# Virtual Reality Games: Wie Sound Design die Richtungslokalisation verbessert

vorgelegt von Michael Dmoch an der Hochschule der Medien Stuttgart · am 30. März 2017 zur Erlangung des akademischen Grades B.Eng. Audiovisuelle Medien

Erstprüfer: Prof. Oliver Curdt

Zweitprüfer: Prof. Uwe Schulz

# **Abstract**

This work shows what aspects of sound design have an impact on localization in virtual reality games that use binaural synthesis. In current research of binaural hearing, it examines those factors that can be influenced in the sound design process.

A game scenario, derived from VR and gaming theory frameworks, is being used to demonstrate a possible sound design process with localization factors in mind. The resulting sounds are then tested in a VR lab and for the most part confirm the effectiveness of the process.

# Zusammenfassung

Die vorliegende Arbeit zeigt auf, welche Aspekte im Sound Design sich auf die Lokalisation in Virtual Reality Games, die Binauralsynthese einsetzen, auswirken. Aktuelle Forschung zum binauralen Hören wird auf solche Faktoren hin untersucht, die im Sound Design beeinflussbar sind.

Ein Spielszenario, erarbeitet aus Frameworks der VR und Spieltheorie, wird genutzt, um einen möglichen Gestaltungsprozess vorzuführen, bei dem Lokalisationsfaktoren maßgeblich sind. Die so entstehenden Klänge werden in einem VR Labor getestet und bestätigen überwiegend die Effektivität der Vorgehensweise.

# Eidesstattliche Erklärung

Hiermit versichere ich, Michael Dmoch, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: "Virtual Reality Games: Wie Sound Design die Richtungslokalisation verbessert" selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

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

Stuttgart, den 29. März 2017		
Michael Dmoch		

# Inhaltsverzeichnis

A	bstract		
Z	usamme	enfassung	
E	idesstat	tliche Erklärung	11
Α	bkürzur	ngsverzeichnis	ν
Α	bbildun	gsverzeichnis	V
T	abellen	verzeichnis	V
1	Einle	eitung	1
2	Räu	mliches Hören	2
	2.1	Interaurale Cues	2
	2.2	Monaurale Cues	
	2.3	Entfernungshören	
	2.4	Kopfbewegungen	
	2.5	Visuelle Cues	
	2.6	Im-Kopf-Lokalisation (IKL)	3
	2.7	Vertauschungen	3
3	Bina	ouralsynthese	4
	3.1	Quellsignale	4
	3.2	HRTFs	5
	3.3	Faltung	6
	3.4	Raumreflexionen	7
	3.5	Kopfhörerkalibrierung	7
	3.6	Dynamische Binauralsynthese	8
4	Wirl	kung der Klangeigenschaften auf die Lokalisation	9
	4.1	Effekte des Signalspektrums	9
	4.2	Effekte der Hüllkurve	12
	4.3	Signaldauer	
	4.4	Vorkenntnis & Lerneffekte	
	4.5	Zusammenfassung	15
5	Virt	ual Reality Games	16
	5.1	Technik der VR	16
	5.2	Virtual Environments (VE)	
	5.3	Erlebnis der VR	18
	5.4	Games	
	5.5	Zusammenfassung	20

6 S	Szenario	21
6.1	Kriterien	21
6.2	Auswahl eines VR-Games	22
6.3	Definition der erfolgreichen Lokalisation	24
7 S	Sounddesign	26
7.1	Software	26
7.2		
7.3	Spektrale Bearbeitung	27
7.4	Amplitudenmodulation	32
7.5	Zusammenfassung	37
8 E	rhebung	38
8.1	Eigenschaften der Erhebung	38
8.2		
8.3	Durchführung	48
8.4	Datenverarbeitung	49
8.5	Ergebnisse & Interpretation	51
8.6	Fazit	56
9 Z	Zusammenfassung	58
Literat	turverzeichnis	59
Anhan	ng / Datenträger	62

# Abkürzungsverzeichnis

**BRIR** Binaural Room Impulse Response

**DAW** Digital Audio Workstation

**EQ** Equalizer

**FFT** Fast Fourier Transformation

**HdM** Hochschule der Medien

**HE** Hörereignis

**HER** Hörereignisrichtung

**HMD** Head Mounted Display

**HpTF** Headphone Transfer Function

**HRIR** Head Related Impulse Response

**HRTF** Head Related Transfer Function

IKL Im-Kopf-Lokalisation

**ILD** Interaural Level Differences

ITD Interaural Time Differences

**ITD-HF** Interaural Time Differences in High Frequencies

**ITD-LF** Interaural Time Differences in Low Frequencies

NPC Non Player Character

**RPG** Role Playing Game

**SE** Schallereignis

**SER** Schallereignisrichtung

**SWE** Selbstwirksamkeitserleben

**VAD** Virtual Auditory Display

**VE** Virtual Environment

**VP** Versuchsperson

VR Virtual Reality

VST Virtual Studio Technology

# Abbildungsverzeichnis

Abbildung 3: Spektrum von Klang C
Abbildung 4: Spektrum von Klang D
Abbildung 5: Spektrum von Klang E
Abbildung 6: Spektrum von Klang F
Abbildung 7: Transienten des Helikopter-Samples
Abbildung 8: Spektrum von Klang G
Abbildung 9: Wellenform von Klang G
Abbildung 10: Wellenform von Klang H
Abbildung 11: Spektrum von Klang H
Abbildung 12: Spektrum von Klang I
Abbildung 13: Wellenform von Klang I
Abbildung 14: Versuchsanordnung der virtuellen Schallquellen in der Medianebene 40
Abbildung 15: Versuchsanordnung der virtuellen Schallquellen in der Horizontalebene 40
Abbildung 16: Das reale Labor
Abbildung 17: Das VR-Labor
Abbildung 18: Fehllokalisationen
Abbildung 19: Szenario-relevante Fehllokalisationen
Abbildung 20: Vorne-Hinten-Vertauschungen
Abbildung 21: Oben-Unten-Vertauschungen
Abbildung 22:Hörereignisse ohne Richtung
Abbildung 23: Hörereignisse mit mehrdeutiger Richtung
Tabellenverzeichnis
Tabelle 1: Vorne-Hinten-Lokalisation bei gefiltertem Rauschen
Tabelle 2: Vorauswahl VR-Games im Gameslab

# 1 Einleitung

Virtual Reality (VR) Games steht in den Startlöchern, zum Massenprodukt der nächsten Jahre zu werden. Beim überwiegenden Teil dieser Games wird der Ton über Kopfhörer konsumiert, wobei Binauralsynthese räumliches Hören ermöglicht. Dies ist für ein überzeugendes VR-Erlebnis essentiell.

Ein zentrales Qualitätsmerkmal der Binauralsynthese ist, wie gut sich Hörereignisse lokalisieren lassen. Dies hängt von einer Reihe von Faktoren ab. Aktuelle Forschung dazu konzentriert sich vorwiegend auf die Faktoren der Verarbeitungs- und Wiedergabetechniken.

Obwohl auch die Klangeigenschaften der Inhalte die Lokalisation beeinflussen, gibt es unter Sound Designern noch wenig Bewusstsein dafür, was das für ihre gestalterische Arbeit bedeutet.

Deshalb stellt die vorliegende Arbeit als zentrale Forschungsfrage:

Was können Sound Designer beachten, damit ihre Klänge in VR-Games besser lokalisiert werden?

Zunächst stellt die vorliegende Arbeit die Grundlagen des räumlichen Hörens in der Natur sowie die Binauralsynthese vor.

Der Stand der Forschung zum binauralen Hören soll dahingehend betrachtet werden, welche Wirkung Klangeigenschaften auf die Lokalisation haben können.

Außerdem wird besprochen, welche Rolle Ton bei den Mechanismen der VR-Games spielt. Daraus soll ein Spielszenario entwickelt werden.

Für dieses Spielszenario wird anschließend ein Sounddesign erarbeitet, bei dem die Erkenntnisse zur Wirkung der Klangeigenschaften maßgeblich sein sollen.

Inwiefern die dadurch entstandenen Klänge besser lokalisierbar sind, soll letztendlich in einer Erhebung in VR geprüft werden.

\_

<sup>&</sup>lt;sup>1</sup> vgl. Statista GmbH Hamburg 2016, S. 6–18

# 2 Räumliches Hören

Es wird unterschieden zwischen dem Schallereignis (SE), also dem physikalischen, objektiven, messbaren Entstehen von Schallwellen, und dem Hörereignis (HE), also der sensorischen, subjektiven Wahrnehmung derer.<sup>2</sup>

Lokalisation benötigt Cues: Merkmale des wahrgenommenen Schalls, von denen aus man auf die Richtung und Entfernung des HE schließen kann.

#### 2.1 Interaurale Cues

Bei einem SE seitlich vom Kopf entstehen Unterschiede zwischen den Ohrsignalen. Das Gehör eines Menschen kann diese Unterschiede verarbeiten, und daraus darauf schließen, ob sich ein SE links oder rechts von ihm befindet. Man unterscheidet zwischen Laufzeit- und Pegelunterschieden.

Laufzeitunterschiede (*Interaural Time Differences*, ITD) existieren, wenn der Schall das zugewandte Ohr früher erreicht als das abgewandte.

Pegelunterschiede (*Interaural Level Differences*, ILD) entstehen, weil bei einem seitlichen Schalleinfall der Kopf das abgewandte Ohr akustisch teilweise abschattet.

Die Verarbeitung der interauralen Cues ist geteilt in zwei voneinander unabhängige Mechanismen:<sup>3</sup>

Der eine wertet den Zeitversatz der Schwingungen der Ohrsignale aus, vergleicht also ihre Phasen. Dieser Mechanismus funktioniert jedoch nur für den Teil des Signals unterhalb von etwa 1,6 kHz.

Der andere arbeitet nur oberhalb von etwa 1,6 kHz und wertet dort die ILD und ITD aus. Im Falle der ITD werden jedoch nicht die Phasen verglichen, sondern der Zeitversatz der Hüllkurven des Signals.

#### 2.2 Monaurale Cues

Dadurch, dass Schallwellen auf dem Weg zum Ohr vom Torso, Kopf und besonders den Außenohren (lat. *Pinna*) reflektiert und gebeugt werden, entstehen Veränderungen im Frequenzspektrum, die sog. monauralen oder auch spektralen Cues.

Sie sind ausschlaggebend beim Hören in der Medianebene<sup>4</sup>, also der Unterscheidung zwischen oben und unten, sowie vorne und hinten.

# 2.3 Entfernungshören

Das Einschätzen der Entfernung eines HE basiert auf dessen Lautstärke, die sich mit der Entfernung verringert; auf dem Spektrum, da HE mit der Entfernung Höhen verlieren; außerdem auf den zeitlichen Eigenschaften der frühen Raumreflexionen und dem Verhältnis des Direktsignals zu den diffusen Raumreflexionen.<sup>5</sup>

<sup>&</sup>lt;sup>2</sup> vgl. Dickreiter et al. 2014, S. 118

<sup>&</sup>lt;sup>3</sup> vgl. Blauert 1974, S. 139

<sup>&</sup>lt;sup>4</sup> Oft wird auch von Saggitalebenen gesprochen; Die Medianebene ist diejenige Saggitalebene, die durch die Mitte des Kopfes verläuft.

<sup>&</sup>lt;sup>5</sup> vgl. Dickreiter et al. 2014, S. 132

## 2.4 Kopfbewegungen

Ein Kopf lässt sich um drei Achsen rotieren: Drehen, Nicken und Kippen. Um besser zu lokalisieren, führt ein Mensch unbewusst kleine Kopfrotationen aus; sein Gehör lässt dann die Unterschiede in den Cues, die dadurch entstehen, und seine Propriozeption<sup>6</sup> in die Lokalisation mit einfließen.<sup>7</sup>

#### 2.5 Visuelle Cues

Lokalisation ist ein multimodaler Vorgang. Ist das Objekt oder Ereignis, von dem das SE ausgeht, zu sehen, ist das ein starker Cue für die Lokalisation, der deutlich über die akustischen Cues dominiert.

