats per minute (BPM) on different intervals in time and written to a log file. The heart rate means were calculated using the Log Analysis Tool [124], developed by the TU Delft. For the heart rate baseline, only the last 2 minutes were taken.
Statistical Analysis

By looking at the box plot in figure 4-8, the values of baseline, mono and dolby surround seem slightly higher than stereo and 3D audio, but all differences seem rather small.

When using an ANOVA with Repeated Measures with sphericity assumed, the mean heart rate was not statistically significantly different between audio techniques ($F(3, 60) = 2.41, P = 0.076$).

4-2-5 Correlations

Presence vs SUD Score

Presence and SUD score were not significantly correlated for mono ($r(19) = 0.41, p = 0.07$), stereo ($r(19) = 0.37, p = 0.1$), dolby surround ($r(19) = 0.35, p = 0.126$) and 3D audio ($r(19) = 0.25, p = 0.270$).
From the results we can conclude that audio can elicit anxiety, measured by SUD score. No significant differences on heart rate were found.

We can also conclude that for VRET with audio only, 3D audio has a larger effect on overall presence compared to mono and dolby surround. Compared to stereo, the overall presence was not significantly higher, but the presence sub-scale "involvement" was. For SUD score, a significantly larger effect of 3D audio was found compared to dolby surround but not compared to mono and stereo. For spatial perception, 3D audio only had a significant larger effect compared to mono, but no significant differences were found compared to stereo and dolby surround. No significant differences were found between audio techniques on heart rate, although the p-value was quite low. In table 4-4 an overview is given of the effect of 3D audio on all dependent variables compared to all audio techniques.

Table 4-4: Significant larger effects for $\alpha = 0.009$ of 3D audio on presence, SUD score, heart rate and spatial perception compared to mono, stereo and dolby headphones

Although no significant differences were found for the effect of 3D audio on SUD score compared to mono and stereo, their p-values were quite small and the median ($Q_1$ to $Q_3$) differed, indicating a trend for a higher SUD score for 3D audio. Also, for presence the median ($Q_1$ to $Q_3$) differed compared to stereo and 3D audio, indicating a trend for a higher presence value for 3D audio.

It was surprising that dolby headphones performed worse than stereo. This is probably a consequence of the algorithms used to "simulate" a sense of dolby surround, while dolby surround is actually reconstructed using multiple loudspeakers. Also, stereo is the most commonly used audio technique, so
it is possible that people are more "used" to this type of audio technique and therefore preferring it more than dolby headphones.

Another result was that no significant difference in spatial perception was found between 3D audio and stereo or dolby surround, although the median score indicated a trend in favour of 3D audio. This is possibly because both stereo and dolby surround provide some of the important information cues about localization for azimuth and for distance perception. So, even while every movement path of all different techniques was the same, people probably perceived them differently but experienced them as "just as good". The spatial perception questionnaire was no real pin-point measurement tool trying to find the exact location of the wasp. Instead, it tried to measure a sense of spatial perception as an experience factor. For the 3D audio technique, the perception of the movement path of the wasp was probably the most accurate, but this was not the aim of this experiment.

In the stimuli, only the wasp sound was used. No environmental context or reverberation properties were added. This is only realistic when the wasp is inside a reverberant dead listening room, which is a very uncommon case. But, the experiment was done inside a listening room with extremely little reverberation. The participants were also told that they could let their imagination run free, but if they have difficulties imagining a virtual world represented by the audio fragment (because there was none), they could imagine if the wasp was existing in the experimental room in which they were in. But since this is an uncommon scenario, the experience could be affected by the fact that there was no environmental context. But the absence of environmental context was a constant factor for all audio conditions. Only the dolby surround added room simulation to the sound scene, which was an unwanted effect, but impossible to disable from the dolby headphones emulator, because the horizontal sound source positioning and externalization outside one’s head was achieved partly because of these reverberation properties.
From the ABX test it has become clear that there is a perceptual difference between 3D audio and the audio techniques mono, stereo and dolby surround for the given wasp stimuli. Also, without using visuals, 3D audio has a significantly larger effect on presence and SUD Score compared to mono and dolby surround. Differences between stereo and 3D audio were less clear. However, since stereo is the most commonly used audio technique, the focus of this experiment lies on the comparison of stereo and 3D audio. Also, the condition no audio has been included in order to test the addition of sound itself in a VRET environment. Dolby surround was excluded because the simulation algorithms used for the headphones were doubtful. Mono was also excluded in order to focus on the more subtle differences between stereo and 3D audio. This chapter gives an answer to our final research question:

Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET with using visuals?

First, the method of the experimental setup is explained in section 5-1 and the results are presented in section 5-2, followed by a conclusion and discussion in section 5-3.

5-1 Method

This section contains all experimental details related to the experiment "3D audio for VRET with visuals". The experimental design is explained in subsection 5-1-1. The participants are described in subsection 5-1-2. The stimuli used are discussed in subsection 5-1-3. An overview and explanation of all equipment, software and experimental room used is given in subsection 5-1-4. The measurement instruments used are discussed in subsection 5-1-5 and finally the whole procedure is presented step by step in subsection 5-1-6.

5-1-1 Experimental design

A Within-subject design, also called a Repeated Measure design, was chosen for this experiment. The conceptual model is visualized in figure 5-1. The effect of one independent variable audio condition,
with the three audio conditions no audio, stereo and 3D audio, on five separate dependent variables Presence, Spatial perception, SUD Score, Heart rate and Galvanic skin response was studied.

Also for this experiment a within-subject design was chosen to make it less resource heavy. To reduce the carry-over effects, a Balanced Latin square was used for the audio conditions. Since there was an odd number of three audio conditions, two Balanced Latin squares were needed in order to avoid carry-over effects. The first was created exactly the same way as described in section 4-1-1 and the second one is a mirror image of the first. The result is presented in table 5-1

In order to avoid expectation during the different exposures, each stimulus was constructed in a way that the movement path of the wasp consisted of four different scenarios: Far away in front of the observer, close in front of the observer landing on left ear, close in front of the observer landing on right ear and wasp sitting and walking on the table. These scenarios were also counter balanced using a balanced Latin square, see table 5-2. The final result of the ordering is presented in table 5-3.
Table 5-3: Audio conditions and stimuli ordered using a balanced Latin square [12].

5-1-2 Participants

Since the participants of the previous experiments were already exposed by the flying wasp stimuli quite a bit, a new group of participants was used for this experiment. The group consisted of 25 non-patient participants. Most of them were PhD, post-doc, staff or student, while the rest was general public, with an average age of 28 ($SD = 8.2$) and 16 were men. One participant was suffering from a known hearing impairment "30dB loss on both ears". This person was not excluded from the dataset since the impairment was in both ears, only in global loudness level, so a sense of 3D should be present. This participant was no outlier. Three other participants mentioned a small hearing impairment "Left ear a little noise", "I have problems hearing directions of the sound" and "I'm a little hard of hearing", which should not have a large effect on the 3D audio effect. All others stated that they were not aware of any hearing impairment at the time.

5-1-3 Stimuli

The audio consisted of one wasp flying and crawling for 51 seconds in an in-door environment without environmental sounds in the following three audio conditions: no audio, stereo and 3D audio. The stimuli were constructed from the same wasp sound as in the preceding experiments. The movement path of the wasp in stereo and 3D audio was exactly the same, but represented in its corresponding audio technique.

Stereoscopic 3D visuals were used since the match with the 3D audio should be most optimal in this case. The visual 3D representation of the wasp was an open source realistic 3D model of a wasp [125]. An illustration of the wasp model after image mapping of the skin can be found in figure 5-2. This model was suitable for being adapted in Autodesk 3DS Max and being exported to Worldviz Vizard. Since the wasp was flying most of the time, the wasp model had to be adapted to get a realistic flying posture, which is illustrated in figure 5-3. The environment was a realistic in-door environment, which can be seen in figure 5-4. The environment is a 360 degrees stereoscopic panorama picture made by Peter Murphy of Australia used as an environment map to the virtual world in Vizard. The virtual world was not interactive. The position and orientation of the observer was fixed and also no head movement was supported. This was done since the chosen 3D audio software did not support interactive rotation of the observer’s head and it was not possible to make a predefined flying path and make the wasp turn along according to the head orientation. Also, there would have been a delay of
5-1 Method

approximately 100ms which could cause a mismatch when the participant moved his or her head very fast, which would have been a normal response to a near flying wasp. This restriction was already covered in section 3-1-4.

As mentioned before in section 5-1-1, there were four different scenarios for the movement path of the wasp: 1) Far away in front of the observer, 2) close in front of the observer landing on left ear, 3) close in front of the observer landing on right ear and 4) wasp sitting and walking on the table. The movement paths of the visuals were constructed using predefined coordinates in Worldviz Vizard with Cubic Bezier interpolation [126]. The movement paths including the implemented code can be found in the appendix D. The audio had to match the movement paths of the visuals. This was done by moving the mono flying wasp sound in SoundLocus with a PS3 joypad that was adjusted with the right sensitivity, to create a basic movement path in SoundLocus. After this, the movie was loaded into SoundLocus and the audio movement path was manually adjusted by hand in order to match the precise movement to the visuals.

For the baseline measurement, a neutral world was presented, being an informative stereoscopic movie about mountain goats, played on a television [9]. A screen-shot of the neutral world is given in figure 5-5.

5-1-4 Listening room, hardware & software

The experiment was done at the Delft University of Technology, Mekelweg 4 in the VRET lab on the second floor, which is a segment of the ISight lab. The VRET lab was not acoustically isolated from the outside world as was the listening room of the previous experiments, but it is a more realistic environment for real VRET treatment. The experimental setup in the room consisted of two main areas: The experimenter control area and the participant area or experimental area. Two desks were located in front of each other, with an isolating screen in between that separated the experimental room into two sections.

On the participant exposure part, there was a Beyerdynamic DT 770 headphone, a Sony HMZ-T2 Head Mounted Display, a Mobi8 hardware device with heart rate and galvanic skin conductance sensors, the Titmus Fly Stereopsis test [127], a pile of books that functioned as an adaptive chin rest, a mounted Samsung Galaxy S2 for participant illumination, a video camera, an experiment pack, a pen and two chairs. The Sony HMZ-T2 HTMD had two OLED Panels, a display resolution of 1280 x 720 pixels, an aspect Ratio of 16:9, a gradation of RGB 24bit, a field of view of 45 degree and a virtual
image size of 750 inch in 20m distance. The technological properties of the Beyerdynamic DT 770 and the Mobi8 were already described in section 4-1-4. A picture of the participant area is given in figure 5-6.

On the experimenter control part, there was a Macbook pro 15 inch with Windows 7 and an external 19 inch LCD monitor for external monitoring the visuals on the HMD. The stimuli were played using the software Worldviz Vizard. The light conditions of the room were controlled such that the participant could see as few external aspects of the VRET lab as possible. However, the face of the participant had to be illuminated in order to capture behavioural aspects with video. Therefore a direct LED light of the Samsung Galaxy S2 was illuminating the face of the participant from the front. This light was not supposed to distract the participant since the light was dimmed by the HMD. A picture of the experimenter control area can be seen in figure 5-7.

5-1-5 Measurement instruments

Most measurement instruments that were used in the experiment "3D audio for VRET without visuals", described in section 4-1-5, were also used in this experiment. They include a descriptive questionnaire, the Fear of Wasps Questionnaire, the IGroup Presence Questionnaire, the Spatial Perception Questionnaire, the Immersive Tendency Questionnaire, the SUD Score, heart rate and galvanic skin conductance. Three more measurement instruments were added: The Simulator Sickness Questionnaire (SSQ), video & audio recordings using a camera for behavioural observation and a stereopsis test in order to validate if the participant could actually perceive stereoscopic visuals. All questionnaires can be found in the appendix C.
5-1-6 Procedure

The flow and structure of this experiment is visualized in figure 5-8. The whole experiment took approximately 55 minutes with a minimum of 45 minutes and a maximum of 60 minutes. Prior to the experiment, participants were provided with the experimental pack, containing a description of the experiment, which had to be signed when the experiment took place. The experimental pack contained a consent form that was approved by the University ethics committee. The language was either Dutch, if the participant was native Dutch speaking, or English. All questionnaires were in English. Before the experiment, people had to fill in the Fear of Wasps Questionnaire, which took approximately 15 minutes.

Intro

The participant was welcomed and asked to sign the experimental pack if all content was read, clear and accepted. An instruction was given about the experiment and if all was clear, the participant was asked to fill in the descriptive questionnaire. After the descriptive questionnaire, the stereopsis test was done, followed by the simulator sickness questionnaire.

Baseline

Before the baseline measurement, the chin rest was adjusted to its optimal height and the Mobi8 sensors, the HMD and headphones were adjusted and mounted. If all equipment was mounted cor-
Figure 5-5: Neutral world used for baseline measurement heart rate and galvanic skin response. Left half is screen-shot of the VR world used and right half is a picture of the actual room, source: [9].

Figure 5-6: Participant area of experimental setup of the 3D audio for VRET experiment.

...rectly and the participant was feeling comfortable, the baseline measurement of the neutral world was started. After 3:00 minutes, the baseline measurement was done and the HMD and headphones were unmounted. For the first time the SUD score was asked. The experimenter placed flags at the beginning and ending of the baseline for later analysis.

Exposure 1, 2, 3 and 4

The real experiment started. The HMD and headphones were again mounted on the participant’s head and the participant was asked whether he or she felt comfortable. If so, the experimenter told the participant that he or she was about to be exposed in 5 seconds. This was because flags could be appropriately placed on accurate time intervals in order to be able to do the data analysis more easily. The duration of the exposure was 0:51 seconds and when the exposure ended, the experimenter placed another flag that the exposure was ended. The headphones and HMD were again unmounted and the chin rest was moved to the side. The participant was asked for the SUD score. After the SUD score was noted, both the IPQ presence questionnaire and the Spatial Perception Questionnaire were given to the participant. The participant was instructed to start with the IPQ presence questionnaire. This was done because presence was of higher importance and presence was measured by doing a "memory task”. The experimenter pointed out that the participant should keep good notice on the scale of the IPQ, since it was sometimes reversed. Also, if the participant had any questions, he or she could ask them at any time. The experimenter left the participant alone filling in the questionnaires. When the participant completed both questionnaires, the SUD score was asked. After noting the SUD score,
the next exposure was followed, until all three stimuli of the wasp in different audio conditions were presented.

Debriefing

When all three exposures were done, again a simulator sickness questionnaire was filled in by the participant, followed by the immersive tendency questionnaire. Finally, the participant was thanked and guided to the exit door.

---

**Figure 5-7:** Control area of experimental setup of the 3D audio for VRET experiment.

**Figure 5-8:** Flow and structure of the 3D audio for VRET experiment
5-2 Results

In this section all statistical analysis of the data is presented in order to answer the main research question: *Can 3D audio more than other commercially available audio techniques create presence, anxiety and spatial perception in VRET with using visuals?*

5-2-1 Stereopsis and simulator sickness

All participants succeeded for the stereopsis test, so all participants were able to perceive the stereoscopic visuals.

Before and after the VR exposures, simulator sickness was measured. There was no statistically significant difference in simulator sickness before and after by running a Friedman test, \( \chi^2(1) = 0.077, p = 0.782 \). Running a similar Friedman test for the sub-scales nausea, oculomotor and disorientation also resulted in no significant differences.

5-2-2 The effect of 3D audio on presence for VRET

**Data preparation & reliability**

The IPQ data was prepared for analysis as discussed in section 4-2-3.

A reliability test of the overall sense of presence for each audio technique separately resulted in Cronbach’s alpha levels between 0.85 and 0.92, which is higher than 0.7 and therefore an acceptable level of reliability [123]. A reliability test for the four sub-scales of spatial presence, involvement, experienced realism and global presence for each audio technique separately resulted in Cronbach’s alpha levels of between 0.68 to 0.88, which is close to or higher than 0.7 and therefore an acceptable level of reliability [123], except for involvement, having Cronbach’s alpha levels between 0.55 and 0.60 for all different audio techniques.

**Statistical analysis**

By looking at the box plots given in figure 5-9, there only seems to be a trend of higher scores for presence of both stereo and 3D audio compared to no audio. The differences between stereo and 3D audio are very small with sometimes even a slightly higher score for stereo.

Histograms and Q-Q plots of the presence data are given in the appendix B in figures B-1, B-2, B-3 and B-4. By looking at the histograms and Q-Q plots, and by doing Shapiro-Wilk normality tests, it was clear that most data was significantly different from a normal distribution. Non-parametric tests were therefore used to test for the effect of audio techniques on presence.

There was a statistically significant difference in the overall sense of presence depending on audio condition whilst running a Friedman test, \( \chi^2(2) = 24.15, p < 0.001 \). Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.017. Median (Q1 to Q3) experienced presence for the audio conditions no audio, stereo and 3D audio were -0.29 (-1.14 to 0.32), 0.64 (-0.46 to 1.14) and 0.43 (-0.11 to 0.79) respectively. There was
Spatial Presence

<table>
<thead>
<tr>
<th>audio conditions</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D audio vs. no audio</td>
<td>-3.73</td>
<td>0.001*</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-0.31</td>
<td>0.754</td>
</tr>
<tr>
<td>stereo vs. no audio</td>
<td>-3.30</td>
<td>0.001*</td>
</tr>
</tbody>
</table>

χ²(2) = 17.61, p < 0.001*

Involvement

<table>
<thead>
<tr>
<th>audio conditions</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D audio vs. no audio</td>
<td>-2.20</td>
<td>0.028</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-0.84</td>
<td>0.401</td>
</tr>
<tr>
<td>stereo vs. no audio</td>
<td>-2.69</td>
<td>0.007*</td>
</tr>
</tbody>
</table>

χ²(2) = 11.07, p = 0.004*

Experienced realism

<table>
<thead>
<tr>
<th>audio conditions</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D audio vs. no audio</td>
<td>-3.14</td>
<td>0.002*</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-0.88</td>
<td>0.380</td>
</tr>
<tr>
<td>stereo vs. no audio</td>
<td>-3.00</td>
<td>0.003*</td>
</tr>
</tbody>
</table>

χ²(2) = 12.78, p = 0.002*

Global presence

<table>
<thead>
<tr>
<th>audio conditions</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D audio vs. no audio</td>
<td>-2.77</td>
<td>0.006*</td>
</tr>
<tr>
<td>3D audio vs. stereo</td>
<td>-0.45</td>
<td>0.655</td>
</tr>
<tr>
<td>stereo vs. no audio</td>
<td>-2.55</td>
<td>0.011*</td>
</tr>
</tbody>
</table>

χ²(2) = 11.51, p = 0.003*

Table 5-4: Summary of Friedman test results regarding the effect of all audio conditions on overall presence and IPQ sub-scales for "3D audio for VRET with visuals". Significant results are marked with an asterisk *.

A significant difference between no audio and stereo (Z = -3.54, p < 0.001), and between no audio and 3D audio (Z = -3.79, p < 0.001). There was no significant difference between stereo and 3D audio (Z = -0.35, p = 0.728).

A Friedman test with post-hoc Wilcoxon Signed-Rank tests and a Dunn-Sidak corrected alpha level of 0.017 was also done for all four IPQ sub-scales separately, see table 5-4. No significant difference was found between 3D audio and stereo on the four IPQ sub-scales. Significant differences were found on almost all sub-scales between no audio and stereo, and between no audio and 3D audio.
Figure 5-9: Box plots of all presence data for audio with visuals for VRET: 14-attributes Overall Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio
5-2 Results

Figure 5-10: Box plots of the SPQ and SSPQ data for all audio conditions.

5-2-3 Effect of 3D audio on spatial perception for VRET

Data preparation & reliability

For preparation, the SPQ and the SSPQ, as discussed in section 4-2-2, was used. The SPQ and SSPQ were only filled in for the stereo and the 3D audio condition since the Spatial Presence Questionnaire explicitly concerns spatial attributes of the audio, which is impossible if there is no audio.

A reliability test of the SPQ and the SSPQ for each audio technique separately resulted in Cronbach’s alpha levels between 0.93 and 0.96, which is higher than 0.7 and therefore an acceptable level of reliability [123].

Statistical analysis

By looking at the box plots given in figure 5-10, the differences between the effect of stereo and 3D audio on spatial perception seem very small. Histograms and Q-Q plots of the SPQ and SSPQ data are given in the appendix B in figures B-5, B-6 and B-7. By looking at the Q-Q plots, histograms and by doing Shapiro-Wilk normality tests, it is clear that the data for stereo audio differ significantly from a normal distribution. Non-parametric analyses were therefore used to test for the effect of audio conditions on spatial perception.

For the SPQ data, there was no statistically significant difference in the spatial perception depending on audio condition whilst running a Friedman test, $\chi^2(1) = 0.2, p = 0.655$. For the SSPQ data, there was also no statistically significant difference found in the sense of spatial perception, $\chi^2(1) = 0.73, p = 0.394$.

5-2-4 Effect of audio technique on SUD score

Data preparation & reliability

The SUD score was measured on a scale from 0 to 10 measuring the sense of anxiety and/or discomfort. This scale including its attributes can be found at the appendix C "Questionnaires". Two SUD
Score measures were analysed: The SUD Score measured after the exposure, which is called the SUD Score Post measurement, and the difference of the SUD Score before and after the exposure, called the SUD Score Increment.

### Statistical Analysis

By looking at the box plots given in figure 5-11, there seems to be a trend of higher scores for SUD Score Post and SUD Score Increment of both stereo and 3D audio compared to no audio. The differences between stereo and 3D audio are very small.

Histograms and Q-Q plots of the SUD Score Post and SUD Score Increment are given in the appendix B in figures B-8 and B-9. By looking at the Q-Q plots, histograms and by doing Shapiro-Wilk normality tests, it is clear that some data for both SUD Score Post and SUD Score Increment differ significantly from a normal distribution. Non-parametric analysis was therefore used to test for the effect of audio conditions on both SUD Score measures.