Wird dabei das HE an einer anderen Stelle lokalisiert als das visuelle Ereignis, tritt der sog. Bauchredner-Effekt ein: Die Diskrepanz wird von der Wahrnehmung aufgelöst, indem der Hörereignisort so umgedeutet wird, dass sich die beiden decken. Es wurde beobachtet, dass dies noch bei Abständen von über 30° auftritt.<sup>8</sup>

## 2.6 Im-Kopf-Lokalisation (IKL)

Die IKL beschreibt den Effekt, dass der Hörereignisort als innerhalb des Kopfes befindlich wahrgenommen wird.

Er tritt auf, wenn die für das Entfernungshören relevanten Cues widersprüchlich sind und sich nicht eindeutig auswerten lassen. Somit kann man die IKL wie eine Lokalisation mit einer Entfernung von 0 betrachten.

Die Wissenschaft ist sich uneinig, ob Kopfbewegungen bei der Minimierung der IKL helfen.<sup>9</sup>

## 2.7 Vertauschungen

Während die Lokalisation über temporale Cues relativ zuverlässig ist, ist es die über spektrale Cues weniger. So kann es zu der Situation kommen, dass die Links-Rechts-Lokalisierung (Lateralisation) zustande kommt, gleichzeitig aber falsch zwischen vorne und hinten, und/oder auch zwischen oben und unten unterschieden wird.

Dieses Phänomen nennt man Vorne-Hinten-Vertauschung bzw. Oben-Unten-Vertauschung. Es kann durch Kopfbewegungen gemindert werden. 10

<sup>10</sup> vgl. Wenzel 1993, S. 122

<sup>&</sup>lt;sup>6</sup> Propriozeption ist der Sinn für Bewegung und Lage des Körpers im Raum oder von Körperteilen zueinander.

<sup>&</sup>lt;sup>7</sup> vgl. Blauert & Braasch 2008, S. 88

<sup>&</sup>lt;sup>8</sup> vgl. Thurlow 1973, nach Brimijoin et al. 2013, S. 8

<sup>&</sup>lt;sup>9</sup> vgl. ebd., S. 2

# 3 Binauralsynthese

Binauralsynthese ist ein Verfahren, ein Signal, dessen Klang keine binauralen Richtungsinformationen besitzt, mit solchen zu versehen.

Dabei wird das Eingabesignal mit den der gewünschten Richtung entsprechenden *Head Related Transfer Function* (HRTF) mittels Faltung verrechnet, jeweils eine HRTF pro Ohr und Richtung. <sup>11</sup> Um den Rechenaufwand geringer zu halten, wird vor der Faltung eine invertierte Fouriertransformation auf das Signal angewendet, weil Faltung in der Frequenzdomäne aus einer einfachen Multiplikation besteht; nach der Faltung wird das Signal wieder fouriertransformiert.

Diese Technik bereichert das Eingabesignal um den erwünschten Richtungseindruck. Das Ergebnis ist ein zweikanaliges Signal, das über Kopfhörer wiedergegeben wird.

## 3.1 Quellsignale

Als Quellsignale eignen sich prinzipiell Mono-Signale, Mehrkanal-Signale und Ambisonics. Kunstkopfaufnahmen eignen sich dagegen nicht, da sie bereits binaural sind und sich die binaurale Information auch nicht wieder entfernen oder verändern lässt.

#### 3.1.1 Mono

Ein Signal in Mono stellt den einfachsten Fall dar: Es wird, durch die Binauralsynthese mit Richtungsinformation versehen, zur Punktschallquelle im binauralen Raum.

#### 3.1.2 Mehrkanalsignale

Prinzipiell wird ein Mehrkanalsignal einfach als mehrere Mono-Signale behandelt, die über virtuelle, um den Hörer aufgestellte Lautsprecher wiedergegeben werden, die jeweils eine separat binauralisierte Punktschallquelle darstellen.

Da Mehrkanalaufnahmen fast immer für ein gängiges Surround-Format wie z.B. Dolby 5.1 aufgenommen wurden, werden die virtuellen Lautsprecher dem Standard entsprechend aufgestellt; damit einher geht, je nach Format, dass Richtungsinformation in der Vertikalen entweder gar nicht vorhanden ist, oder schlechter aufgelöst als in der Horizontalebene. Außerdem werden die Richtungen zwischen den Lautsprechern unschärfer abgebildet als die, die sich mit den Lautsprechern decken.

Mit Stereosignalen wird ähnlich verfahren; die beiden virtuellen Lautsprecher werden im sog. Stereo-Dreieck zum Hörer positioniert. Hier fehlt nicht nur die Höheninformation; das Signal kann grundsätzlich nur aus den Richtungen der virtuellen Lautsprecher und denen zwischen ihnen kommen.

#### 3.1.3 Ambisonics

Ambisonics-Aufnahmen müssen für ihr Ausgabeformat gerendert werden; für räumliche Wiedergabe über Kopfhörer bedürfen sie also auch der Binauralsynthese. Dazu werden die Daten als Zwischenschritt zunächst für eine beliebige Lautsprecheraufstellung zu mehreren Kanälen gerendert;

<sup>&</sup>lt;sup>11</sup> vgl. Nicol 2010, S. 30

der weitere Prozess entspricht dem für Mehrkanalsignale. <sup>12</sup> Im Gegensatz zu diesen ist hier die Verteilung der virtuellen Lautsprecher jedoch gleichmäßig und ihre Anzahl theoretisch frei wählbar – wenn auch abgewägt werden muss zwischen der Richtungsauflösung der Quelle (die bei First Order Ambisonics geringer ausfällt als bei Higher Order Ambisonics), der dafür sinnvollen Anzahl an virtuellen Lautsprechern und der zur Synthese zur Verfügung stehenden Rechenkapazität. <sup>13</sup>

#### 3.1.4 Eignung für Games

Mehrkanalaufnahmen und Ambisonics besitzen bereits seit der Aufnahme bzw. dem Mixdown Richtungsinformationen, über das Panning der Phantomschallquellen zwischen den Kanälen sind sie fest ins Signal geprägt.

Dadurch lassen sie sich in der Binauralsynthese höchstens noch um die Hörposition rotieren, das verschieben einzelner Teile ist nachträglich nicht mehr möglich. Für z.B. Kinofilme oder Musikproduktionen ist das unproblematisch; anders ist es jedoch bei interaktiven Medien, die dynamisch zur Laufzeit generiert werden.

Hier hat sich deshalb diese Herangehensweise etabliert:

- Mono-Signale werden für diskrete Ereignisse, Objekte und Akteure im Spiel verwendet; also alles, was eine diskrete Punktschallquelle darstellt, die jederzeit beliebig und unabhängig von anderen Schallquellen im Raum bewegt werden soll.
- Mehrkanalformate<sup>14</sup> und zunehmend Ambisonics werden auch genutzt, jedoch für Atmos; also immer dann, wenn nur Information über eine Umgebung vermittelt und ein Gefühl des diffusen "Umhülltsein mit Klang" erzeugt werden soll, aber einfache Rotation des Gesamtsignals mit der Kopfrichtung genügt.

#### 3.2 HRTFs

Die HRTF ist definiert als das Verhältnis der Fouriertransformation des Schalldrucks am Ohr zur Fouriertransformation des Schalldrucks an der Stelle, an der die Mitte des Kopfes wäre, wenn der Hörer nicht anwesend wäre. Sie beschreibt also die Unterschiede, die durch Pinnae, Kopf und Torso des Hörers entstehen, wobei sie sowohl spektrale Unterschiede, als auch über ihre Phase die ITD abbildet. 16

Sollen HRTFs gemessen werden, stellt man eine Person in einen reflexionsarmen Raum und führt ihr jeweils ein kleines Mikrofon in die beiden Ohrkanäle ein. Nun beschallt man sie mit einem breitbandigen Impuls aus einem Lautsprecher. <sup>17</sup> So erhält man zwei Impulsantworten, die auch als Head-Related Impulse Responses (HRIR) bezeichnet werden. Durch Fouriertransformation der HRIRs erhält man ein Paar HRTFs. Dieses ist spezifisch für das Gehör der gemessenen Person; und für genau eine Richtung - der Prozess muss also für eine Vielzahl von Richtungen wiederholt werden. <sup>18</sup>

<sup>12</sup> vgl. Algazi & Duda 2011, S. 39

<sup>&</sup>lt;sup>13</sup> vgl. ebd., S. 39

<sup>&</sup>lt;sup>14</sup> Trotz genannter Nachteile vor allem wegen dem dank der Filmindustrie sehr großen Angebot hochwertiger Mehrkanalaufnahmen.

<sup>&</sup>lt;sup>15</sup> vgl. ebd., S. 35

<sup>&</sup>lt;sup>16</sup> vgl. Nicol 2010, S. 32

<sup>&</sup>lt;sup>17</sup> Blauert & Braasch 2008, S. 90

<sup>&</sup>lt;sup>18</sup> Somit erhält man damit einen ganzen Datensatz vieler HRTFs, im allgemeinen Sprachgebrauch und in dieser Arbeit wird aber der Einfachheit halber oft HRTF geschrieben, wenn eigentlich ein ganzer solcher Datensatz gemeint ist.

#### 3.2.1 Nichtindividuelle HRTFs

Mit individuell gemessenen HRTFs wurde eine hohe Genauigkeit in der Lokalisation festgestellt. <sup>19</sup> Der Prozess ist jedoch aufwändig: Typische HRTF-Datenbanken beinhalten Messungen für 700 bis 2500 Richtungen. <sup>20</sup> Da sie damit für den Massenmarkt wenig geeignet sind, werden üblicherweise, und auch in Binauralsyntheselösungen gängiger VR-Games, sogenannte nichtindividuelle HRTFs verwendet. Dabei handelt es sich um "allgemeingültige" HRTFs, die z.B. über das Mitteln vieler verschiedener individueller HRTFs erlangt werden. In anderen Worten: Sie erzielen für die meisten Hörer passable, aber keine optimalen Ergebnisse.

Aber auch mit nichtindividuellen HRTFS kann man, mit Einschränkungen, gut lokalisieren. <sup>21</sup> Die interauralen Cues liefern Ergebnisse, die mit denen von individuellen HRTFs vergleichbar sind, während spektrale Cues jedoch schlechter abschneiden. <sup>22</sup> So verschlechtert sich die Lokalisation in der Medianebene, und tritt das Phänomen der Vorne-Hinten-Vertauschungen deutlich häufiger auf als bei individualisierten HRTFs. <sup>23</sup> Am stärksten ausgeprägt sind die Probleme bei den Cues, die durch das Außenohr erzeugt werden, also oberhalb von etwa 3kHz, weil sich das Außenohr besonders stark von Mensch zu Mensch unterscheidet. <sup>24</sup>

#### 3.2.2 Glättung

Damit die Signalfärbung durch die HRTF weniger auffällig ist, werden HRTFs auch oft geglättet. Es konnte gezeigt werden, dass eine Glättung bis zu einer halben Oktave sich nicht merklich auf die Lokalisation auswirkt.<sup>25</sup>

#### 3.2.3 Interpolation

Auch wenn HRTFs tausende gemessene Richtungen beinhalten, muss zwischen ihnen interpoliert werden, sobald sich ein SE und/oder der Kopf bewegen.<sup>26</sup>

Empfehlungen für eine ausreichende Anzahl von diskreten Messpunkten für HRTFs, zwischen denen interpoliert wird, reichen von 150 (entspricht etwa Abständen zwischen den diskreten Richtungen von 20°) bis 1100.<sup>27</sup>

Interpolation hat kaum negative Auswirkungen auf Genauigkeit der Lokalisation, aber kann Vorne-Hinten-Vertauschungen häufiger auftreten lassen.<sup>28</sup>

# 3.3 Faltung

Hauptprobleme des Rendering mittels HRTF sind Rechenleistung und Latenz. 29

Wird eine Faltung statt in der Zeit- in der Frequenzdomäne durchgeführt, besteht sie aus einer einfachen Multiplikation und stellt dadurch deutlich weniger Anforderungen an Rechenleistung, wenngleich auch etwas mehr Latenz entsteht.

<sup>&</sup>lt;sup>19</sup> vgl. Wightman & Kistler 1989, S. 866

<sup>&</sup>lt;sup>20</sup> vgl. Anish Chandak 2014, S. 29

<sup>&</sup>lt;sup>21</sup> vgl. Wenzel 1993, S. 121

<sup>&</sup>lt;sup>22</sup> vgl. ebd., S. 111

<sup>&</sup>lt;sup>23</sup> vgl. Møller et al. 1996, S. 463-464

<sup>&</sup>lt;sup>24</sup> vgl. Algazi & Duda 2011, S. 34–35

<sup>&</sup>lt;sup>25</sup> vgl. Nicol 2010, S. 31

<sup>&</sup>lt;sup>26</sup> vgl. ebd., S. 43

<sup>&</sup>lt;sup>27</sup> vgl. ebd., S. 44

<sup>&</sup>lt;sup>28</sup> vgl. ebd., S. 44

<sup>&</sup>lt;sup>29</sup> vgl. Algazi & Duda 2011, S. 38

Deshalb wird das Signal, das binauralisiert werden soll, vor der Faltung per Fouriertransformation in die Frequenzdomäne überführt, dann mit der HRTF multipliziert, und anschließend per inverser Transformation zurück in die Zeitdomäne gebracht, um das binauralisierte Signal zu erhalten.

#### 3.4 Raumreflexionen

Ein binauralisiertes HE hat zwar eine Richtungsinformation, entspricht aber dem Hören in einem reflexionsarmen Raum, was weder natürlich noch angenehm ist. Und gerade das Erkennen der Entfernung eines SE gelingt erst mittels Reflexionen an Wänden und Objekten, denn ihr Fehlen deutet unser Gehör als sehr geringe Entfernung vom SE. In reflexionsreichen Umgebungen funktioniert wiederum die Richtungslokalisation weniger gut.<sup>30</sup>

Nicht zuletzt vermitteln Reflexionen dem Hörer Informationen über die akustischen Eigenschaften seiner Umgebung.

In den meisten Audiolösungen, die auf Binauralsynthese basieren, werden die frühen Reflexionen entweder per Physical Modelling berechnet, oder per Faltung des Direktsignals mit einer Binauralen Raumimpulsanwort (Binaural Room Impulse Response, BRIR) erzeugt.<sup>31</sup>

Die BRIR wird gemessen, ähnlich einer HRIR, und gibt dann spezifisch die akustischen Eigenschaften nur des gemessenen Raums wieder, und ist somit erheblich weniger flexibel in dynamischen Anwendungen. Auch ist der diffuse Nachhall in der BRIR bereits enthalten. Dadurch ist die BRIR um ein vielfaches länger als eine HRTF, und damit die Anforderungen an die Rechenkapazität deutlich höher, weshalb für Echtzeitanwendungen auf Consumer-PCs meist von dieser Lösung abgesehen wird.<sup>32</sup>

Im Fall des Physical Modelling dagegen können die frühen Reflexionen unter Berücksichtigung der geometrischen Eigenschaften jedes Raumes berechnet werden.

In diesem Fall wird der Nachhall separat mittels eines einfachen Hallalgorithmus erzeugt, wie er in vielen anderen Medien mit nichtbinauralem Ton oft eingesetzt wird. Da er keine Richtungsinformation besitzt, ergibt es keinen Sinn, ihn zu binauralisieren.

# 3.5 Kopfhörerkalibrierung

Ein weiterer Aspekt eines gut funktionierenden Binauraltons über Kopfhörer ist dessen Kalibrierung.

Die Kopfhörertransferfunktion (Headphone Transfer Function, HpTF) beschreibt Verzerrungen, die aufgrund des Frequenzgangs des Kopfhörers sowie der akustischen Kopplung zwischen Kopfhörer und Ohr entstehen.<sup>33</sup>

Als Lösung dient die Kopfhörerkalibrierung: Bei ihr wird die HpTF gemessen; durch invertierte Filterung der Kopfhörersignale mit ihr werden fast alle beschriebenen Effekte ausgeglichen.

Die HpTF ist spezifisch für die gemessene Person in Kombination mit dem gemessenen Kopfhörer. Somit ist der Aufwand der Kalibrierung hoch, und gilt als ungeeignet für den Massenmarkt.

Ohne diese Kalibrierung können jedoch mehr Fehllokalisationen des Erhebungswinkels und mehr Vorne-Hinten-Vertauschungen auftreten.

<sup>30</sup> vgl. Anish Chandak 2014, S. 17

<sup>&</sup>lt;sup>31</sup> vgl. Xie et al. 2013, S. 369

<sup>32</sup> vgl. Algazi & Duda 2011, S. 37-38

<sup>&</sup>lt;sup>33</sup> vgl. Nicol 2010, S. 5

Ein Faktor, der jedoch nicht durch die Kalibrierung ausgeglichen werden kann, ist, wie die Kopfhörermuschel auf dem Außenohr sitzt, was bei jedem Außsetzten anders sein kann. Dadurch entstehen Verzerrungen, die bei Frequenzen unter 6 kHz moderat bleiben, darüber aber eine Standardabweichung von 10 dB verursachen können. Dafür gibt es bisher keine praktische Lösung; immerhin sind ohrumschließende Kopfhörer weniger empfänglich für den Effekt als aufliegende.<sup>34</sup>

## 3.6 Dynamische Binauralsynthese

Die bisher beschriebene Binauralsynthese ist statisch: Die Klangszene folgt starr jeder Drehung des Kopfes mit. Für einige Anwendungen außerhalb von VR kann das noch akzeptabel sein, es ist jedoch als Hörsituation ausgesprochen unrealistisch. Sollen Schallquellen ihre Position im Raum halten, ist der Einsatz von Head-Tracking erforderlich, damit nutzt man dann die sog. Dynamische Binauralsynthese. 35

Beim Head-Tracking wird fortlaufend die Rotation und meist auch Position des Kopfes festgestellt. <sup>36</sup> Wird die Position über einen Bereich gemessen, der so groß ist, dass der Nutzer sich einige Schritte weit frei bewegen kann, spricht man auch von Room-Tracking.

Nicht nur wird die Klangszene durch Head-Tracking realistischer, auch lassen sich erst dadurch die positiven Effekte von Kopfbewegungen auf die Richtungslokalisation auch für das Hören eines binauralen Kopfhörersignals nutzen.

Spezifisch bei Binauralsynthese (mit nichtindividuellen HRTFs) reduziert Head-Tracking die Häufigkeit von Vorne-Hinten-Vertauschungen von 50% auf 28%.<sup>37</sup> Ob sich die IKL ebenfalls verringert, ist umstritten.<sup>38</sup> Außerdem mildert Head-Tracking die negativen Effekte ab, die nichtindividuelle HRTFs mit sich bringen.<sup>39</sup>

Aber auch eine Verschlechterung durch Head-Tracking kann eintreten: Thaner stellte fest, dass Nickbewegungen des Kopfes zu schlechterer Lokalisation im hinteren Teil der Medianebene (90° bis 210°) führen.<sup>40</sup>

Insgesamt überwiegen jedoch die Vorteile von Head-Tracking eindeutig. Außerdem lässt sich anmerken, dass die Feststellung von Ausrichtung und Position des Spielerkopfs für die Grafikausgabe (und meist auch für das Gameplay) Voraussetzung ist. Somit ist Head- bzw. Room-Tracking ein essentielles Element von VR und deshalb standardmäßig in HMDs vorhanden.

<sup>35</sup> vgl. ebd., S. 42

<sup>&</sup>lt;sup>34</sup> vgl. ebd., S. 5

<sup>&</sup>lt;sup>36</sup> vgl. VRNerds Blog 2016

<sup>&</sup>lt;sup>37</sup> vgl. Xie et al. 2013, S. 374

<sup>&</sup>lt;sup>38</sup> vgl. Brimijoin et al. 2013, S. 2

<sup>&</sup>lt;sup>39</sup> vgl. Algazi & Duda 2011, S. 40

<sup>&</sup>lt;sup>40</sup> vgl. Thaner 1999, S. 50

# 4 Wirkung der Klangeigenschaften auf die Lokalisation

Während in wissenschaftlichen Untersuchungen weitestgehend einfache Signale wie Rauschen oder Sinustöne verwendet werden, kommen im Sounddesign vorwiegend komplexe Signale zum Einsatz, die sich zeitlich, dynamisch und spektral stark unterscheiden können. In diesem Abschnitt soll näher betrachtet werden, bei welchen dieser Klangeigenschaften Effekte auf die Lokalisation zu erwarten sind.

## 4.1 Effekte des Signalspektrums

Blauert konstatiert: Es sei für die Lokalisation in der Medianebene wesentlich, dass das Signalspektrum eine möglichst große Bandbreite aufweist. <sup>41</sup> Diese Aussage ist jedoch ausgesprochen allgemein gehalten – für ein differenziertes Vorgehen im Sounddesign wäre es wünschenswert, spezifischere Erkenntnisse über die Eignung verschiedener Spektren zu erlangen.

#### 4.1.1 Bandbreite des Signalspektrums

Grundlage für diesen Abschnitt ist die Überlegung, inwiefern es problematisch sein könnte, wenn ein Signal nicht Inhalte über die volle Bandbreite des Spektrums vorweisen kann.

#### 4.1.1.1 Vorne-Hinten-Unterscheidung

Für die Unterscheidung von vorne und hinten betrachtet Blauert als notwendig, dass das Signal in den Höhen breitbandig bis zu einer unteren Grenzfrequenz zwischen 2 und 8 kHz sei. 42 Die Ergebnisse seiner Erhebung mit Hoch- und Tiefpassrauschen (siehe Tabelle 1) bestätigen diesen Schluss: Erst wenn spektrale Anteile oberhalb von etwa 2 kHz vorhanden sind, konnten die Versuchspersonen (VP) beim Tiefpassrauschen vorne und hinten gut unterscheiden. Tiefpassrauschens gut. Kommen dagegen im Signal nur noch spektrale Anteile unterhalb von etwa 2 kHz vor, funktioniert die Vorne-Hinten-Unterscheidung zwar noch passabel, aber mit einem erheblichen Anteil an Vertauschungen. Somit ließe sich bei Klängen, die wenig bis keine Anteile über 2 kHz haben, eine Verbesserung erzielen, indem man solche Anteile hinzufügt.

Immerhin funktioniert diese Unterscheidung anscheinend auch über die spektralen Anteile unter 2 kHz – eine unverzichtbare Rolle spielen sie jedoch nicht: Sobald nämlich Anteile über 2 kHz vorhanden sind, ist die Unterscheidung fast so gut wie bei Rosa Rauschen mit voller Bandbreite; somit könnte man die Anteile unterhalb von 2 kHz ohne erhebliche Einbuße (zumindest den Aspekt Vorne-Hinten-Unterscheidung betreffend) entfernen.

Ein weiterer Schluss aus den Daten: Beim Hochpassrauschen werden die Werte erst dann schlechter, wenn die Grenzfrequenz so hoch ist, dass nur noch spektrale Anteile oberhalb von 8 bis 15,5 kHz übrig bleiben – Signale, die ausschließlich aus so hochfrequentem Material bestehen, sind im Sounddesign ausgesprochen selten, sodass dieser Makel in dieser Arbeit nicht weiter berücksichtigt werden soll.

-

<sup>&</sup>lt;sup>41</sup> vgl. Blauert 1969, S. 106

<sup>&</sup>lt;sup>42</sup> vgl. ebd., S. 57

Signalart	Grenzfrequenz	Korrekte Lok	alisation (%)
	(Hz)	SER vorne	SER hinten
Rosa Rauschen	Ohne	100	100
	2160 - 3000	98	99
	3000 - 4200	97	90
Hachnass Dauschan	4200 - 5800	94	83
Hochpass-Rauschen	5800 - 8100	97	96
	8100 - 11200	90	93
	11200 - 15500	84	76
	420 - 580	72	65
	580 - 810	71	65
	810 - 1120	77	87
	1120 - 1550	55	98
Tiefpass-Rauschen	1550 - 2160	50	80
	2160 - 3000	91	83
	3000 - 4200	97	70
	4200 - 5800	85	71
	5800 - 8100	74	89

Tabelle 1: Vorne-Hinten-Lokalisation bei gefiltertem Rauschen. 43
Die VP hatten als Antwort nur zwei Möglichkeiten, "vorne" und "hinten";
ein Wert von 50% deutet also auf ein zufälliges Ergebnis.

#### 4.1.1.2 Oben-Unten-Unterscheidung

Die Lokalisationsunschärfe des Erhebungswinkels liegt allgemein bei 4° (bei weißem Rauschen) bis etwa 17° (bei der Stimme eines unbekannten Sprechers) – dies gilt in der Vorwärtsrichtung der Medianebene. 44 Mit zunehmender Erhebung (und damit Orientierung weg von der Vorwärts- hin zur Oben-Richtung) steigt sie jedoch auf etwa das Dreifache. 45 Bei SE in Oben-Richtungen sind also erhebliche Fehllokalisationen zu erwarten; in diesem Abschnitt soll es darum gehen, inwiefern dies von der Bandbreite abhängig ist.