There was a statistically significant difference in the SUD Score Post depending on audio condition whilst running a Friedman test, \( \chi^2(2) = 20.60, p < 0.001 \). Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.017. Median \((Q_1 \text{ to } Q_3)\) experienced SUD Score Post for the audio conditions no audio, stereo and 3D audio were 3.00 (2.00 to 4.00), 4.00 (3.00 to 5.00) and 4.00 (3.00 to 5.00), respectively. There was a significant difference between no audio and stereo \((Z = -3.446, p = 0.001)\), and between no audio and 3D audio \((Z = -3.559, p < 0.001)\). There was no significant difference between stereo and 3D audio \((Z = -0.24, p = 0.808)\).

There was a statistically significant difference in the SUD Score Increment depending on audio condition whilst running a Friedman test, \( \chi^2(2) = 12.22, p = 0.002 \). Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.017. Median \((Q_1 \text{ to } Q_3)\) experienced SUD Score Increment for the audio conditions no audio, stereo and 3D audio were 0.00 (0.00 to 1.00), 1.00 (0.00 to 2.50) and 1.00 (0.00 to 2.00), respectively. There was a significant difference between no audio and stereo \((Z = -2.68, p = 0.007)\), and between no audio and 3D audio \((Z = -3.04, p = 0.002)\). There was no significant difference between stereo and 3D audio \((Z = -0.27, p = 0.788)\).

So the addition of sound on visuals could elicit more anxiety. In order to validate the stimuli as fearful, the participants were divided into two groups for which "fear for wasps" was either "yes" or "no". Since the median of the Fear of Wasp score over all participants was 3, participants with score 0 to 3 were categorized as "no", and participants with a score of 4 or higher were "yes". There was a statistically significant difference in median between these two groups for 3D audio SUD Score Post \((Z = -2.49, p = 0.013)\) and for 3D audio SUD Score Increment \((Z = -1.99, p = 0.047)\). No significant differences were found between the two groups for the audio conditions no audio and stereo. So, for 3D audio the chosen wasp stimulus was able to elicit more fear with participants who were more afraid of wasps than with participants who were less afraid.
5-2-5 Effect of audio technique on heart rate

Data preparation & reliability

The heart rate was measured in beats per minute (BPM) on different intervals in time and written to a log file. The heart rate means were calculated using the Log Analysis Tool [124], developed by the TU Delft. For the heart rate baseline, only the last two minutes were taken from the baseline period of the experimental session. The heart rate data of the first five participants was not stored successfully, so these cases were excluded from the analysis. Also, participant 11 was an extreme outlier for the "no audio", probably caused by an excessive amount of anticipation anxiety since the heart rate was much higher during the baseline and first exposure, which had the "no audio" condition. This case was also excluded from the analysis.

Statistical Analysis

By looking at the box plots given in figure 5-12, there seem to be a slightly lower heart rate value for no audio compared to stereo, 3D audio and the baseline measurement, although the differences between all conditions seem very small. Histograms and Q-Q plots of the heart rate data are given in the appendix B in figures B-10 and B-11. By looking at the Q-Q plots, histograms and by doing Shapiro-Wilk normality tests it is clear that the "no audio" data differ significantly from a normal distribution. Non-parametric analysis was therefore used to test for the effect of audio conditions on heart rate.

There was a statistically significant difference in the heart rate depending on audio condition whilst running a Friedman test, $\chi^2(2) = 9.79, p = 0.007$. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Dunn-Sidak applied, resulting in a corrected alpha level of 0.017. Median ($Q_1$ to $Q_3$) heart rate for the audio conditions no audio, stereo and 3D audio were 70.41 (68.03 to 73.59), 73.73 (67.36 to 79.04) and 71.15 (68.83 to 76.13), respectively. There was a significant difference between no audio and stereo ($Z = -2.58, p = 0.010$). There was no significant difference between no audio and 3D audio ($Z = -2.01, p = 0.044$) for $\alpha = 0.017$. There was also no significant difference between stereo and 3D audio ($Z = -1.21, p = 0.227$).
5-2-6 Effect of audio technique on galvanic skin response

Data preparation & reliability

The galvanic skin response was measured on different intervals in time and written to a log file. The galvanic skin response means were calculated using the Log Analysis Tool [124], developed by the TU Delft. For the galvanic skin response baseline, only the last two minutes were taken from the baseline period of the experimental session. The galvanic skin response data of the first five participants was not stored successfully, so these cases were excluded from the analysis. Also, participant 25 was an extreme outlier, probably because of the fact that during the first exposure with audio the participant had a sudden frightening reaction when the exposure started. This could have caused an extreme GSR during all successive exposures. This case was also excluded from the analysis.

Statistical Analysis

By looking at the box plots given in figure 5-13, there seem to be lower galvanic skin response values for the baseline measurement compared to no audio, stereo and 3D audio. Differences between no audio, stereo and 3D audio seem very small.

Histograms and Q-Q plots of the galvanic skin response data are given in the appendix B in figures B-12 and B-13. By looking at the Q-Q plots, histograms and by doing Shapiro-Wilk normality tests, it is clear that all data differ significantly from a normal distribution. Non-parametric analysis was therefore used to test for the effect of audio conditions on galvanic skin response.

There was no statistically significant difference in the galvanic skin response depending on audio condition whilst running a Friedman test, $\chi^2(2) = 0.316, p = 0.854$. 

Figure 5-12: Box plot of the heart rate data for all audio conditions and the baseline measurement.
5-2-7 Video & audio recordings

Audio and video recordings were made of the participant’s face only during the exposures in order to see if there were any behavioural effects of the flying wasp. In most videos, only the mouth and some global head movements could be seen and the audio was heard. Patterns were that for the "no audio" condition, there were minimal responses and most of the participants stayed neutral. For stereo and 3D audio there were behavioural responses by smiling or even some fright reactions when the wasp was moving very close to the head. Some subtle reactions could be doubted if they were actually random, but some of them were clearly indicating a response on the wasp. Figure 5-15 illustrates a participant that clearly showed behavioural responses to the stimuli and figure 5-16 reveals some more commonly behavioural responses. Some participants didn’t respond to any audio condition.

5-2-8 Correlations

Presence vs SUD Score

Presence and SUD score were not significantly correlated for no audio ($r(23) = 0.08, p = 0.711$), stereo ($r(23) = 0.39, p = 0.057$) and 3D audio ($r(23) = 0.23, p = 0.267$).

5-3 Conclusion & Discussion

From the results we can conclude that for VRET with visuals, no significant difference was found on presence, anxiety and spatial perception for 3D audio compared to stereo. In addition, the median values were quite comparable, and so, clinical relevance for 3D audio can be doubted.

But, the addition of sound itself combined with visuals was found to have a significantly larger effect on presence and anxiety compared to no audio. A significantly larger effect was found on presence for stereo compared to no audio and for 3D audio compared to no audio. Also, a significantly larger
Figure 5-14: Video and audio recordings of behaviour participants during VR exposure.

Table 5-5: Significantly larger effects for $\alpha = 0.017$ of 3D audio on presence, SUD score, heart rate, galvanic skin response and spatial perception compared to no audio and stereo.

<table>
<thead>
<tr>
<th>3D audio vs.</th>
<th>Presence</th>
<th>SUD Score</th>
<th>Heart rate</th>
<th>Galvanic skin response</th>
<th>Spatial Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>no audio</td>
<td>found</td>
<td>found</td>
<td>not found</td>
<td>not found</td>
<td>not tested</td>
</tr>
<tr>
<td>stereo</td>
<td>not found</td>
<td>not found</td>
<td>not found</td>
<td>not found</td>
<td>not found</td>
</tr>
</tbody>
</table>

Effect was found on SUD score Post for stereo compared to no audio and for 3D audio compared to no audio, and on the SUD Increment for stereo compared to no audio and for 3D audio compared to no audio. Even for heart rate there was a significant increase from no audio to stereo. No significant difference was found for galvanic skin response between all audio conditions.

Table 5-6: Significantly larger effects of audio on presence, SUD score, heart rate and galvanic skin response compared to no audio.

<table>
<thead>
<tr>
<th>audio vs.</th>
<th>Presence</th>
<th>SUD Score</th>
<th>Heart rate</th>
<th>Galvanic skin response</th>
</tr>
</thead>
<tbody>
<tr>
<td>no audio</td>
<td>found</td>
<td>found</td>
<td>found</td>
<td>not found</td>
</tr>
</tbody>
</table>
Figure 5-15: Clear fright responses followed by laughing in the audio conditions.

Figure 5-16: More common subtle behaviour of laughing when wasp was very close to the head for stereo and 3D audio. For no audio the response was mostly neutral.
The results of the ABX test indicate a clear perceived difference between all audio techniques used, i.e., mono, stereo, dolby surround and 3D audio. Hence, differences in performance obtained for different audio techniques applied in a VRET environment can be associated with the audio technique.

From the results we can conclude that sound itself can create anxiety. Also, for VRET with audio only, 3D audio has significantly higher values for overall presence compared to mono and dolby surround. Compared to stereo, the overall presence was not significantly higher, but the presence sub-scale "involvement" was. For SUD score, significantly higher values were found for 3D audio compared to dolby surround but not compared to mono and stereo. For spatial perception, 3D audio only had significantly higher values compared to mono but not compared to stereo and dolby surround. So, 3D audio had a significantly larger effect on presence and anxiety compared to mono and dolby surround, but compared to stereo 3D audio only tended to have a larger effect on presence and anxiety.

In the subsequent experiment with visuals, the comparison was more focused on the difference between 3D audio and stereo, but also the condition "no audio" was added. Also, it was more applied to the VRET domain by integrating 3D visuals using an HMD inside a real VRET treatment environment. No significant differences between 3D audio and stereo were found during this experiment. But, there was a significantly higher effect of audio compared to no audio on presence, SUD score and heart rate.

From the results of both experiments it is clear that the differentiating effect between 3D audio and stereo is reduced after adding visuals. The significantly larger effect of 3D audio on the presence sub-scale involvement disappeared and the trend of 3D audio having a larger effect on overall presence and anxiety was reduced. This is probably because the audio was overruled by the visuals when combining the information. Vision is known to be the dominant sense, and when combining audio with visuals, visuals "take over". As a consequence, different audio techniques may result in a similar experience, a phenomenon also illustrated by the McGurk effect [128], where slightly different visual information
influences the perceived auditory information, resulting in a change in the perceived sound from the real auditory input. But, the flying wasp stimuli used were most of the time within the line of sight. Outside the line of sight, audio is known to be the dominant sense [31]. Therefore, it is possible that when the stimuli are outside the line of sight and visuals are used, the 3D audio effect becomes higher, since both modalities supplement each other. Follow-up studies are needed to evaluate this hypothesis.

From our conclusions, the added value of 3D audio for VRET may be doubted. The effect of audio itself combined with visuals is significantly higher on the main moderating variables presence and anxiety compared to no audio. It is therefore advised to incorporate audio into a VRET system. Stereo and dolby surround are the most commonly used audio techniques [31]. However, dolby surround is doubtful when using headphones instead of loudspeakers. Since 3D audio is more expensive and time-consuming to realize and since it hardly showed any added value compared to stereo, the most logical choice in a clinical setting is stereo. Furthermore, it may be more effective to increase the number of matching modalities instead of optimizing only one in particular.

In order to get more insight into the effect of 3D audio compared to stereo, follow-up studies can be done. Some factors can be improved with respect to the experiments described in this report. It should be mentioned that all experiments were done with non-patients. Using real patients could generate a more pronounced effect of technological differences on perceived presence and anxiety, and therefore, it would be useful to repeat at least the last experiment with real patients. It must also be mentioned that the used 3D audio technique, i.e. using signal processing with HRTFs and human hearing modelling, is probably not as convincing as 3D audio recorded by a binaural microphone. No current literature exists with significant results, but claims are being made in anecdotal reports, websites and through listening demo’s [129]. If these claims are true, possibly better results can be obtained using the binaural recording technique. The main disadvantage for this technique is that it is not flexible since the movement behaviour of the sound sources is stigmatized in the recording.

Also, if 3D audio is combined with visuals, it should be more explicitly tested if the 3D audio effect is increased when the sound sources are outside the line of sight. This is plausible, since the modalities supplement each other and audio is known to be the dominant sense outside the line of sight. The stimuli could also be made more "realistic" by adding reverberation and environmental sounds. There is also need for developing a standardized spatial perception questionnaire that is able to measure the performance of 3D audio spatial attributes, since current evaluation techniques do not include this kind of properties.

There are of course other 3D audio reproduction techniques, but since 3D audio through headphones, with the possibility of head movement interaction, has high potential for commercial use because of its practical use and potentially low costs, it is suggested to study 3D audio through headphones. Meanwhile 3D audio reproduction techniques using signal processing and human hearing modelling should be further developed until maximal performance and maybe even individualized 3D audio reproduction techniques are possible that approach the performance of the actual human hearing. 3D audio studies may also concern other applications than VRET, for example movies and games.
Appendix A

Exploratory Data Analysis - 3D audio for VRET without visuals
Figure A-1: Box plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for all audio techniques.
Figure A-2: Histograms of the 14-attributes Total Presence data
Figure A-3: Histograms of the IPQ sub-scales Spatial Presence and Involvement data
Figure A-4: Histograms of the IPQ sub-scales Experienced Realism and Global Presence data
Figure A-5: Q-Q plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for all audio techniques.
**Figure A-6:** Box plots of the 10-item and 13-item Spatial Perception data for all audio techniques.

**Figure A-7:** Q-Q plots of the 13-item and 10-item Spatial Perception data for all audio techniques.
Figure A-8: Histograms of the 13-item and 10-item Spatial Perception data
Figure A-9: Box plot, histograms and Q-Q plots of the SUD score data for all audio techniques.
Figure A-10: Box plot and histograms of the heart rate means data for all audio techniques including the baseline.
Figure A-11: Box plot and histograms of the baseline-corrected heart rate data for all audio techniques.
Figure A-12: QQ plots of the heart rate means data for all audio techniques including the baseline.
Figure A-13: QQ plots of the baseline-corrected heart rate data for all audio techniques.
Appendix B

Exploratory Data Analysis - 3D audio for VRET with visuals
Figure B-1: Box plots of all Presence data for audio with visuals for VRET: 14-attributes Overall Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio
Figure B-2: Histograms of the 14-attributes Overall Presence data for audio with visuals for no audio, stereo and 3D audio
Figure B-3: Histograms of the four IPQ sub-scales data for audio with visuals for VRET: Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio
Figure B-4: Q-Q plots of all Presence data: 14-attributes Total Presence, Spatial Presence, Involvement, Experienced Realism and Global Presence for no audio, stereo and 3D audio.
Figure B-5: Box plots of the 10-item (SPQ) and 13-item (SSPQ) Spatial Perception data for all audio conditions.

Figure B-6: Histograms of the (SPQ) 13-item and 10-item (SSPQ) Spatial Perception data for all audio conditions.
Figure B-7: Q-Q plots of the 13-item (SPQ) and 10-item (SSPQ) Spatial Perception data for all audio conditions.
Figure B-8: Box plot, histograms and Q-Q plots of the SUD score data measured after the exposure for all audio conditions.
Figure B-9: Box plot, histograms and Q-Q plots of the increment in SUD score during the exposure data for all audio conditions.
Figure B-10: Box plot and histograms of the heart rate data for all audio conditions.
Figure B-11: Q-Q plots of the heart rate data for all audio conditions.
Figure B-12: Box plot and histograms of the Galvanic Skin Response data for all audio conditions.
Figure B-13: Q-Q plots of the Galvanic Skin Response data for all audio conditions.
Appendix C

Questionnaires
**Igroup presence questionnaire (IPQ)**

Please indicate whether or not each statement applies to your experience. You can use the whole range of answers. **And please remember: Answer all these questions only referring to this one experience.**

How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

<table>
<thead>
<tr>
<th>Extremely aware</th>
<th>Moderately aware</th>
<th>Not aware at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How real did the virtual world seem to you?

<table>
<thead>
<tr>
<th>Completely real</th>
<th>Moderately real</th>
<th>Not real at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I had a sense of acting in the virtual space, rather than operating something from outside.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th>Moderately disagree</th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How much did your experience in the virtual environment seem consistent with your real world experience?

<table>
<thead>
<tr>
<th>Not consistent</th>
<th>Moderately consistent</th>
<th>Very consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

About as real as an imagined world

<table>
<thead>
<tr>
<th>Indistinguishable from the real world</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
</tr>
<tr>
<td>-2</td>
</tr>
<tr>
<td>-1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>+1</td>
</tr>
<tr>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
</tr>
</tbody>
</table>

I did not feel present in the virtual space.

<table>
<thead>
<tr>
<th>Felt present</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
</tr>
<tr>
<td>-2</td>
</tr>
<tr>
<td>-1</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>+1</td>
</tr>
<tr>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
</tr>
</tbody>
</table>
I was not aware of my real environment.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the computer generated world I had a sense of "being there".

<table>
<thead>
<tr>
<th>Not at all</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Very much</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Somehow I felt that the virtual world surrounded me.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I felt present in the virtual space.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I still paid attention to the real environment.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The virtual world seemed more realistic than the real world.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I felt like I was just perceiving pictures.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I was completely captivated by the virtual world.

<table>
<thead>
<tr>
<th>Fully disagree</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Fully agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
<td>+2</td>
<td>+3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fear of Wasps Questionnaire

Please select how you feel towards wasps. A typical wasp is a yellow-black striped flying insect that can sting with venom coming out of their "stinger" tail.

* Required

What is your full name? *

If I came across a wasp now, I would get help from someone else to remove it. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Currently, I am sometimes on the lookout for wasps. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If I saw a wasp now, I would think it will harm me. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I now think a lot about wasps. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I would be somewhat afraid to enter a room now, where I have seen a wasp before. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I now would do anything to try to avoid a wasp. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Currently, I sometimes think about getting stung by a wasp. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If I encounter a wasp now, I wouldn’t be able to deal effectively with it. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If I encountered a wasp now, it would take a lot time to get it out of my mind. *

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>If I came across a wasp now, I would leave the room.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I saw a wasp now, I would think it will try to sit on me.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I saw a wasp now, I would ask someone else to kill it.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I encountered a wasp now, I would have images of it trying to get me.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I saw a wasp now I would be afraid of it.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I saw a wasp now, I would feel very panicky.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wasps are one of my worst fears.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would feel very nervous if I saw a wasp now.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If I saw a wasp now I would probably break out in a sweat and my heart would beat faster.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Do you check the lounge for wasps before sitting down? *
- Yes
- No

Do you ever make plans in case you come across a wasp? *
- Yes
- No

Do you sometimes look at the corners of the room for wasps? *
- Yes
- No

When watching television, would you notice a wasp crawling across the floor elsewhere in the room? *
- Yes
- No

Do you check the bedroom for wasps before going to sleep? *
- Yes
- No

Would your mind be a lot easier if wasps didn’t exist? *
- Yes
- No

Are you always on the lookout for wasps? *
- Yes
- No

Have you a “plan for action” in case you find a wasp in the kitchen? *
- Yes
- No

Do you make very certain there are no wasps around before taking a bath? *
- Yes
- No

Do you sometimes sense the presence of a wasp without actually seeing it? *
- Yes
- No

If there’s a wasp in the house, are you the most likely person to find it? *
- Yes
- No

Can you spot a wasp out of the corner of your eye? *
- Yes
- No
Do you sometimes dream about wasps? *
☐ Yes
☐ No

Do you think a lot about wasps? *
☐ Yes
☐ No

Do you worry more about wasps than most people? *
☐ Yes
☐ No

When you imagine a wasp, can you see parts of it in great detail? *
☐ Yes
☐ No

Do you ever find yourself thinking about wasps for no reason? *
☐ Yes
☐ No

Do you sometimes find it an effort to keep thoughts of wasps out of your mind? *
☐ Yes
☐ No

Do you often think about particular parts of wasps, for example the fangs? *
☐ Yes
☐ No

Are you sometimes distracted by thoughts of wasps? *
☐ Yes
☐ No

Are you sometimes haunted by thoughts of wasps? *
☐ Yes
☐ No

When watching television do you think more about the danger of there being a wasp in the room than about the programme? *
☐ Yes
☐ No

Have you had nightmares about wasps? *
☐ Yes
☐ No

Can you deal effectively with wasps yourself when you find them? *
☐ Yes
☐ No
Do you get other people to get rid of wasps when you find them? *
  Yes
  No

Would you know how to cope with wasps in the bath? *
  Yes
  No

Do you sometimes use a book or a newspaper to deal with a wasp? *
  Yes
  No

Do you feel a lot more secure if someone else is in the house, in case you come across a wasp? *
  Yes
  No

When you find a wasp in a room, would you avoid going in that room until someone else had removed it? *
  Yes
  No

Would you get help if you came across a wasp? *
  Yes
  No

If you find a wasp in the bath, would you, say, use a shower to wash the wasp down the plughole? *
  Yes
  No

If you discover a wasp in the room, do you leave the room straight away? *
  Yes
  No

Would you think about using a broom to deal with a wasp in the kitchen? *
  Yes
  No

When imagining a wasp, is it always the same one or kind? *
  Yes
  No

Do you ever lie in bed at night and listen out for wasps? *
  Yes
  No

If you thought you saw a wasp would you go for a close look? *
  Yes
  No
When you see a wasp, does it take a long time to get it out of your mind? *
- Yes
- No

Are you slightly scared to enter a room, say a bathroom, where wasps have been in the past? *
- Yes
- No

Are wasps insects? *
- Yes
- No

Do wasps have six legs? *
- Yes
- No

Are wasps solely meat eaters? *
- Yes
- No

Have you a good idea whereabouts wasps are likely to appear? *
- Yes
- No

Do you know when (what time of year) you are likely to come across a wasp? *
- Yes
- No
Immersive Tendency Questionnaire
This questionnaire measures differences in tendencies of individuals to experience presence.