Roffler & Butler fanden heraus, dass Klänge Frequenzen über 7kHz enthalten müssen, damit das Gehör den Erhebungswinkel in der Medianebene bestimmen kann. Sie haben in Versuchen mit gefiltertem Rauschen festgestellt, dass, sobald Signalanteile über 7kHz vorhanden sind, die Unschärfe der Lokalisation erheblich sinkt. Folglich bereiten Klänge ohne Spektralanteile über 7 kHz Schwierigkeiten bei der Lokalisation in der Vertikalen.

Diese Ergebnisse von Roffler & Butler traten im Übrigen auch dann ein, wenn das Rauschen gar keine Anteile unterhalb von etwa 8 kHz aufwies - diese sind somit vernachlässigbar. 46

Dies deckt sich in etwa mit Blauerts Beobachtungen, der in seinen Richtungsbestimmenden Bändern das einzige Oben-Band bei etwa 8 kHz ausmachte.<sup>47</sup>

In eine andere Richtung deuten jedoch Carlile & Pralongs Untersuchungen von HRTFs: Sie zeigen, dass Pinna, Kopf und Torso Spitzen und Kerben im Frequenzgang erzeugen, die sich, abhängig von der Elevation, im Frequenzbereich zwischen etwa 3 und 8 kHz, und bei manchen VP bis 12 kHz,

<sup>45</sup> vgl. Dickreiter et al. 2014, S. 132

<sup>&</sup>lt;sup>43</sup> in Anlehnung an ebd., S. 56–57

<sup>&</sup>lt;sup>44</sup> vgl. Blauert 1974, S. 35

<sup>&</sup>lt;sup>46</sup> vgl. Roffler & Butler 1968, S. 1257–1258

<sup>&</sup>lt;sup>47</sup> vgl. Blauert & Braasch 2008, S. 94

bewegen.<sup>48</sup> Jedoch bedeutet das nicht notwendigerweise, dass diese vom Gehör für die Lokalisation in der Vertikalen auch verwendet werden; hierzu ist weitere Forschung nötig.

Klänge sind somit in der Vertikalen gut lokalisierbar, selbst wenn sie keine Signalanteile unter 8 kHz haben.

#### 4.1.2 Koinzidenzen mit Richtungsbestimmenden Bändern

Blauert fordert, dass das Spektrum von Signalen frei von stärkeren Einbrüchen oder Spitzen sein müsse, die mit den richtungsbestimmenden Bändern zusammenfallen könnten. <sup>49</sup> Solche Verzerrungen stellen widersprüchlich betonte Bänder da, die zu Diffuslokalisation führen, oder zum Zerfall in zwei Anteile, die in unterschiedlichen Richtungen lokalisiert werden. <sup>50</sup>

In seinem Versuch hatte er die VP (wohlgemerkt mit fixiertem Kopf) mit Stimuli beschallt, die aus Rosa Rauschen bestanden. Dieses hatte er jedoch so vorverzerrt, dass die mit den richtungsbestimmenden Bändern korrespondierenden Terzbänder um 30 bis 40 dB angehoben waren; für einen Stimulus in den v-Bändern, für den anderen in den h-Bändern. Unabhängig davon, ob das SE vorne oder hinten lag, lokalisierten sämtliche VP die Stimulus mit betonten v-Bändern vorne, und den Stimulus mit betonten h-Bändern hinten. Diese Anhebungen hatten also die Fähigkeit des Gehörs, richtig zu lokalisieren, vollständig ausgehebelt.

Da solche Spitzen (und Kerben) im Spektrum jedoch auch das Resultat von Klangbearbeitungen im Sounddesign sein können, sind dort ebensolche Koinzidenzen mit den richtungsbestimmenden Bändern zu erwarten. Hier könnte sich somit lohnen, das Sounddesign daraufhin zu analysieren und eventuell so zu verändern, dass diese Merkmale nicht mehr oder zumindest weniger ausgeprägt vorkommen.

Einen Hinweis darauf, wie breit und tief solche Spitzen und Kerben sein müssten, um einen negativen Effekt zu haben, erbringt die Untersuchung von Macpherson & Middlebrooks: Darin wurden Stimuli verwendet, deren Spektren Wellen aufwiesen. Es zeigte sich, dass solche Wellen im Frequenzgang die Zahl der Lokalisationsfehler erhöhten, wenn sie bei etwa eine Welle pro Oktave und einem Hub von mehr als 20dB lagen. Mehr als zwei oder weniger als eine halbe Welle pro Oktave hatten keine Auswirkungen mehr.

Auch Asano et al. meinen, dass makroskopische Eigenschaften, also die Energie in einem Oktavband relativ zu anderen, entscheidende Cues darstellen. Lediglich in Frequenzen unter 2 kHz fänden sich Cues für die Vorne-Hinten-Unterscheidung, die eher mikroskopisch, also über Spitzen und Kerben, arbeiten.<sup>53</sup>

#### 4.1.3 Reine Töne und schmalbandige Signale

Dass reine Töne sowie Signale mit einer Bandbreite von weniger als ein bis zwei Terzen zu extremen Fehllokalisationen in der Medianebene führen, belegt Blauert. <sup>54</sup> Solche Signale sind zwar im Sounddesign eher selten, gerade in Spielen mit Science-Fiction-Thema können sie aber z.B. in Form von Warnklängen oder Klängen von Interfaces und digitalen Maschinen durchaus auch vorkommen. Gerade bei ihnen ist also zu prüfen, ob sie sich breitbandiger und mit mehr Rauschanteil gestalten lassen.

<sup>51</sup> vgl. ebd., S. 92–93

<sup>&</sup>lt;sup>48</sup> vgl. Carlile & Pralong 1994, S. 3453

<sup>&</sup>lt;sup>49</sup> vgl. Blauert 1969, S. 106

<sup>&</sup>lt;sup>50</sup> vgl. ebd., S. 95

<sup>&</sup>lt;sup>52</sup> vgl. Macpherson & Middlebrooks 2003, S. 430

<sup>&</sup>lt;sup>53</sup> vgl. Asano et al. 1990, S. 167

<sup>&</sup>lt;sup>54</sup> vgl. Blauert 1974, S. 36

#### 4.2 Effekte der Hüllkurve

Wegen der unterschiedlichen Auswertungsmechanismen interauraler Cues<sup>55</sup> verwendet die vorliegende Arbeit zur einfachen Unterscheidung die Begriffe *Interaural Time Difference in Low Frequencies* (ITD-LF) für die Spektralanteile unterhalb, und *Interaural Time Difference in High Frequencies* (ITD-HF) für die oberhalb von etwa 1,6 kHz.

#### 4.2.1 Steilheit der Transienten

Im Frequenzbereich unterhalb von 1,5 kHz spielt die Steilheit der Transienten eine Rolle, wenn man davon ausgeht, dass eine Hüllkurve prinzipiell eine Amplitudenmodulation darstellt:

Dietz et al. stellen fest, dass für die Lokalisation in der Horizontalebene in Hörumgebungen, die viel Hall und/oder viele Störgeräusche aufweisen, der frühe, ansteigende Teil einer Amplitudenmodulation wichtig ist. <sup>56</sup> Dann liefern scharfe Steigungen im Frequenzbereich der ITD-LF die zuverlässigsten interauralen Cues. Das wird damit erklärt, dass diese frühen Signalanteile mit größter Wahrscheinlichkeit direkt von der Schallquelle kommen, während spätere Anteile von Raumreflexionen überdeckt werden können und so die Richtungsinformation widersprüchlich wird. <sup>57</sup>

Aber auch in reflexionsarmen Hörumgebungen konnte eine Relevanz der Hüllkurve nachgewiesen werden, hierbei im Frequenzbereich der ITD-HF: Klein-Henning et al. haben die einzelnen Hüllkurvensegmente (Attack, Hold, Decay) eines Stimulus mit 4 kHz-Centerfrequenz auf ihren Einfluss auf die Lateralisation hin untersucht. Während die Hold-Phase (Segment zwischen Attack und Decay, mit anhaltender Amplitude) und die Decay-Phase (Abschwellen des Klangs) wenig bis keinen Unterschied machten, führten kürzere Attackzeiten zur Verbesserung der Lateralisation. <sup>58</sup>

Der Effekt war bei den Attackzeiten bis zu 2,5ms zu beobachten; bei noch kürzeren Zeiten stellte sich keine Verbesserung mehr ein.

Allerdings betrug die längste Attackzeit, auf die getestet wurde, lediglich 10 ms. Für die Fragestellung dieser Arbeit wäre es interessant gewesen, auch für längere Attackzeiten Ergebnisse vorliegen zu haben. Ohne diese lässt sich nicht sagen, ob sich die Korrelation von Attackzeit und Lateralisationsgenauigkeit auch auf längere Zeiten, wie sie im Sounddesign häufig vorkommen, erweitern lässt; ob also die Genauigkeit mit längeren Zeiten als 10 ms noch schlechter wird. Hierzu wäre zukünftige Forschung wünschenswert.

Zusammenfassend kann gesagt werden, dass die Eigenschaften der Transienten die Lokalisation in komplexen Hörumgebungen im Frequenzbereich der ITD-LF beeinflussen, und in reflexionsarmen Umgebungen im Frequenzbereich der ITD-HF.

#### 4.2.2 Abstand der Transienten

Santala et al. haben die Lokalisation von Signalen zwischen 2 und 8,2 kHz in der vorderen Horizontalebene in einem reflexionsarmen Raum untersucht und einen Zusammenhang mit den Abständen zwischen Transienten festgestellt.<sup>59</sup> Die Werte lagen dabei zwischen 10 und 166 ms, wobei sich die Lokalisation verbesserte, je länger der Abstand war.

In gewisser Hinsicht wiederspricht dem allerdings die Studie von Klein-Henning et al., die den Zusammenhang der Lateralisation von Stimuli mit 4 kHz-Mittenfrequenz in reflexionsarmen Räumen

<sup>55</sup> vgl. Abschnitt 2.1

<sup>&</sup>lt;sup>56</sup> vgl. Dietz et al. 2013, S. 15151

<sup>&</sup>lt;sup>57</sup> vgl. ebd., S. 15154

<sup>&</sup>lt;sup>58</sup> vgl. Klein-Hennig et al. 2011, S. 3870

<sup>&</sup>lt;sup>59</sup> vgl. Santala et al. 2013, S. 7–8

mit der Länge der Pausen zwischen den Stimuli untersuchte. <sup>60</sup> Sie kam zu dem Ergebnis, dass sich oberhalb von 8,75 ms keine Verbesserung mehr einstellt. <sup>61</sup> Da es jedoch unklar ist, ob Pausen zwischen Signalen mit Abständen zwischen Transienten gleichzusetzen sind, und weil die untersuchten Pausen mit Werten zwischen 0 und 17 ms kurz gewählt waren, bewertet diese Arbeit den Wiederspruch als wohl irrelevant.

Es ist also anzunehmen, dass längere Abstände zwischen Transienten bei der Lokalisation in der Horizontalebene helfen.

#### 4.2.3 Ausgeprägtheit der Transienten

Auch der sog. Franssen-Effekt weist auf die Wichtigkeit von Hüllkurven für die Lokalisation in reflexionsreichen Umgebungen hin. Ihm zufolge sind dort Klänge nur bei Einsetzen des Klanges (dem sog. *Onset*) zu lokalisieren, im weiteren Signalverlauf jedoch nicht mehr, wenn dieser unmoduliert ist.<sup>62</sup> Im Umkehrschluss eignet sich also ein mit ausgeprägten Transienten versehender Signalverlauf besser. Daneben konnten Santala et al. in ihrer bereits erwähnten Arbeit zeigen, dass die Lokalisation desto besser funktioniert, je prominenter die Transienten aus dem Gesamtsignal herausragen. <sup>63</sup>

## 4.3 Signaldauer

Auch die Signaldauer nimmt in zweierlei Hinsicht Einfluss:

#### 4.3.1 Informationsmenge

Es liegt nahe, dass mit einer längeren Dauer eines Signals dem Gehör auch mehr Richtungsinformation übermittelt wird. Für die Richtungslokalisation in der Horizontalebene konnte dieser Einfluss nachgewiesen werden: Sowohl Lokalisationsfehler als auch Vorne-Hinten-Vertauschungen nehmen mit zunehmender Signallänge erheblich ab, von 22° Fehllokalisation und 13% Vorne-Hinten-Vertauschungen bei Stimuli von 64 ms Dauer auf 11° und 5% bei Stimuli von 2000 ms Dauer.