1. Do you easily become deeply involved in movies or TV dramas?
   1 2 3 4 5 6 7
   Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

2. Do you ever become so involved in a television program or book that people have problems getting your attention?
   1 2 3 4 5 6 7
   Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

3. How mentally alert do you feel at the present time?
   1 2 3 4 5 6 7
   Not alert ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very alert

4. Do you ever become so involved in a movie that you are not aware of things happening around you?
   1 2 3 4 5 6 7
   Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

5. How frequently do you find yourself closely identifying with the characters in a story line?
   1 2 3 4 5 6 7
   Not frequently ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very frequently

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?
   1 2 3 4 5 6 7
   Not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

7. How physically fit do you feel today?
   1 2 3 4 5 6 7
   not fit ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very fit
8. How good are you at blocking out external distractions when you are involved in something?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>not good</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very good</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

9. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>not at all</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very much</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

10. Do you ever become so involved in a daydream that you are not aware of things happening around you?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>not at all</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very much</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

11. Do you ever have dreams that are so real that you feel disoriented when you awake?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>not at all</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very much</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

12. When playing sports, do you become so involved in the game that you lose track of time?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>not at all</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very much</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

13. How well do you concentrate on enjoyable activities?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not well</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very well</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

14. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not often</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Very often</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
15. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16. Have you ever gotten scared by something happening on a TV show or in a movie?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not scared</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17. Have you ever remained apprehensive or fearful long after watching a scary movie?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18. Do you ever become so involved in doing something that you lose all track of time?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Intro Questionnaire

Do you suffer from any kind of hearing impairment? If yes, please describe.

........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................

Do you have a fear of wasps? *

0 1 2 3 4 5 6 7 8 9 10
No fear at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

How scared are you to do this experiment? *

0 1 2 3 4 5 6 7 8 9 10
Not scared at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very much

What is your age?
........................................................................................................................................

What is your educational level?
........................................................................................................................................

What is your gender?
........................................................................................................................................
Simulator Sickness Questionnaire

Please answer all these questions only referring to the LAST VIRTUAL REALITY you experienced. Please report the degree to which you experience each of the symptoms as one of “doesn't feel anything”, “a little”, “Medium” and “a lot”. These are scored respectively as 0, 1, 2 and 3.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>0-‘doesn’t feel anything’</th>
<th>1-‘a little’</th>
<th>2-‘medium’</th>
<th>3-‘a lot’</th>
</tr>
</thead>
<tbody>
<tr>
<td>General discomfort</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Headache</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Eyestrain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Difficulty focusing</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Increased salivation</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sweating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Nausea</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Difficulty concentrating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Fullness of head</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Blurred vision</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Dizzy (eyes open)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Dizzy (eyes closed)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Vertigo</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Stomach awareness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Burping</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
**Spatial Sound Perception Questionnaire**

Please indicate whether or not each statement applies to your experience. You can use the whole range of answers. **And please remember: Answer all these questions only referring to this one experience.**

| The sonic environment sounds ... | enveloping | | | | | very |
|---------------------------------|------------|
| a little                        | -3 -2 -1 0 +1 +2 +3 |

| I feel ... immersed in the sonic environment | | | | | very |
|---------------------------------------------|------------|
| a little                                    | -3 -2 -1 0 +1 +2 +3 |

<table>
<thead>
<tr>
<th>The representation of the sonic environment was ...</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor</td>
<td>-3 -2 -1 0 +1 +2 +3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The sense of externalization of the scene was ...</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor</td>
<td>-3 -2 -1 0 +1 +2 +3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The sense of localization of the scene was ...</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor</td>
<td>-3 -2 -1 0 +1 +2 +3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The sense of distance of the scene was ...</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor</td>
<td>-3 -2 -1 0 +1 +2 +3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The sense of movement of the scene was ...</th>
<th>good</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor</td>
<td>-3 -2 -1 0 +1 +2 +3</td>
</tr>
</tbody>
</table>
The sense of directions of the scene was ...  
poor  
-3 -2 -1 0 +1 +2 +3  
good

The sense of depth of the scene was ...  
poor  
-3 -2 -1 0 +1 +2 +3  
good

The sense of space of the scene was ...  
poor  
-3 -2 -1 0 +1 +2 +3  
good

The sense of realism of the scene was ...  
poor  
-3 -2 -1 0 +1 +2 +3  
good

The naturalness, true to life, was ...  
not truthful  
-3 -2 -1 0 +1 +2 +3  
truthful

The quality of the reproduction was ...  
poor  
-3 -2 -1 0 +1 +2 +3  
good

Comments (if any):

........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
........................................................................................................................................................................
Appendix D

Designing and realizing the auditory virtual environment

D-1 Sound selection

VRET is done by gradually exposing people to their fear in order to decrease their anxiety for certain situations in real-life [14]. Since our participants were all not suffering from a known anxiety disorder, the presented stimuli should create a universal sense of fear. The selection for an applicable scenario for this experiment was made by formulating requirements, brainstorming about possible scenarios and an evaluation of these scenarios with respect to the requirements.

The requirements for the auditory fragments are listed below:

Requirements

**R1:** The audio fragment has to cause a universal discomfort and/or fear to people that is not known to be suffering from an anxiety disorder.

**R2:** The possibility to play back the audio fragments through all four audio techniques mono, stereo, dolby surround and 3D audio, using headphones, with the exact same auditory landscape, sound source position and sound source movement.

**R3:** The possibility to realize the audio fragments in high quality, with sample rate of at least 44.100 Hz, a bit depth of at least 16 bits and a bit rate of at least 312 Kb / sec, recorded or modelled with high quality hardware and/or software, with minimal signal distortion and other perceivable unwanted artefacts caused by anything other than the sound source itself.

**R4:** The moving sound source must have both high-frequency and low-frequency components.

**R5:** The audio fragment should include a stressor that has some dynamic movement properties, horizontally, vertically and in distance. Preferable also sometimes close to the head, with a distance of around 1 cm from ones ear.
R6: The audio fragment should include a stressor that emits sound with a loudness level that is easily perceivable or distinguishable from possible other sound sources if they are present in the audio fragment.

R7: The audio fragments may not cause a panic reaction.

A motivation for these requirements is given below:

R1: Since no real patients suffering from an anxiety disorder participated in this experiment, the audio fragment has to create a universal sense of discomfort and/or fear to the average listener in order to increase the meaning of anxiety.

R2: The exact same scenario should be presented in terms of sound source movement, sound source position, etc. in order to reduce variance caused by stimuli inconsistency. Ideally, the sound should be recorded in all four techniques at the same time, or be modelled once in one piece of software and then be exported to each file format separately, causing differences to be minimized.

R3: The quality of the audio in terms of recording and modelling techniques, sample and bit rate should be high in order to exclude unwanted distractions. The human hearing can perceive from around 20 to 20,000 Hz. According the sampling theorem, if the audio is sampled at twice the highest frequency, the signal can be reconstructed perfectly without any loss. Since 20,000 Hz is the maximum frequency, a sampling rate of 44,100 Hz is above this threshold, allowing perfect reconstruction. Why 44,100 Hz is the standard sampling rate for CD quality "perfect" audio reconstruction can be found in [126].

R4: This requirement was formulated since some frequencies are known to be less hard to localize [31]. Both low- and high-frequency components should be present for optimal performance.

R5: A stressor should contain dynamic movement properties in all directions in order to get a good comparison of the spatial properties of all audio techniques. Especially localization very close to one’s head, with a distance of around 1 cm to ones ear, have been reported to be very effective in causing fight or flight reactions [31].

R6: The loudness level of a stressor should be easily perceivable in loudness. If for example the sound of a walking spider is presented with a realistic loudness level, probably nothing will be perceived.

R7: A panic reaction can over-stimulate emotion and therefore perception, making the results less reliable for especially the "spatial perception" parts. Also, a panic reaction can be caused using a simple mono sound, for example, a sudden unexpected loud growl of a lion next to ones ear. It can be interesting to see in what degree all four audio techniques differ in relation to cause panic reactions, but this would be another experiment. Also, VRET is based on the principle of gradual exposure, making subtle increments in sense of fear. A panic reaction is a thing that should be avoided.
D-1-1 Sound selection

A brainstorm was held in order to come up with a list of possible suitable scenarios for the audio fragments. The stimuli should provoke a global sense of fear and/or discomfort, since all participants were not known to be phobia patients. Some scenarios that were found to be suitable were: flying bee / wasp, flying mosquito, barking / growling dog, growling lion and unknown beast at night in a forest and an intimidating person approaching from behind with gasping in the participants neck. These scenarios were evaluated by rating them according to the requirements formulated earlier, and the scenario of the flying wasp was rated as highest.

D-1-2 Sound processing

The requirement R2 states that all sound scenarios should be represented the same in terms of auditory landscape, sound source position and sound source movement, realized with all four audio techniques. This implies that a method has to be used that maximizes this consistency. Two methodologies are discussed below. These methods were designed by literature research, online search and logical reasoning. The methods are being explained and evaluated in terms of practicability, costs, quality, cross-over audio technique constancy, flexibility and controllability. The first method to be discussed is "Recording through microphones with natural play-back" and the second is "Modelled play-back using signal processing".

Recording through microphones with natural play-back

This method regards the recording of the real-life wasp sound using a microphone setup that is able to record the sound for each audio technique at once. After being recorded, post-processing has to be done for each audio technique in order to make them as consistent as possible. It is important that all recording techniques are recording the wasp at the exact same time in order to maximize the consistency between the behaviour and sound properties of the flying wasp. An ideal situation of this method would be the following:

Release a wasp in a desirable environment, which can be an anechoic chamber in order to isolate the wasp sound and exclude environmental audio cues, or in a room with desirable acoustical properties. Record the wasp by setting up multiple microphone arrays for the different audio techniques, by taking into account the exact position of the listener. For recording 3D audio, a binaural dummy head microphone like the Neumann KU 100, illustrated in figure D-1, or in-ear microphones be used. By doing this the HRTFs are present in the recording itself, immediately realizing 3D audio if the recorded sound is played back through normal headphones. The recording of a binaural microphone is not compatible with dolby surround, stereo and mono play-back, since two channels are being recorded, with the HRTF present in the recording itself. It is impossible to cancel these binaural properties caused by head-shadowing, interaural differences and reflectional properties of the pinnae and the torso. For dolby surround a setup can be used of multiple microphones correctly positioned according to the Dolby standards [130], a dolby decoder and a mixing device. There are many recording techniques, all requiring different positioning of multiple microphones, with no real "standard", but there are guidelines for good dolby surround recording practising. Some of them are mentioned in [131]. A method has to be chosen where the post-processing is as little needed as possible, in order to keep good consistency with the other recording techniques for mono, stereo and 3D audio. Some of
For best practices, also a separate recording technique for stereo and mono have to be applied. For recording stereo, there are also several methods, each trying to get an "as good as possible" stereo image of the sound field by only using two channels. The most popular ones used by many professional audio engineers are described in [132]. The AB Stereo method can be used for this purpose. For recording mono, only one omnidirectional polar patterned microphone has to be used and positioned on the exact same spot of the listener. The omnidirectional microphone differs from a cardioid-pattern microphone, presented in figure D-3, in a way that the polar pattern is more or less equal in every angle. This results in a realistic mono representation of the sound source since the movement and position is captured equally in every angle, creating accurate distance cues for only one channel. Note that the azimuth and elevation cues are not present in mono recordings.