Betrachtet man jedoch unmodulierte Signale in reflexionsreichen und/oder komplexen Umgebungen, verhält es sich anders: Hier spielt der Onset die dominante Rolle, im Gegensatz zum unmodulierten Teil. Es lässt sich folgern: Wenn sich aus dem unmodulierten Teil kaum nützliche Information beziehen lässt, ändert sich dies wahrscheinlich auch nicht wesentlich mit der Signallänge.

#### 4.3.2 Latenzen

Die Signaldauer ist auch relevant für die Wirksamkeit des Headtracking: So konstatiert Blauert, dass der Mensch im Mittel 350 ms braucht, um auf ein Signal mit einer Kopfbewegung zu reagieren.<sup>64</sup>

Dieser Wert gilt für das natürliche Hören – in VR ist jedoch auch noch die Systemlatenz hinzu zu rechnen, bestehend aus den Latenzen des Headtrackings und der Binauralsynthese: Die Headtracker moderner VR-Systeme wie der HTC Vive erzeugen lediglich Latenzen von unter 5 ms. <sup>65</sup> Zur Binauralsynthese ist die Latenz bei den meisten kommerziellen Lösungen unbekannt; veröffentlicht haben ihre Werte hierzu nur Impulsonic, die sie bei ihrer VR-Audio-Lösung

<sup>&</sup>lt;sup>60</sup> vgl. Klein-Hennig et al. 2011, S. 3870

<sup>61</sup> vgl. ebd., S. 3858

<sup>&</sup>lt;sup>62</sup> Franssen, Nico V. 1960, zitiert nach Dietz et al. 2013, S. 15151

<sup>&</sup>lt;sup>63</sup> vgl. Santala et al. 2013, S. 8

<sup>&</sup>lt;sup>64</sup> Wilmowsky 1960, zitiert nach Blauert 1974, S. 79

<sup>65</sup> vgl. Kreylos 2016

RealSpace3D mit 21ms angeben. 66 Es ist jedoch davon auszugehen, dass die meisten Lösungen unter einem Wert von ungefähr 50 ms bleiben, um marktfähig zu sein, da Latenzen darüber für den Hörer auffällig sind. 67 68

Somit ist damit zu rechnen, dass die Vorteile von Headtracking nur bei Signalen, die länger als etwa 400 ms sind, zur Geltung kommen. Es liegt jedoch nahe, dass diese Grenze nur für Signale gilt, die einmalig oder in sehr großem Zeitabstand zu hören sind; sobald ein so kurzes Signal in schneller Abfolge mehrmals zu hören ist, gibt dies wiederholte Gelegenheit zur Lokalisation, und ist insofern wohl wie ein längeres Signal zu betrachten.

#### 4.4 Vorkenntnis & Lerneffekte

Erfahrung besteht meist bereits bei Signalen, die dem Hörer im Alltag oft begegnen – z.B. die Stimmen bekannter Menschen, Hundebellen, Schritte, etc. – aber sie lässt sich auch gewinnen, indem ein unbekanntes Signal mehrmals vorgespielt wird.<sup>69</sup>

#### 4.4.1 Entfernungshören

Wie bereits beschrieben schätzt das Gehör Entfernungen mittels mehrerer Mechanismen ein. Zumindest der Mechanismus, der über Lautstärke- und spektrale Veränderungen funktioniert, ist abhängig von der Vertrautheit mit der Quelle, weil unser Gehör dabei die Lautstärke und das Spektrum eines HE mit seiner Erfahrung vergleicht. <sup>70</sup> Fehlt ihm diese jedoch, weil er das Signal noch nie gehört hat, ist der Vergleich nicht möglich, und somit auch nicht feststellbar, ob die Lautstärke und das Spektrum des HE bereits im Signal des SE vorhanden sind oder durch die Entfernung zustande kommen. In anderen Worten: Ist ein SE unbekannt und sein HE z.B. leise und dumpf, lässt sich nicht unterscheiden, ob dies eine Eigenschaft des SE ist, oder ob das HE aufgrund der Entfernung leise und dumpf ist.

Dies ist nicht zuletzt für das Problem der IKL relevant, weil unklare Entfernungsinformationen dazu beitragen.<sup>71</sup>

#### 4.4.2 In der Medianebene

Auch in der Medianebene ist es naheliegend, dass das Spektrum des HE verglichen wird mit bekannten Mustern, um festzustellen, ob spektrale Eigenschaften vom Signal oder von der Filterung durch die HRTF des Hörenden herrühren. Blauerts Erkenntnisse bestätigen dies für die Lokalisation mit fixiertem Kopf.<sup>72</sup>

Dies relativiert sich jedoch, wenn man bedenkt, dass Hören mit fixiertem Kopf eine äußerst ungewöhnliche Situation ist. Und sobald sich der Kopf und/oder die Schallquelle bewegen, so argumentiert Han, variieren diese Klangeigenschaften des HE; dadurch kann unterschieden werden, welche spektralen Eigenschaften vom Signal und welche von der HRTF herrühren. So gesehen wird auch hier mit dem Erinnerungsvermögen gearbeitet, jedoch wird quasi nicht mit langfristigen Erinnerungen verglichen, sondern mit dem Klangeindruck vor der Kopfdrehung.

-

<sup>66</sup> vgl. Anish Chandak 2014, S. 37

<sup>&</sup>lt;sup>67</sup> vgl. Brungart et al. 2006, S. 42–43

<sup>&</sup>lt;sup>68</sup> vgl. Stitt et al. 2016, S. 4

<sup>&</sup>lt;sup>69</sup> vgl. Blauert 1969, S. 62

<sup>&</sup>lt;sup>70</sup> vgl. Coleman 1962, S. 345–346

<sup>&</sup>lt;sup>71</sup> vgl. Dickreiter et al. 2014, S. 133-134

<sup>&</sup>lt;sup>72</sup> vgl. Blauert 1974, S. 35

Han schlussfolgert, dass bei freien Kopfbewegungen vorherige Kenntnis des ursprünglichen Spektrums des Signals nicht mehr notwendig sei. 73 Er selbst relativiert jedoch, dass Signalkenntnis wiederum wichtig werde, sobald Menschen ihren Kopf (freiwillig) nicht bewegen, oder Cues widersprüchlich sind. 74

Zusammenfassend lässt sich sagen, dass Vertrautheit mit dem Signal eine Rolle spielt; wie sehr, hängt jedoch von den Umständen ab.

Dieser Aspekt wäre interessant für weitere Forschung; da dies jedoch den Rahmen dieser Arbeit sprengen würde, soll sie jedoch für das Sounddesign in dieser Arbeit keine weitere Rolle spielen.

## 4.5 Zusammenfassung

Es wurden spezifische Aspekte festgestellt, die im kommenden Abschnitt bei der Klanggestaltung berücksichtigt werden sollen.

So wurde festgestellt, das bestimme Frequenzbereiche für die verschiedenen Cues relevanter sind als andere, und sich manche ignorieren lassen.

Auch Amplituden- und Zeitmerkmale der Hüllkurve oder die Länge des Signals sind Faktoren, die bei manchen Signalen und Hörsituationen berücksichtigt werden können.

Um mögliche Probleme zu identifizieren, zeigt sich auch der Ansatz als dienlich, das Signalspektrum auf Spitzen und Kerben hin zu untersuchen, die mit Richtungsbestimmenden Bändern zusammenfallen.

-

<sup>&</sup>lt;sup>73</sup> vgl. Han 1992, S. 3

<sup>&</sup>lt;sup>74</sup> vgl. ebd., S. 14

# 5 Virtual Reality Games

Im Folgenden soll dargestellt werden, was Virtual Reality Games sind und welche Rolle die Richtungslokalisation in ihnen spielen kann.

#### 5.1 Technik der VR

VR-Systeme, wie sie aktuell für den Konsumentenmarkt erhältlich sind, bestehen üblicherweise aus einem sogenannten Head Mounted Display (HMD), einem Eingabegerät und einem Trackingsystem. Exemplarisch wird hier die Vive (Consumer Edition) von HTC, die im April 2016 auf den Markt kam, vorgestellt.

#### 5.1.1 Head Mounted Display

Ein HMD, auch VR-Brille genannt, wird auf dem Kopf getragen, und stellt nah vor beiden Augen ein stereoskopisches Bildschirmsystem bereit. So ersetzt sie als visuelles Ausgabegerät den bei Spiele-PCs üblichen Monitor.

Sie lässt von außen kein Licht an die Augen, sodass der Benutzer von der "realen" Welt um ihn herum visuell komplett abgeschottet wird.

Die Vive hat pro Auge jeweils einen Bildschirm mit 1080 x 1200 Pixeln. <sup>75</sup> Genaue Angaben über den Blickwinkel macht der Hersteller nicht. Kreylos <sup>76</sup> hat ihn gemessen und kommt auf 113° in der Vertikalen und 110° in der Horizontalen. <sup>77</sup>

#### 5.1.2 Trackingsystem

Die Aufgabe des Trackingsystems ist es, die Bestimmung der Position und Orientierung des HMDs und der Controller zu ermöglichen.

Dazu sendet das *Lighthouse* getaufte Trackingsystem der Vive über zwei sogenannte Basisstationen, die an gegenüberliegenden Punkten am Rande des Trackingbereichs aufgestellt werden, Infrarotblitze aus, die von Photosensoren am Headset und an den Controllern registriert werden. Über ihre Laufzeiten kann die Position und Rotation relativ zu den Basisstationen berechnet werden.<sup>78</sup>

Lighthouse kann dies über ein bis zu 25m² großes "Spielfeld". So wird nicht nur die Rotation des HMDs festgestellt, um die virtuelle Kamera der VR-Anwendung in die richtige Richtung auszurichten; es ermöglicht dem Benutzer auch, sich im Rahmen des Spielfeldes frei zu bewegen. In anderen Worten: Er bewegt sich auf eigenen Füßen fort, statt sich z.B. (wie es in früheren VR-Systemen der Fall war) schwebend fortzubewegen, während er in der Realität auf einem Stuhl sitzt und die Richtung per Joystick bestimmt.

<sup>76</sup> vgl. Oliver Kreylos 2016

<sup>&</sup>lt;sup>75</sup> vgl. Wikipedia 2017

<sup>&</sup>lt;sup>77</sup> Kreylos merkt jedoch an, dass sich der horizontale Wert wegen der variierenden Überlappung der Sichtfelder der beiden Augen und anderen Faktoren wie individuellem Augenabstand nicht sehr genau messen ließe.

<sup>78</sup> vgl. Sean Buckley 2015

Für Distanzen, die über den Rand des Trackingbereichs hinausgehen, muss eine alternative Fortbewegungsmechanik implementiert werden, wie z.B. das *Teleporting*, bei dem der Benutzer auf einen weiter entfernten Bereich zeigt und schlagartig dorthin versetzt wird.

#### 5.1.3 Handcontroller

Als Eingabegerät setzt HTC zwei Controller ein, jeweils einen für jede Hand. Auch ihre Position und Rotation wird über das Trackingsystem erfasst. In VR sind sie auch visuell repräsentiert, durch virtuelle Objekte gleichen Aussehens.

Zwar bieten die Handcontroller mehrere Knöpfe und ein Trackpad; ihre Hauptfunktion ist aber, durch ihr bloßes Bewegtwerden Eingaben zu machen – z.B. durch Zeigen, Berühren virtueller Objekte, oder auch durch Gesten wie Werfen.

#### 5.1.4 Audiosystem

Die Vive besitzt, im Gegensatz zu einigen Konkurrenzmodellen, keine eigenen Kopfhörer, und stellt bisher Softwareentwicklern auch keine eigene binaurale Audio-Lösung. Ein Kopfhöreranschluss ist jedoch am HMD angebracht; der Nutzer soll eine eigene Wahl an Lautsprechern einsetzen. Entwickler haben ohnehin die Möglichkeit, ein eigenes oder eingekauftes Binauralsynthese-Plug-In für das Audio-System der Game Engine einzusetzen.