Many microphones are used and positioned at specific places in order to maximize the performance for recording each different technique, but because all four techniques have to be recorded simultaneously, some microphones have to be positioned at the same place, which is impossible, resulting in some inconsistency in the recordings. Also, the binaural dummy head will probably affect the recordings of mono, stereo and dolby surround channels in some degree, because the reflections from the dummy head are also picked up by the microphones recording the other techniques. But, since this dummy head is positioned in the room and is therefore "part of" the recording environment, so consistency should not be lost.

**Modelled play-back using signal processing**

This approach uses a 1-channel mono audio source with the pure direct sound of the desired sound source object, in our case the wasp, without any other informational cues like environmental context, sound source location and distance. Ideally, the sound source is recorded using one microphone that is located on a static position relative to the sound source position, in an anechoic room [133]. In an anechoic room all walls are being completely covered with absorbing material and even the listener has to float on a metal rack in order to achieve such extreme effects. Only a few of such
anechoic chambers exist because they have to meet very strict properties that are practically very hard to achieve at high costs. An image of an anechoic room is given in figure D-2. These chambers are mainly used for testing audio products or research projects that demand such extreme circumstances. But even when releasing a wasp in an anechoic chamber, there is still the problem of the static position of the microphone compared to the moving wasp. It is practically impossible to attach a microphone to a flying wasp. When a microphone is recording the wasp, but is not attached to the wasp, some undesired distance cues are already present in the recording. These distance cues can be partially eliminated using the appropriate sound post processing tools like compression or expanders, which we will discuss later on.

A more practical method is to use "close miking" with one microphone in an environment with limited reflections and background noises. Close miking is a technique in which there is one microphone is positioned as close as possible, aimed to the sound source object in order to get a more direct or "close" sound and hereby reducing background noises like reflections and/or external sounds. Preferably a microphone with a cardioid pattern is used in order to reduce environmental sounds and/or reflections even more. A cardioid patterned microphone picks up the most signal that is right in front of the aiming position and picks up less signal in the other directions, according to its pattern. A regular cardioid pattern is given in figure D-3. Even the remaining reflections and/or background noises can be further reduced using signal enhancement and noise reduction techniques [134]. But, it must be mentioned that these noise reduction techniques can also affect the shape of the sound of the sound source itself, which can be an unwanted side-effect.

It is also possible to completely model the sound source using analysis - synthesis audio modelling techniques [135]. If done accurately, it can give enormous flexibility about the behaviour and nature of the sound and it would eliminate all problems discussed in the paragraph above like background noise reduction, loudness level compression. But, because it is a modelled simulation, it will probably sound less natural or realistic if the model is slightly inaccurate. A demonstration of a modelled wasp
can be listened on the website [136].

Once the mono sound source is realized, the sound source movement, positioning and distance can be modelled in a 360° sphere around the listeners head using software tools that apply signal processing with models of the human hearing system, for spatial placement and movement of the sound source. Different software tools are available that promise to achieve this task, but all in a different way. They differ in HRTFs, acoustical models of the environment, GUI’s with varying options and functionality, and possibilities to convert the auditory landscape into different formats. A list of requirements was formulated for the software tools in order to find the most suitable software needed for this experiment. A search study was done to find all currently available 3D audio processing tools and they were compared by evaluating the tools according to the formulated requirements. The formulated requirements, according to the MoSCoW [137] method are stated below:

**Must haves**

**R_M1**: Placing the mono sound source into a 3D sphere around the listeners head in terms of azimuth, elevation and distance (from now on called x, y and z).

**R_M2**: The ability to move the sound source by either defining a pre-defined movement path or by using the mouse or a joy stick in the locations x, y and z for experimental consistency and controllability. Preferable is a pre-defined path.

**R_M3**: The ability to convert the auditory virtual environment to the four audio techniques mono, stereo, dolby surround and 3D audio through headphones with a maximized consistency in
terms of sound source movement, location, distance and environmental context between all four audio techniques.

**R_M4:** Convincing in the 3D audio representation.

**Should haves**

**R_S5:** Compatibility of controlling sound source position, distance and movement by other software tools in order to maximize compatibility with VRET systems.

**R_S6:** Sound source movement in the locations x, y and z are being done in real-time for allowing interactive head movement in the future.

**R_S7:** The ability to add other sound sources or environmental sounds to the auditory virtual environment by multi-track mixing.

**R_S8:** The ability to control auditory properties like duration, volume, dynamics, spectral filters, and possible even environmental room acoustics.

**R_S9:** Easy to use

**R_S10:** Low cost

No "could haves" and "would haves" are formulated. In R_Sx, the R stands for "Requirement", the M for "Must have", the S for "Should have" and the number x for the requirement number.

A search was done for 3D modelling realization tools by keyword searching on Google and Google scholar. This resulted in some complete software packages that promised to do most of the requirements. Some tools were more aimed on the commercial business, others more scientific. The most promising software packages that could deliver real 3D audio using HRTFs were CLAM, SoundLocus and Longcat 3D audio. All three software platforms are described and evaluated according to the requirements below.

**CLAM**

CLAM (C++ Library for Audio and Music) is a full-fledged software framework for research and application development in the Audio and Music Domain. It offers a conceptual model as well as tools for the analysis, synthesis and processing of audio signals [138]. Stable releases of the basic CLAM platform were pre-compiled for Linux, Mac OS and Windows and it is open source and free to use under the GLU license. In order to use spatial 3D audio, the spatialization plugin has to be used. This is done by re-compiling the CLAM distribution on a Linux Ubuntu 11.04 system together with the additional plugin files. The spatialization plugin can place a mono audio file into 3D space by making use of real HRTFs. It is also possible to include or combine different HRTFs, even individualized HRTFs if they are present. CLAM makes use of a Network Editor, which is a GUI providing basic audio signal processing elements in order to make a network of basic transformations realizing complex audio processing chains. An illustration of the CLAM Network Editor is given in figure D-4. Mono sound sources can be placed in the 3D space by allocating values to the x, y and z parameters of the spatialization plugin. There is a high quality "offline" mode, but also a lesser quality real-time mode, in order to combine 3D audio with games or head-tracking devices. It is also possible to realize the auditory virtual environment to all four audio techniques mono, stereo, dolby surround and 3D.
audio. There are also plenty of other additional plugins, for example for room modelling. A demo can be listened on [139], where CLAM is used to generate real-time 3D audio in a game YoFrankie developed in the game engine Blender [140]. CLAM is able to be used in combination with other platforms. For example, a python wrapper has been made in order to use CLAM together with python based products. Also, it is extendible in C++. CLAM is very promising for audio signal processing education and project development. One disadvantage is that at the moment of writing, the source code was not compiling and the source code was not kept up-to-date for more than one year.

SoundLocus

SoundLocus is an authoring tool for realizing binaural 3D audio by placing pre-recorded or modelled sound sources in the 3D space by making use of HRTFs and a GUI to control the sound source position, distance and movement path relative to the listener. The software tool was commercially developed by the Japanese company Arnis Sound Technology for Windows and Mac aiming at the film, gaming and music branch. SoundLocus is recently available in the Netherlands since begin 2012. No scripting languages are needed; it is a full compiled GUI based software tool with lots of functionalities with a metaphor inspired on modern professional audio production software packages like ProTools and Cubase. SoundLocus is most noticeable because GUI with lots of manners to place mono sound sources in the 3D space and even defines a path of the movement of each sound source in time. This path is editable and configurable in different ways. An impression of the GUI is given in figure D-5. SoundLocus is capable of exporting the created auditory virtual environment to 2-channel binaural 3D audio and 5.1 dolby surround. Stereo can be gathered by only using the two front speakers and making a new stereo file by making use of another audio production tool like Cubase. It is also compatible with other software tools by controlling certain sound properties such as position and

Figure D-4: The CLAM Network Editor.
movement using MIDI. MIDI is an electronic musical instrument industry specification that enables a wide variety of digital musical instruments, computers and other related devices to connect and communicate with one another [141]. A demo can be listened at [136]. The main drawback of SoundLocus is that it is a commercial product and therefore it comes at a price for around €750.

LongCat 3D audio

LongCat is a French company developing 3D audio tools for audio and music production, but also to be combined with VR systems. There are three main products capable of delivering 3D audio: AudioStage, H3D plugin and SoundPack for Virtools. AudioStage is a full GUI based 3D audio software tool that is capable of placing pre-recorded or modelled one-channel sound sources in 3D space using the GUI and your mouse or another external controlling device using MIDI. It is also compatible with all major audio software: ProTools, Pyramix, Nuendo, Cubase and Logic Audio. AudioStage is an advances tool capable of modelling multiple virtual rooms with each its own acoustical characteristics. The sound sources can be moved in these rooms, updating the environmental acoustics along its path, even with transition algorithms between the rooms. A screenshot of AudioStage can be seen in figure D-6 and a demo can be viewed and listened on [142]. AudioStage is capable of exporting the auditory virtual environment to binaural 3D audio, 5.1 dolby surround, stereo and mono and it uses HRTFs for the 3D audio. AudioStage comes at a price of €250 for the Stereo edition, with the possibility to add surround capabilities for a fixed time.

The H3D plugin uses the very same audio engine as AudioStage, but with a limited amount of settings. It can be used it in a real-time audio environment, such as PureData, Max/MSP, Bidule or AudioMulch and there are also options to control the sound source positions with MIDI. This plugin is suitable if only simple 3D audio spatialization in sound source location, distance and movement are needed without complex room modelling options. For this experiment, this will probably be the case. The H3D plugin comes at a price of €82.78 and is compatible with Windows XP, Vista or Seven and OSX 10.6 or 10.7. A screenshot of the GUI of the H3D plugin can be seen at figure D-7.
There is another product of LongCat audio: SoundPack for VirTools. VirTools is a tool to create and model 3D environment for simulation and VR applications. SoundPack also uses the same audio engine as AudioStage, but is solely developed for using in combination with VirTools. The SoundPack Player Edition is available for €250 and can only playback VirTools compositions using spatialized 3D renderings in OfficeXE or standalone Virtools players. The SoundPack Dev Edition is available for €2500 and allows to create, edit and playback Virtools compositions using spatialized 3D renderings from within the Virtools Dev environment. VirTools itself is not available for sale any more, making it a dependency that is practically almost unachievable if one does not already own VirTools.

**Figure D-6:** A glance of the GUI of Longcat AudioStage.

**Figure D-7:** A glance of the GUI of Longcat H3D plugin.

**Evaluation, selection and realization**

All software tools are evaluated according to the formulated requirements R_M1, ..., R_S10 on a 5-point Likert scale by the author of this document, where the "must haves" are weighted double and the "should haves" weighted once. The results are presented in table D-1. SoundLocus has the highest score, mostly because of its 3D audio performance and having the most flexible options with the highest "ease of use" compared to the other software packages, but it comes at the highest cost. CLAM
Table D-1: Sound scenario’s evaluated according to the formulated requirements. All requirements are weighted equally except for R_M1, which is weighted double.

<table>
<thead>
<tr>
<th></th>
<th>R_M1</th>
<th>R_M2</th>
<th>R_M3</th>
<th>R_M4</th>
<th>R_S5</th>
<th>R_S6</th>
<th>R_S7</th>
<th>R_S8</th>
<th>R_S9</th>
<th>R_S10</th>
<th>µ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAM</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>++</td>
<td>3.79</td>
</tr>
<tr>
<td>Longcat</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3.79</td>
</tr>
<tr>
<td>SoundLocus</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>4.07</td>
</tr>
</tbody>
</table>

and Longcat have the same score but have different weaknesses and strengths. CLAM is the only package that has no boundary on the distance of the sound source, having a higher score for R_M1, it is free and primarily developed for research, making it an attractive choice. But, on the other hand, it has a very steep learning curve and the current branch is unusable due to compilation errors. Longcat was less convincing than SoundLocus on many parts, but is was much cheaper. Finally SoundLocus was chosen for these experiments. But, because SoundLocus can only export the auditory virtual environment to 5.1 dolby surround and binaural 3D audio, other tools were needed in order to realize also stereo and mono. For realizing stereo, the exported 5.1 dolby surround channels were loaded into Cubase, which is a multi-track audio editing software tool, and the channels for left front and left rear were panned to the left, the centre channel was panned to the centre and the channels for right front and right rear were panned to the right. This resulted in a stereo audio reproduction. For realizing mono, the stereo sound was simply exported in mono, making it only one channel in the centre.

The recorded wasp sound was flying around the microphone, resulting in loudness differences in time. These differences were eliminated using audio compression techniques, which reduces the dynamics in audio fragments, resulting in a more constant flying wasp sound.
Appendix E

Source code 3D audio for VRET with visuals
import viz
viz.setMultiSample(16)
viz.fov(90)
viz.go(viz.STEREO)
#   viz.setOption("viz.fullscreen.monitor", 2)
#   viz.go(viz.FULLSCREEN|viz.STEREO_HORIZ)

#Define constants
MOVE_SPEED = 350
TURN_SPEED = 1000

evRight = viz.add(viz.ENVIRONMENT_MAP,'townhall_R.jpg')
skyRight = viz.add('skydome.dlc')
skyRight.texture(envRight)

evLeft = viz.add(viz.ENVIRONMENT_MAP,'townhall_L.jpg')
skyLeft = viz.add('skydome.dlc')
skyLeft.texture(envLeft)

skyLeft.disable(viz.RENDER_RIGHT)
skyRight.disable(viz.RENDER_LEFT)

skyLeft.setEuler([170,0,0])
skyRight.setEuler([170,0,0])

#env = viz.add(viz.ENVIRONMENT_MAP,'townhall_L.jpg')
#sky = viz.add('skydome.dlc')
#sky.texture(env)

#Initialize wasp and place view behind it

wasp = viz.add( 'wasp.ive' )
wasp.setScale(0.7,0.7,0.7)
wasp.setPosition([0,0,10])
#wasp.billboard(viz.BILLBOARD_YAXIS)
viz.MainView.setPosition([0,0,0])

# Far = 1
# Table = 2
# Close back = 3
# Close front = 4

# Balanced latin suare:
#  1 2 3 4 5
#  6 7 8 9 0
# 1 2 3 4 5
# 6 7 8 9 0

# Scenario 1: Far | Table | Close front | Close back
# Scenario 2: Table | Close Back | Far | Close front
# Scenario 3: Close back | Close front | Table | Far
# Scenario 4: Close front | Far | Close back | Table

pathScenario1 = viz.addAnimationPath()
pathScenario2 = viz.addAnimationPath()
pathScenario3 = viz.addAnimationPath()
pathScenario4 = viz.addAnimationPath()

sound1_S = viz.add('Scenario1_Stereo.wav')
sound1_3D = viz.add('Scenario1_3D.wav')
sound2_S = viz.add('Scenario2_Stereo.wav')
sound2_3D = viz.add('Scenario2_3D.wav')

sound3_S = viz.add('Scenario3_Stereo.wav')
sound3_3D = viz.add('Scenario3_3D.wav')

sound4_S = viz.add('Scenario4_Stereo.wav')
sound4_3D = viz.add('Scenario4_3D.wav')

stubs = [
    # close right front
    [0.3,0.5,0],
    [0.2,-0.02,0,1],
    [-0.2,-0.02,-0.1,1],
    [0.1,-0.02,-0.1,1],
    [-0.2,-0.02,0,1],
    [0.2,-0.02,0.1,1],
    [-0.2,-0.02,0,1],
    [0.2,-0.02,0.1,1],
    [-0.2,0.02,-0.1,1],
    [0.2,0.02,0,1],
    [-0.2,0.02,0,1],
    [0.2,-0.02,0,1],
    [-0.2,0.02,0,1],
    [0.2,0.02,0,1],
    [-0.1,0.02,0,1],
    [0.1,0.02,0,1],
    [-0.2,0.02,0,1],
    [0.2,0.02,0,1],
    
    # side
    [-0.1,0,0,0],
    [0.7,-0.1,0.1],
    [-0.7,-0.1,0.1],
    [0.6,-0.1,0.1],
    [-0.8,-0.1,0.1],
    [0.6,-0.1,0.1],
    [-0.6,-0.1,0.1],
    [0.8,-0.1,0.1],
    [-0.6,0.1,0.1],
    [0.6,0.1,0.1],
    [-0.7,0.1,0.1],
    [0.9,-0.1,0.1],
    [-0.6,-0.1,0.1],
    [0.6,0.1,0.1],
    [-0.7,0.1,0.1],
    [0.9,-0.1,0.1],
    
    # roam
    [1,1,0,1],
    [0.5,-1,0.5,1],
    [-0.5,-1,0.5,1],
    [-2,2,0,1],
    [-1,-2,0,1],
    [2,0,0,1],
    [-1,1,0,1],
    [1,0,0,1],
]

def scenario1(x,y,z):

# all increment numbers: x, y, z, s
movement = [
    # far front
    [0.1, 0, 3, 0],
    [-0.3, 0, 0, 1],
    [0.2, -0.01, 0, 1],
    [-0.2, -0.01, 0, 1],
    [0.2, -0.01, 0, 1],
    [-0.2, 0.01, 0, 1],
    [-0.3, -0.01, 0, 1],
    [0.3, -0.01, 0, 1],
    [-0.2, 0.01, 0, 1],
    [0.2, 0.01, 0, 1],
    [-0.3, 0.01, 0, 1],
    [0.2, -0.01, 0, 1],
    [-0.2, 0.01, 0, 1],
    [0.2, 0.01, 0, 1],
    [0.2, -0.01, 0.1],
    [0.2, 0.2, -0.1, 1],
    [0.1, 0.1, -0.1, 1],
    [0.1, 0.05, -0.1, 1],
    [0.1, 0.1, 0, 1],
    [0.1, 0.1, 0.1, 1],
    [0.1, 0.2, 0, 1],
    [0.1, -0.2, -0.1],
    [0.1, -0.1, -0.1, 1],
    [0.2, -0.1, -0.1],
    [0.05, -0.1, 0, 1],
    [0.1, -0.2, -0.1, 1],
    [0.1, -0.05, -0.1],
    [0.1, -0.1, 0, 1],
    [0.1, -0.1, 0.1, 1],
    [0.1, 0.1, 0, 1],
    [0.1, 0.1, 0, 1],
    [0.2, -0.05, 0, 1],
    [-0.4, 0.125, -0.1],
    [-0.2, 0.115, -0.1, 1],
    [-0.2, -0.105, -0.1],
    [-0.3, 0.105, 0.1, 1],
    [-0.4, 0.125, 0.1, 1],
    [-0.5, 0.125, 0.1],
    [0.1, -0.125, 0.1],
    [0.1, -0.125, 0.1],
    [0.2, 0.1, 0.1, 1],
    [0.05, -0.2, 0, 1],
    [0.1, -0.1, -0.1, 1],
    [0.05, 0.1, 0, 1],
    [0.1, 0.2, 0, 1],
    # far to table transition
    [0.2, -0.05, 0.1, 1],
    [-0.1, -0.05, 0.1, 1],
    [0.1, 0.1, 0, 1],
    [-0.1, -0.1, -0.1, 1],
    [0.1, 0.1, 0, 1],
    [0.1, 0.1, -0.1, 1],
]
```python
oldPos = [x,y,z]
positions = []
for x,pos in enumerate(movement):
    if pos[3] == 0:
        newPos = [pos[0],pos[1],pos[2]]
    else:
p
```
#table to close transition
[-0.1, 0.01, 0, 1],
[0.1, 0.01, 0, 1],
[-0.1, 0.02, 0, 1],
[0.1, 0.02, 0, 1],
[-0.1, 0.01, 0, 1],
[0.1, 0.01, 0, 1],
[-0.1, 0.01, 0, 1],
[0.2, 0.01, 0, 1],
[-0.1, 0.01, 0, 1],
[0.2, 0.01, 0, 1],
[-0.05, 0.1, -0.1, 1],
[0.05, 0.1, 0, 1],
[0.05, 0.1, -0.2, 1],
[0.3, 0.05, -0.2, 1],
[-0.1, 0.05, 0, 1],
[0.3, 0.1, -0.2, 1],
[-0.1, 0.1, -0.1, 1],
[0.3, 0.1, 0, 1],
[-0.1, 0, -0.2, 1],

#on right ear
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],
[0.1, 0, 0, 0],