## 5.2 Virtual Environments (VE)

Die "Substanz" eines VE besteht für Sherman & Craig aus Geographie, Benutzeroberflächen, Objekten und Agenten.  $^{79}$ 

#### 5.2.1 Geographie

Die Geographie eines VE beinhaltet sowohl das Terrain als auch verschiedene Regionen der Virtuellen Welt und wie sie miteinander verbunden sind. Akustisch sind hier vor allem Atmos relevant; aber auch entscheidet sich hier, welche Raumakustik vorhanden (oder nicht) ist, je nachdem, ob man sich in einem Gebäude (und welchem) oder draußen befindet.

#### 5.2.2 Benutzeroberfläche

In Nicht-VR-Spielen werden Benutzeroberflächen oft als Overlay außerhalb der Spielwelt eingeblendet; in VR werden sie dagegen häufig von Objekten im VE repräsentiert. Somit ergibt es auch Sinn, ihre Klänge mit Richtungsinformation wiederzugeben. Denkbar ist, dass solch ein Interface eine eher zweitrangige Rolle spielt (z.B. Einstellungen), oder aber zentrales Spielelement darstellt. So oder so sind hier alle möglichen Klänge denkbar, wobei jedoch der Fokus auf effektivem Feedback auf Spielerinteraktion liegen dürfte.

#### 5.2.3 Objekte

Objekte sind alle Dinge im VE, vom Gänseblümchen bis zu Wolkenkratzern. Dinge, die sich beobachten und eventuell auch manipulieren lassen. Objekte können aktiv oder passiv sein, relevant für den Spieler, oder nicht. Ihre Vertonung muss sich selbstverständlich an ihrer Natur orientieren; sind die Objekte jedoch keine Agenten, dienen sie wahrscheinlich eher der Anreicherung des VE als zentralen Handlungen; und spielen so eine ebensolche Rolle.

\_

<sup>&</sup>lt;sup>79</sup> vgl. Sherman & Craig 2003, S. 406–410

#### 5.2.4 Agenten

Unter Agenten versteht man autonom handelnde Elemente. Es können Charaktere sein, aber auch Objekte, wenn sie Verhalten aufweisen, das sie als eigenständige Intelligenz erscheinen lassen, und somit lebendig. Das Konzept ähnelt den Non Player Characters (NPC) die man in Spielen findet. Agenten dürften oft die größte Aufmerksamkeit des Spielers erfahren und Objekt seiner Interaktion sein, wodurch auch ihre Vertonung eine wichtige Rolle einnimmt.

#### 5.3 Erlebnis der VR

Das Medium Virtual Reality besteht aus interaktiven Computersimulationen, die die Position und Handlungen des Nutzers registrieren und das Feedback an seine Sinne erweitern oder ersetzen; dabei erzeugen sie das Gefühl, in der Simulation präsent zu sein.<sup>80</sup>

#### 5.3.1 Presence

Eben dieses Gefühl wird als *Presence* bezeichnet; also das Erlebnis, mental an einem Ort zu sein, obwohl man physisch an einem anderen ist. <sup>81</sup> Sie stellt ein zentrales Element für die Effektivität des VR-Erlebnisses dar. Ist sie vorhanden, tritt ein Effekt namens *Willing Suspension of Disbelief* ein: Die Bereitschaft des Nutzers, das Wissen, dass es sich beim VE um eine Illusion handelt, vorübergehend auszublenden; sich also mental auf sie einzulassen. <sup>82</sup>

Nach Witmer & Singer sind für Presence zwei Teilaspekte notwendig: Immersion und Involvement.

Immersion<sup>83</sup> ist die Wahrnehmung, von einem Virtual Environment (VE) umgeben, Teil von ihm zu sein und mit ihm zu interagieren.<sup>84</sup>

Involvement meint einen mentalen Zustand, den der Nutzer erlebt, wenn er seine Aufmerksamkeit und Energie auf Wahrnehmungen, Handlungen und Ereignisse konzentriert.<sup>85</sup> Je relevanter diese für ihn sind, desto involvierter ist er.

#### 5.3.2 Einflussfaktoren auf die Presence

Verstärkt sich Immersion und/oder Involvement, entsteht auch intensivere Presence. Witmer & Singer nennen 11 Faktoren in 4 Gruppen, die sich auf Involvement und/oder Immersion auswirken, und damit auf die Presence.<sup>86</sup>

Darunter gibt es die beiden Gruppen der Ablenkungs- und der Kontrollfaktoren; aus einigen dieser Faktoren lässt sich auf den Einfluss von nicht-gestalterischen Aspekten wie Headtracking, Latenz, sowie akustischer Isolation von der realen Welt mittels (geschlossener) Kopfhörer auf Presence schließen.

Die anderen beiden Gruppen sind mit Hinblick auf die Richtungslokalisation relevant, sie sollen deshalb hier detaillierter besprochen werden.

In der Gruppe der Realismus-Faktoren finden sich:

81 vgl. Witmer & Singer 1998, S. 225

<sup>&</sup>lt;sup>80</sup> vgl. ebd., S. 13

<sup>82</sup> vgl. Dörner et al. 2013, S. 8

<sup>&</sup>lt;sup>83</sup> Immersion, wie sie bei Witmer & Singer definiert ist, ist nicht zu verwechseln mit anderen, oft verwendeten Definitionen, z.B. mit der von Slater & Wilbur, die in ihr keinen subjektiven Zustand sondern vor allem eine technischen, objektiven Aspekt sehen.

<sup>84</sup> vgl. Witmer & Singer 1998, S. 227

<sup>85</sup> vgl. ebd., S. 227

<sup>86</sup> vgl. ebd., S. 228-230

Scene realism: Der Realismus eines VE trägt ebenfalls zur Presence bei. Ihr Inhalt muss dabei nicht realitätsnah sein, Fantasy-Elemente z.B. sind also durchaus erlaubt. Realismus bezieht sich hier auf die die Qualität der multimodalen Darstellung. Eine schlechte Lokalisation stört die Presence somit auch.

Consistency of information with the objective world: Je mehr sich das Erleben des VE deckt mit Erfahrungen aus der realen Welt, desto eher entsteht Presence. Nutzer erwarten, das Richtungshören so funktioniert, wie sie das außerhalb von VR gelernt haben.

Zur Gruppe der sensorischen Faktoren gehören:

Interface awareness: Wird dem Nutzer ins Bewusstsein gerufen, dass er sich in einer Simulation befindet, sei es z.B. durch Schweißbildung im HMD oder Artefakte in der Darstellung, stört dies das Gefühl von Presence. Auch die Fehllokalisation ist als solches Artefakt zu betrachten; tritt sie auf, wird sich der Spieler der Technik bewusst, die die Simulation ermöglicht.

Multimodal Presentation: Die Kohärenz, mit der die Sinne des Nutzers angesprochen werden, beeinflusst seine Immersion und sein Involvement und darüber seine Presence. Kohärenz bedeutet, dass sich die Information eines Kanals decken soll mit der eines anderen Kanals. Wenn der Nutzer also ein Objekt visuell lokalisiert, ein dazugehöriges HE aber woanders, leidet die Presence.

#### 5.4 Games

Games und VR scheinen in vielerlei Hinsicht hervorragend zusammen zu passen: Beide stellen den Nutzer und sein Handeln in den Mittelpunkt. Auch in Games sind Involvement und Immersion wichtige Aspekte. Und Games lassen sich betrachten als Interaktion zum Zweck der Unterhaltung; insofern lässt sich die Aussage, Interaktion sei in VR direkter und bedeutungsvoller, auf die Interaktion in VR-Spielen übertragen.

Klimmt beschreibt die Wirkung von Games über drei Aspekte:

#### 5.4.1 Selbstwirksamkeitserleben (SWE)

Im echten Leben haben viele Handlungen nicht die gewünschte Auswirkung; zu viele Faktoren, die man nicht in der Hand hat, spielen eine Rolle. Manchmal zeigen sich die Auswirkungen auch mit großer Verzögerung, sodass man fast nicht mehr sicher ist, ob wirklich das eigene Handeln den Ausschlag gab. Kurz: Menschen sind nicht so mächtig, wie sie gerne wären.

SWE beschreibt das positive Gefühl, das man bekommt, wenn die eigenen Handlungen direkte Auswirkungen zeigen. Es ist besonders groß, wenn die Auswirkungen unmittelbar eintreten.<sup>87</sup>

Computerspiele sind mit ihren unmittelbaren Aktions-Reaktions-Schleifen besonders gut geeignet, SWE hervorzurufen. Sie kann jedoch durch technische Mängel nicht nur reduziert werden, sondern sogar ins Negative umgekehrt werden. 88 "Hakt" die Steuerung etwa, ist statt SWE mit erheblichem Frust beim Spieler zu rechnen, da er nun nicht mehr die faire Chance hat, mit seinem Handeln direkte Wirkung zu erreichen. Fehlerhafte Richtungslokalisation eines HE wird in dieser Arbeit als eben solcher technischer Mangel angesehen.

-

<sup>87</sup> vgl. Klimmt 2010, S. 76-77

<sup>&</sup>lt;sup>88</sup> vgl. ebd., S. 79

#### 5.4.2 Spannung und Lösung

Mit Spannung beschreibt Klimmt den Wunsch, ein Ziel zu erreichen, und die Ungewissheit, ob dies gelingen wird; unter Lösung versteht er die positiven Gefühle bei einsetzendem Erfolg.<sup>89</sup>

Ein wichtiges dieser Gefühle ist das Kompetenzerleben. Somit ist naheliegend, dass Dynamik aus Spannung und Lösung gestört wird, wenn Lokalisation als Element der Spielmechanik einerseits Voraussetzung zum Erfolg im Spiel ist, andererseits unzuverlässig funktioniert und somit das Scheitern im Spiel nicht an der mangelnden Kompetenz des Spielers, sondern an der technischen Unzulänglichkeit des Spiels liegt.

#### 5.4.3 Simulierte Lebenserfahrungen

Spiele kann man auch als Möglichkeit betrachten, in eine Rolle zu schlüpfen, die man im echten Leben nicht innehat. Neben dem Vergnügen, das es bereitet, besondere Möglichkeiten und auch Wichtigkeit (wie z.B. die eines Feldherren) zu genießen, bieten sie auch lehrreiche Erfahrungen, die Menschen bei der Lebensbewältigung helfen.<sup>90</sup>

Klimmt nennt weiterhin als wichtige Faktoren, damit dieser Aspekt gelingt: Interaktivität, ein lebendiges VE mit autonomen Elementen (Agents), Qualität der Stimuli und damit einhergehende Realitätstreue (Fidelity), sowie Komplexität, Umfang und Tiefe des VE.

Die Qualität der Stimuli und die Realitätstreue dürften auch eine gute Lokalisation einschließen. Des Weiteren liegt auch in diesem Aspekt ein Fokus auf der Interaktion, wobei je nach Spielprinzip die wichtigste Interaktion mit Agents stattfinden dürfte; so wäre es z.B. denkbar, dass man Agents lokalisieren muss, um mit ihnen interagieren zu können.

# 5.5 Zusammenfassung

Grundlegende Aspekte der VR und von Games wurden vorgestellt. Sound und insbesondere eine gute Richtungslokalisation wirken sich auf viele Faktoren aus, die das Presence-Erlebnis in VR und das erfolgreiche Spielerlebnis ermöglichen. Letzteres gilt vor allem dann, wenn Richtungslokalisation Teil der Spielmechanik ist.

-

<sup>89</sup> vgl. ebd., S. 82-84

<sup>90</sup> vgl. ebd., S. 95-98

# 6 Szenario

#### 6.1 Kriterien

Von den bisherigen Erkenntnissen ausgehend lässt sich bestimmen, wann in Games Richtungslokalisation am relevantesten und wirksamsten ist. Dies dient im Folgenden als Grundlage für einen Anforderungskatalog an ein Szenario: Eine prototypische Spielsituation, an der sich der Sounddesign-Prozess dieser Arbeit orientieren kann. Dabei werden auch aktuelle VR-Spiele einbezogen, um typische Spielsituationen ausfindig zu machen.

#### 6.1.1 Gameplay

Interaktion ist ein wichtiger Teil der Wirkung von VR-Games, zu der Richtungslokalisation ebenfalls beiträgt. Es liegt daher nahe, ein Szenario auszuwählen, in dem Interaktion des Spielers zentral ist.