#close right back to far transition
[0.1, 0.01, 0, 1],
[0.1, -0.01, 0.1, 1],
[-0.2, 0.01, 0.1, 1],
[0.1, -0.02, 0.2, 1],
[-0.1, 0.02, 0.1, 1],
[0.1, -0.02, 0.1, 1],
[-0.1, 0.02, 0.1, 1],
[0.2, -0.02, 0.1, 1],
[-0.2, 0.02, 0.1, 1],
[0.1, -0.02, 0.2, 1],
[-0.05, 0.05, 0.1, 1],
[-0.1, -0.05, 0.1, 1],
[0.1, 0.1, 0.1, 1],
[-0.05, -0.1, 0.1, 1],
[0.1, 0.05, 0.1, 1],
[-0.1, -0.05, 0.2, 1],
[0.1, -0.1, 0.1, 1],
oldPos = [x, y, z]
pos = []
for x, pos in enumerate(movement):
    if pos[3] == 0:
        newPos = [pos[0], pos[1], pos[2]]
    else:
pos.append(newPos)
oldPos = newPos
cp = viz.addControlPoint()
```python
cp.setPosition(newPos)
cp.setTranslateMode(viz.CUBIC_BEZIER)
pathScenario2.add(cp,x+1)
wasp.setEuler([170,0,0])

pathScenario2.computeTangents()
pathScenario2.setAutoRotate(viz.OFF)
viz.link(pathScenario2, wasp)
pathScenario2.speed(4)
#pathRoam.setTranslateMode(viz.CUBIC_BEZIER)
pathScenario2.pause()

def scenario3(x,y,z):
    # all increment numbers: x, y, z, s
    movement = [
        #init
        [0.1,0,0,0],
        #init to close right transition
        [-0.1,0.01,0,1],
        [0.1,0.01,0.1,1],
        [-0.1,-0.02,-0.1,1],
        [0.2,0.02,0,1],
        [-0.1,0.01,-0.1,1],
        [0.2,-0.01,-0.1,1],
        [-0.1,0.01,0,1],
        [0.2,-0.01,-0.1,1],
        [-0.1,0.01,0,1],
        [0.3,-0.01,0,1],
        [-0.05,0.1,-0.1,1],
        [0.2,-0.1,0,1],
        [0.1,0.1,-0.2,1],
        [-0.1,-0.05,-0.2,1],
        [0.2,0.05,0,1],
        [-0.1,0.1,-0.2,1],
        [0.4,-0.1,-0.1,1],
        [-0.1,0.1,0,1],
        [0.3,0,-0.2,1],
        #on right ear
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        [0.1,0,0,0],
        #close right back to close left front transition
```
oldPos = \{x,y,z\}

for x,pos in enumerate(movement):
    if pos[3] == 0:
        newPos = \{pos[0],pos[1],pos[2]\}
    else:
        newPos = \{oldPos[0]+pos[0],oldPos[1]+pos[1],oldPos[2]+pos[2]\}

positions.append(newPos)
oldPos = newPos

for x,pos in enumerate(movement):
    if pos[3] == 0:
        newPos = \{pos[0],pos[1],pos[2]\}
    else:
        newPos = \{oldPos[0]+pos[0],oldPos[1]+pos[1],oldPos[2]+pos[2]\}

positions.append(newPos)
oldPos = newPos

for x,pos in enumerate(movement):
    if pos[3] == 0:
        newPos = \{pos[0],pos[1],pos[2]\}
    else:
        newPos = \{oldPos[0]+pos[0],oldPos[1]+pos[1],oldPos[2]+pos[2]\}

positions.append(newPos)
oldPos = newPos

cp = viz.addControlPoint()

cp.setPosition(newPos)

cp.setTranslateMode(viz.CUBIC_BEZIER)

pathScenario3.add(cp,x+1)
```python
wasp.setEuler([170,0,0])

pathScenario3.computeTangents()
pathScenario3.setAutoRotate(viz.OFF)
viz.link(pathScenario3, wasp)
pathScenario3.speed(4)
#pathRoam.setTranslateMode(viz.CUBIC_BEZIER)
pathScenario3.pause()

def scenario4(x,y,z):
    # all increment numbers: x, y, z, s
    movement = [
        [0.1,0,3,0],
        #init
        #init to close left transition
        #init to close left transition
        [-0.1,0.01,0,1],
        [0.1,0.01,-0.1,1],
        [-0.2,-0.02,-0.1,1],
        [0.1,0.02,0,1],
        [-0.2,0.01,-0.1,1],
        [0.1,-0.01,-0.1,1],
        [-0.2,0.01,-0.2,1],
        [0.1,-0.01,-0.1,1],
        [-0.2,0.01,-0.1,1],
        [0.2,-0.01,0,1],
        [-0.2,0.1,-0.1,1],
        [0.1,-0.1,-0.2,1],
        [0.1,0.1,-0.2,1],
        [-0.1,-0.05,-0.2,1],
        [0.1,0.05,0,1],
        [-0.2,0.1,-0.2,1],
        [0.2,-0.1,-0.2,1],
        [-0.3,0.1,-0.1,1],
        [0.1,0,-0.2,1],
        #close front left
        [-0.3,0,0.5,0],
        [0.1,-0.02,0,1],
        [-0.2,0.02,-0.1,1],
        [0.1,-0.02,0,1],
        [-0.1,0.02,0,1],
        [0.2,-0.02,0,1],
        [-0.2,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.1,-0.02,0,1],
        [0.2,0.02,0,1],
        [-0.2,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.1,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.1,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.2,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.1,-0.02,0,1],
        [0.2,0.02,0,1],
        [-0.2,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.1,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.2,-0.02,0,1],
        [0.1,0.02,0,1],
        [-0.2,0.02,0,1],
        #on left ear
        [-0.1,0,0,0],
        [-0.1,0,0,0],
        [-0.1,0,0,0],
        [-0.1,0,0,0],
```
# close left front to far transition
[0,0,0.2,1],
[-0.1,0.05,0.2,1],
[0.2,-0.05,0.1,1],
[-0.2,-0.05,0.2,1],
[0.2,0.05,0.1,1],
[-0.2,0.05,0.1,1],
[0.1,-0.05,0.2,1],
[-0.1,-0.05,0.1,1],
[0.1,0.05,0.2,1],
[-0.1,0.05,0.1,1],
[0.1,-0.05,0.1,1],
[-0.1,0.02,1],
[0.2,0.02,1],
# far front
[0.1,0,3,0],
[-0.3,0,0,1],
[0.2,-0.01,0,1],
[-0.2,-0.01,0,1],
[0.2,-0.01,0,1],
[-0.2,0.01,0,1],
[0.2,0.01,0,1],
[-0.2,0.01,0,1],
# far front roam
[0,0,3,0],
[0.2,0.2,-0.1,1],
[0.5,-0.2,-0.1,1],
[0.1,0.1,-0.1,1],
[0.1,0.05,-0.1,1],
[0.1,0.1,0.1,1],
[0.1,0.1,0.1,1],
[0.1,0.2,0.1],
[0.1,-0.2,-0.1],
[0.1,0.1,-0.1,1],
[0.2,-0.1,-0.1],
[0.05,-0.1,0.1],
[0.1,-0.2,-0.1,1],
[0.1,-0.05,-0.1],
[0.1,0.1,0.1],
[0.1,-0.1,0.1,1],
[0.1,0.1,0.1],
[0.1,0.1,0.1],
# far to close right transition
oldPos = [x, y, z]
positions = []
for x, pos in enumerate(movement):
    oldPos = [x, y, z]
if pos[3] == 0:
    newPos = [pos[0], pos[1], pos[2]]
else:
positions.append(newPos)
oldPos = newPos
cp = viz.addControlPoint()
cp.setPosition(newPos)
cp.setTranslateMode(viz.CUBIC_BEZIER)
pathScenario4.add(cp, x+1)
wasp.setEuler([170, 0, 0])

pathScenario4.computeTangents()
pathScenario4.setAutoRotate(viz.OFF)
viz.link(pathScenario4, wasp)
pathScenario4.speed(4)
#pathRoam.setTranslateMode(viz.CUBIC_BEZIER)
pathScenario4.pause()

scenario1(0, 0, 0)
scenario2(0, 0, 0)
scenario3(0, 0, 0)
scenario4(0, 0, 0)
#side(0, 0, 0)

def playScenario1_S():
    pathScenario1.play()
sound1_S.play()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()

    sound2_S.stop()
sound3_S.stop()
sound4_S.stop()
sound1_3D.stop()
sound2_3D.stop()
sound3_3D.stop()
sound4_3D.stop()

def playScenario2_S():
    pathScenario2.play()
sound2_S.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()

    sound1_S.stop()
sound3_S.stop()
sound4_S.stop()
def playScenario3_S():
    pathScenario3.play()
    sound3_S.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario4.pause()
    pathScenario4.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()

def playScenario4_S():
    pathScenario4.play()
    sound4_S.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()

def playScenario1_3D():
    pathScenario1.play()
    sound1_3D.play()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
def playScenario2_3D():
    pathScenario2.play()
    sound2_3D.play()

    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()

    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop()

    sound1_3D.stop()
    sound2_3D.stop()
    sound3_3D.stop()
    sound4_3D.stop()

def playScenario3_3D():
    pathScenario3.play()
    sound3_3D.play()

    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario4.pause()
    pathScenario4.reset()

    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop()

    sound1_3D.stop()
    sound2_3D.stop()
    sound3_3D.stop()
    sound4_3D.stop()

def playScenario4_3D():
    pathScenario4.play()
    sound4_3D.play()

    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()

    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
def playScenario1_N():
    pathScenario1.play()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop()
    sound1_3D.stop()
    sound2_3D.stop()
    sound3_3D.stop()
    sound4_3D.stop()

def playScenario2_N():
    pathScenario2.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    pathScenario4.pause()
    pathScenario4.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop()
    sound1_3D.stop()
    sound2_3D.stop()
    sound3_3D.stop()
    sound4_3D.stop()

def playScenario3_N():
    pathScenario3.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario4.pause()
    pathScenario4.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop
sound1_3D.stop()
sound2_3D.stop()
sound3_3D.stop()
sound4_3D.stop()

def playScenario4_N():
    pathScenario4.play()
    pathScenario1.pause()
    pathScenario1.reset()
    pathScenario2.pause()
    pathScenario2.reset()
    pathScenario3.pause()
    pathScenario3.reset()
    sound1_S.stop()
    sound2_S.stop()
    sound3_S.stop()
    sound4_S.stop()
    sound1_3D.stop()
    sound2_3D.stop()
    sound3_3D.stop()
    sound4_3D.stop()

vizact.onkeydown('1',playScenario1_N)
vizact.onkeydown('2',playScenario2_N)
vizact.onkeydown('3',playScenario3_N)
vizact.onkeydown('4',playScenario4_N)

vizact.onkeydown('q',playScenario1_S)
vizact.onkeydown('w',playScenario2_S)
vizact.onkeydown('e',playScenario3_S)
vizact.onkeydown('r',playScenario4_S)

vizact.onkeydown('a',playScenario1_3D)
vizact.onkeydown('s',playScenario2_3D)
vizact.onkeydown('d',playScenario3_3D)
vizact.onkeydown('f',playScenario4_3D)


[38] D. R. Begault, “AUDITORY AND NON-AUDITORY FACTORS THAT POTENTIALLY INFLUENCE VIRTUAL ACOUSTIC IMAGERY,”


[43] T. Schubert, “Embodied Presence in Virtual Environments 1 The role of psychology in virtual environments 2 Psychological concepts of presence,”


[106] Peter, “ABX Comparator component for Foobar.”


[129] B. Orange, “No Title.”


[133] Y. I. Nobuyuki Ono, Yoshihisa Hayashi, Atsushi Kisuki, “Anechoic chamber and wave absorber.”


[136] SoundLocus, “3D Sound Bee demo.”


[138] CLAM, “CLAM Website.”

[139] U. P. F. Music Technology Group, “CLAM youtube demo,”


[142] Longcat, “Demo, LongCat3DAudio.”

---

Master of Science Thesis A.R.D. Hoekstra