Klimmt teilt die Gameplay-Mechanismen in Computerspielen in zwei Kategorien: Schnelligkeits-Präzisions-Aufgaben und Komplexitäts-Bewältigungs-Aufgaben. <sup>91</sup> Da in Hinblick auf Richtungslokalisation sowohl Schnelligkeit (sichere Lokalisation ohne Zögern) als auch Präzision (Lokalisationsschärfe) geeignet scheinen, soll diese Art Gameplay bevorzugt werden.

Weiterhin soll die Interaktion des Spielers möglichst direkt erfolgen und ihre Auswirkungen unmittelbar eintreten.

Einen dynamischen Verlauf aus Spannung und Lösung sollte das Szenario ebenfalls aufweisen.

Auch soll das Handeln des Spielers besonders sein, also z.B. mehr Macht beinhalten als Menschen im realen Leben haben.

#### 6.1.2 Elemente des VE

Das Terrain des VE sollte *indoor* sein; bei einem Schauplatz in einem Raum wird der Einfluss der Raumakustik berücksichtigt.

Bei der Frage, mit wem der Spieler interagieren soll, stehen die VE-Elemente Benutzeroberflächen, Objekte und Agents zur Wahl. Während die Klänge einer Benutzeroberfläche auch ein interessantes Untersuchungsobjekt wären, scheint es geboten, Interaktion mit Agents den Vorzug zu geben, da die Klänge aus allen Richtungen um den Spieler kommen sollen. Dies kommt am ehesten bei den Agents vor, da sie sich autonom im VE bewegen.

#### 6.1.3 Sound

Die Sounds sollen aus möglichst vielen Richtungen kommen. Unabdingbar ist, dass die Richtungslokalisation von HE elementarer Bestandteil der Spielmechanik ist.

#### 6.1.4 Sonstige Eigenschaften

Für Immersion und Involvement ist relevant, dass sich mehrere Kanäle decken, also die Schallquelle sowohl zu hören ist, als auch ein weiterer Sinn von ihr angesprochen wird. Wenn auch andere Kanäle denkbar sind, kann man davon ausgehen, dass die meisten Spiele die Deckung des Tons mindestens mit dem Bild beinhalten, man also einen Agent sieht, der auch die Schallquelle darstellt.

<sup>&</sup>lt;sup>91</sup> vgl. ebd., S. 81

#### 6.2 Auswahl eines VR-Games

Es soll betrachtet werden, welche VR-Spiele es gibt, die den Kriterienkatalog erfüllen können.

#### 6.2.1 Übersicht VR-Anwendungen

Das Gameslab der HdM bietet ein eigenes VR-Labor mit einem Steam-Account und einer Auswahl an insgesamt 29 Titeln für SteamVR. Aus diesen Titeln wurden von vornherein alle aussortiert, die entweder nicht interaktiv sind oder bei denen Richtungslokalisation von HE nicht wesentlicher Bestandteil der Spielmechanik ist. Außerdem zwei Titel, deren VEs unter Wasser sind und somit eine ungewöhnliche Akustik aufweisen, sowie ein Musikspiel ohne Geräusche.

Die restlichen 11 sind in Tabelle 2 aufgeführt, zusammen mit einer Bewertung, inwiefern sie die Kriterien erfüllen. Zwei Kriterien befinden sich nicht in der Tabelle: Akustisch-visuelle Deckung kommt in sämtlichen Titeln der Liste vor; die klanglichen Makel werden an späterer Stelle besprochen.

#### 6.2.2 Systematische Auswahl

Bei der engeren Auswahl wurde darauf geachtet, möglichst viele oder alle der Kriterien zu bedienen. Nicht sehr bedeutend waren die Kriterien Geographie ("Indoor/Outdoor") und "HE-Richtungen". Denn im Falle des Geographie-Kriteriums "Outdoor" ist es absolut denkbar, das Spielprinzip ohne Einbuße "Indoor" einzusetzen, bzw. umgekehrt; Ausnahme ist hier *Elite Dangerous*, das im Weltall spielt, und deshalb nur in einem VE mit riesigen Ausmaßen und somit nicht "Indoor" funktioniert. Auch beim Kriterium "HE-Richtungen" ist es vorstellbar, die Zahl der Richtungen auf alle zu erweitern und das Spielprinzip beizubehalten.

Somit haben sich sechs Titel herauskristallisiert. Ihre Spielmechaniken lassen sich grob in zwei Kategorien gruppieren:

#### 6.2.2.1 Arena / Zielen & Ausweichen

Die Titel *Lightblade VR*, *Space Pirate Trainer*, *Vive Defender* und der Action-Teil von *Trials on Tattooine* haben, vereinfacht betrachtet, eine ähnliche Spielmechanik, bei der der Spieler in einem Science-Fiction-Scenario auf Gegner schießen und gleichzeitig den Schüssen der Gegnern ausweichen bzw. sie mit einem Schutzschild blockieren muss. Natürlich bestehen im Detail Unterschiede – im Fall von *Lightblade VR* schießt nur der Gegner, der Spieler kann mit einem Schwert zuschlagen, bei *Trials on Tattooine* lenkt der Spieler die Schüsse der Gegner auf sie selbst zurück – gemeinsam haben sie jedoch, dass der Spieler im Wesentlichen zielt und ausweicht, während die Gegner schießen und ihre Position wechseln.

Besonders herausfordernd ist dies bei *Lightblade VR* und *Space Pirate Trainer*, bei denen sich die Gegner in Form von fliegenden Drohnen besonders schnell bewegen, und auch überraschend z.B. hinter dem Spieler auftauchen. Durch die Ähnlichkeit der Interaktion ist also bei diesen Titeln besonders schnelle und gute Lokalisation entscheidend.

#### 6.2.2.2 Dungeon / Nahkampf

Bei den Titeln *Arcan's Call: Act 1* und *Vanishing Realms* handelt es sich um Rollenspiele, in denen sich der Spieler, vereinfacht dargestellt, durch ein Labyrinth fortbewegt, dabei Gegner in Form von Fantasy-Monstern entdeckt, sie mit Schwerthieben bekämpft und gleichzeitig deren Schwerthiebe blockieren muss.

Name	Genre	Aufgaben		Interaktion	ion		Auswirk.	Spannung	In/Out-	Hörereignis
			zentral	zentral besonders direkt mit	direkt	mit	sofort	gnu	door	Richtungen
Abode	Adventure / Puzzle	Komplex Bewältig.	Ja	Nein	Teilw.	Obj. & Interf.	Nein	Nein	ln	Horizontalebene
Arcan's Call: Act 1	RPG / Kampf	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	ī	Horizontalebene
Elena	Adventure / Puzzle	Komplex Bewältig.	Teilw.	Nein	Ja	Obj. & Interf.	Nein	Teilw.	In	Horizontal
Elite Dangerous (Action-Teil)	Weltall-Shooter	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Out	Alle
HoloBall	Geschickl.	Präzision- Schnell.	Ja	Nein	Ja	Obj.	Ja	Ja	'n	Vorne & Oben
Lightblade VR	Kampf / Geschickl.	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Out	Horizontalebene & oben
Trials on Tatooine (Action-Teil)	Kampf / Geschickl.	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Beides	Vorne
Trials on Tatooine (Rätsel-Teil)	Rätsel	Komplex Bewältig.	Ja	Nein	Ja	Interf.	Ja	Nein	Out	Vorne
Vanishing Realms	RPG / Kampf	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Beides	Alle
Vive Defender (Action-Teil)	Tower Defense / Shooter	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Out	Horizontalebene
Screampark VR	Room Escape / Rätsel	Komplex Bewältig.	Ja	Nein	Ja	Obj. & Interf.	Ja	Nein	Īn	Alle
Space Pirate Trainer	Shooter / Geschickl.	Präzision- Schnell.	Ja	Ja	Ja	Agents	Ja	Ja	Out	Horizontalebene; Erhebung: -20° bis 80°

Tabelle 2: Vorauswahl VR-Games im Gameslab<sup>92</sup>

<sup>&</sup>lt;sup>92</sup> Eigene Tabelle

#### 6.2.3 Szenario

Bei *Lightblade VR* und *Space Pirate Trainer* bleibt der Spieler immer am gleichen Ort stehen, während immer neue Gegner aus verschiedenen Richtungen auf ihn zu kommen.

Im Gegensatz dazu kommt in *Arcan's Call* und *Vanishing Realms* der Spieler zum Gegner: Er entdeckt sie während seiner Reise meist schon von weitem, bevor sie ein Geräusch von sich geben; und sie befinden sich meist tendenziell vor ihm. Wird der Spieler von mehreren Gegnern angegriffen, passiert es auch, dass sich ein Gegner hinter ihm befindet, und es vorteilhaft wäre, ihn gut lokalisieren, sozusagen akustisch "im Auge behalten" zu können. Diese Art von Lokalisation erscheint jedoch nicht so essentiell wie die beim Orten der Drohnen bei den anderen beiden Spielen.

Lightblade VR und Space Pirate Trainer sind sich sehr ähnlich; bei ersterem bewegen sich die Gegner deutlich schneller und sie kommen tatsächlich aus allen Richtungen.

Deshalb fällt die Wahl als Vorbild für das Szenario auf Lightblade VR.

Das bedeutet nicht, dass das Szenario identisch damit ist. Vielmehr soll es ihm und den ähnlichen Titeln im Wesentlichen ähneln, aber ideal für die weitere Untersuchung sein; somit sei es wie folgt definiert:

Das Szenario stellt die gleiche Situation dar: Eine Arena, in dessen Mitte der Spieler steht und auf den Agent in Form einer Drohne reagieren muss. Die Arena wird, anders als bei *Lightblade VR*, nach drinnen verlagert.

In der Spielmechanik interessiert für das Szenario nur der Teil, in dem der Spieler die Drohne lokalisiert und sich zu ihr hinwendet; die weiteren Bestandteile wie Ausweichen und Schießen sind denkbar, für das weitere Vorgehen dieser Arbeit aber unwichtig, da sie sich auf den Teil beschränkt, in dem die Lokalisation stattfindet.

Die HER werden von ihrer Einschränkung, wie sie bei Space Pirate Trainer vorkommt, so weit wie möglich befreit.

In *Lightblade VR* erzeugen die Drohnen zwei Klänge: Kurze Laserpistolenschüsse; und einen kontinuierlichen, unmodulierten Klang mit metallischem/gläsernem Charakter, der im Sounddesign tatsächlich oft als "Drohne" umschrieben wird. Letzter ist für den folgenden Abschnitt am besten geeignet, da er die meisten Makel besitzt und sich deshalb eine Vielzahl von Bearbeitungen anwenden lassen.

# 6.3 Definition der erfolgreichen Lokalisation

Es soll eine Grenze für den Winkel der Fehllokalisation festgelegt werden, ab der die Richtungslokalisation als für das Szenario erfolgreich gilt. Grundüberlegung dazu ist, dass der Prozess der Lokalisation für den Spieler so abläuft:

- Er nimmt ein Hörereignis wahr.
- Lokalisiert er es in seinem Kopf, oder diffus (ist also "ahnungslos"), ist die Lokalisation hier bereits gescheitert.
- Falls er es in einer Richtung außerhalb seines Sichtfeldes lokalisiert, wird er sich in diese Richtung umdrehen; falls innerhalb, nicht.
- Ist die Lokalisation nicht exakt richtig, sieht der Spieler keine visuelle Bestätigung für die Existenz einer Schallquelle in der Richtung des Hörereignisses. Dann gibt es drei Möglichkeiten:
  - Der Lokalisationsfehler beträgt unter 30°, es befindet sich also eine sichtbare Schallquelle im Sichtfeld, weniger als 30° von der Hörereignisrichtung (jetzt deckend mit der Blickrichtung) entfernt. Der in Abschnitt 2.5 besprochene Bauchredner-Effekt sorgt dafür, dass der Wiederspruch aufgelöst wird, die Lokalisation ist erfolgreich.

- Der Lokalisationsfehler liegt über 30°, jedoch unter 55°. Bis zu diesem liegt die Schallquelle noch innerhalb des Sichtfeldes, da das Sichtfeld selbst das doppelte, 110° weit ist. Zwar decken sich in diesem Fall Hörereignis und visueller Cue nicht, aber immerhin hat sich der Spieler soweit in die richtige Richtung gedreht, dass er den Gegner im Sichtfeld hat und auf ihn reagieren kann.
- Der Lokalisationsfehler liegt über 55° oder eine Vertauschung liegt vor, das Schallereignis liegt also weiterhin außerhalb des Sichtfeldes. In diesen Fällen ist die Richtungslokalisation für dieses Szenario gescheitert.

Somit liegt bei einem Lokalisationsfehler unter 30° ein Erfolg vor, und unter 55° ein Teilerfolg.

# 7 Sounddesign

In diesem Abschnitt soll das Sounddesign unter Berücksichtigung der Erkenntnisse exemplarisch an einem Klang durchgeführt und dargestellt werden.

#### 7.1 Software

Als DAW für Mischung, Schnitt und als VST-Host stand *Reaper 5.35* von *Cockos* zur Verfügung, für die Klangbearbeitung kamen zum Einsatz:

- Cockos ReaEQ
- mda Envelope
- Klanghelm IVGI
- Audio Assault Defacer
- Audio Assault Transient
- Paul's Extreme Sound Stretch

Zur Visualisierung der Wellenformen wurde *Sony Soundforge 10* und der Spektren *Voxengo Span* verwendet.

# 7.2 Ausgangsklang (Klang A)

Zunächst wurde ein Klang gesucht, der der im Szenario beschriebenen Drohne nahekommt: Metallisch bis gläsern, dem Science-Fiction-Thema entsprechend die Energie eines futuristischen Schwebeantriebs verklanglichend. Darüber hinaus würde es die Rolle der Drohne im Szenario unterstützen, wenn der Sound emotionslos und/oder bedrohlich wirkte.

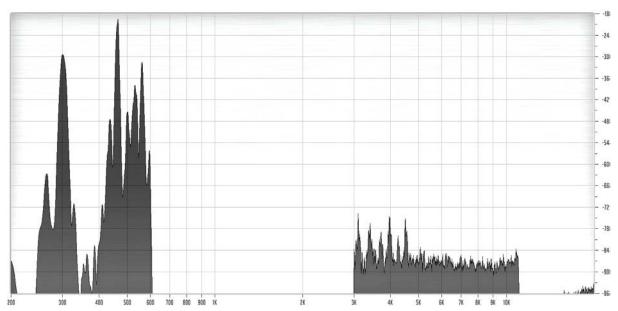


Abbildung 1: Spektrum von Klang A<sup>93</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

 $<sup>^{93}</sup>$  Eigene Grafik, angelehnt an Screenshot von Voxengo Span

Es wurde eine Vorauswahl von geeigneten Klängen aus dem Geräuschearchiv des HdM-Tonstudios *U54* getroffen. Daraus wurde als Grundlage für dieses Kapitel *Klang A* (anzuhören auf dem im Anhang beiliegenden Datenträger) ausgewählt, da er die genannten Anforderungen erfüllt, aber gleichzeitig auch die in dieser Arbeit besprochenen Probleme aufweist:

#### 7.2.1 Problem 1: Bandbreite

Der wesentliche Inhalt besteht aus Signalanteilen im Bereich von etwa 250 bis 600 Hz. Darüber bietet er wenig spektrales Material, das spektrale Cues oder Cues für ITD-HF ermöglichen würde – der in der Spektralanalyse (Abbildung 1) sichtbare Signalanteil von 3 bis 11 kHz ist kaum noch hörbar.

#### 7.2.2 Problem 2: Koinzidenzen

Besagter Haupt-Signalanteil deckt sich mit einem der Richtungsbestimmenden Bänder, dem Vorne-Band (ca. 300 bis 600 Hz). Gleichzeitig existiert eine Lücke zwischen 600 Hz und 3 kHz, in dem sich das Hinten-Band (700 Hz bis 2 kHz) befindet.

#### 7.2.3 Problem 3: Spitzen und Kerben

Klang A hat kaum Rauschanteil, sondern ist – wie bei "gläsernen" Klängen zu erwarten - ausgesprochen tonal; die einzelnen Teiltöne sind in der Spektralanalyse (Abbildung 1) als deutliche Spitzen zu erkennen. Im Frequenzbereich unterhalb von 2 kHz sind solche spektralen Eigenschaften jedoch problematisch.

#### 7.2.4 Problem 4: Dynamikverlauf

Der Wellenformdarstellung (Abbildung 2) lässt sich entnehmen: Eine Modulation findet statt, mit etwa 200 ms langen Abständen zwischen den Maxima. Sie ist in dieser Form jedoch wenig hilfreich für die Lokalisation: Sie findet nur im Frequenzbereich von 250 bis 600 Hz statt, die Erkenntnisse in Abschnitt 4.2.2 zum Einfluss der Maxima-Abstände bezieht sich jedoch auf die ITD-HF, also auf deutlich höhere Frequenzen. Auch ist die Modulation frei von scharfen Anstiegen; vielmehr handelt es sich mehr um ein sanftes Pulsieren, was der Höreindruck bestätigte.

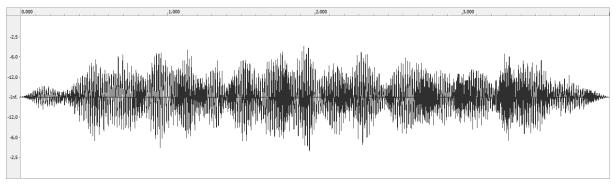


Abbildung 2: Spektrum von Klang A<sup>94</sup> Horizontale Achse: Zeit (s), vertikale: Pegel (dBFS)

# 7.3 Spektrale Bearbeitung

Im Folgenden sollen Strategien betrachtet werden, die ersten 3 Probleme von *Klang A* zu mindern oder zu beseitigen.

<sup>94</sup> Eigene Grafik, angelehnt an Screenshot von Sony Soundforge

#### 7.3.1 Klang B: Entzerrung

Es lag bei unausgeglichenem Frequenzgang zunächst nahe, mit einem Entzerrer/Equalizer zu arbeiten. Inwieweit das erfolgreich ist, hängt aber stark vom Ausgangsmaterial ab. In diesem Fall war bereits bekannt, dass in den höheren Bereichen des Spektrums nur wenig Inhalt vorhanden ist, und auch nur von 3 bis 12 kHz. Weiterhin war zu erwarten, dass der Rauschpegel durch die Bearbeitung deutlich steigt.

Klang A wurde mit ReaEQ, dem internen EQ von Reaper, bearbeitet; ein High-Shelve-Filter wurde bei 1 kHz mit einer Bandbreite von 0.8 Oktaven um 12 dB angehoben.

Das Ergebnis, *Klang B*, war unbefriedigend: Zwar arbeitete der EQ wie gewünscht, und eine Anhebung um 12 dB ist erheblich; aber der Klang war immer noch recht arm an Höhen. Auch die jetzt zum Vorschein getretene Klanginformation passte nur mäßig zum gewünschten Klangeindruck und wirkte etwas wie ein Fremdkörper im Gesamtklang.

Auch sind die Ziele im Wesentlichen nicht erreicht worden: Zwischen 600 Hz und 3 kHz bestand nach wie vor eine große Lücke, Koinzidenzen und auch Spitzen und Kerben sind unverändert.

#### 7.3.2 Klang C: Sättigung

Nichtlineare Verzerrungseffekte verstärken das Signal bis über die Clipping-Grenze, wodurch die Signalform verändert wird und zusätzliche Obertöne entstehen. Es wird also neuer Inhalt generiert, der, so die Hoffnung, das Spektrum so füllt, dass die Probleme beseitigt werden.

Klang A wurde zunächst mit dem VST-Plug-In Klanghelm IVGI bearbeitet, das eine Sättigung nach analogen Vorbildern simuliert. Verzerrungseffekte verändern den Charakter eines Klanges erheblich. Es wurden verschiedene Einstellungen ausprobiert; bei geringen Werten des Hauptparameters *Drive* änderte sich zwar der Klangcharakter, es kamen jedoch nur wenige Anteile in den gewünschten Frequenzbereichen hinzu, und dies eher im Bereich, in dem ohnehin bereits Klanginformation vorhanden war. Bei höheren Werten kamen auch einige hohe Frequenzanteile hinzu, jedoch wurde der Klangcharakter deutlich unangenehm. Ein Wert von 6 stellte einen Kompromiss dar.

Der Klang wurde deutlich verzerrt, wirkte nun auch etwas bedrohlich, was zur gestalterischen Aufgabe passt.

Der Frequenzgang erhielt teilweise neue Klanginformationen, was sich im Spektrum abzeichnet (Abbildung 3). Der Klanganteil oberhalb von 3 kHz war aber nach wie vor sehr leise, mit einem Pegel, der kaum hörbar ist und von dem somit fraglich ist, ob er zur Lokalisation beitragen kann. Außerdem riss bei 11 kHz der Frequenzgang ab, wie auch beim Ausgangsklang. Beides änderte sich auch dann nicht, wenn der Parameter *Response* geändert wurde, der laut Hersteller den Arbeitsbereich der Sättigung auf höhere Frequenzbereiche verlagern soll.

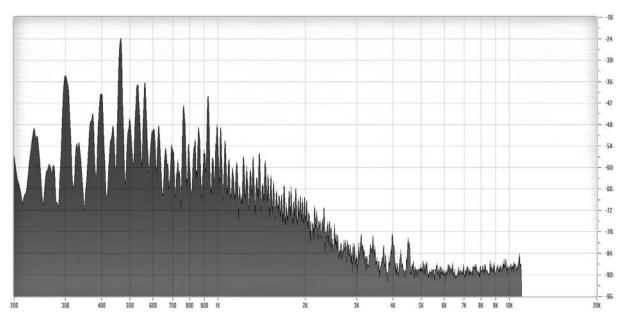


Abbildung 3: Spektrum von Klang C<sup>95</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

Immerhin wurde die Lücke zwischen 600 Hz und 3 kHz, die der Ausgangsklang aufwies, gefüllt; somit ist anzunehmen, dass das Problem der Koinzidenzen gelöst wurde.

Der tiefere Frequenzbereich hat neue Teiltöne dazu erhalten, was die Lücken auffüllte; die markanten Spitzen sind aber nach wie vor vorhanden.

Insgesamt wurde das Ziel der Bearbeitung also nur teilweise erreicht. Das Klangergebnis von Klang C war jedoch positiv, wenn auch zu dem Preis, dass vom Charakter des Ausgangsklanges stark abgewichen wurde.

#### 7.3.3 Klang D: Distortion

Da nichtlineare Verzerrung je nach Algorithmus sehr unterschiedlich ausfallen kann, sollte ein weiteres Plug-In herangezogen werden: *Defacer* der Firma *Audio Assault*. Es wurde auf Klang A angewendet.

Der namensgebende Parameter *Deface* wurde bei der Klangbearbeitung nicht eingesetzt: Er setzt laut Handbuch Sampleblöcke verschiedener Längen in zufälliger Reihenfolge neu zusammen; tatsächlich klingt sie jedoch wie eine Konvertierung in niedrigere Sampleraten, und fügt dem Spektrum erhebliche Aliasing-Artefakte hinzu.

Der Parameter *Distortion* steuert ein Softclipping, und erzeugt dann gute Ergebnisse, wenn man sie mit dem Tiefpassfilter begrenzt, da seine höchsten Verzerrungen unangenehm sind. So werden auch Frequenzanteile hinzugefügt, bei Einsatz des Tiefpassfilters jedoch nur bis 3 kHz.

Die Funktion *Rectify* erzeugt ebenfalls unangenehme Verzerrungen in den Höhen. Die auch von einer Mäßigung mittels Tiefpassfilter profitieren; dieser schneidet hier die hohen Frequenzen jedoch nicht ab, sondern dämpft sie nur. Somit wird der Frequenzgang des Signals bis 20 kHz erweitert.

Durch Mixing von Rectify und Distortion (Parameter *A/B*) sowie dem Originalsignal (Parameter *Mix*) wurde ein Klang erreicht, der dem Originalsignal noch ähnlich ist, dabei aber bedrohlich wirkt; in dieser Einstellung wirkt er überzeugender als *Klang C*.

<sup>&</sup>lt;sup>95</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

Die Spektralanalyse (Abbildung 4) zeigte einen ausgeglichenen Frequenzgang bis 20 kHz ohne Lücken. Ähnlich wie bei der Sättigung sind den Spitzen in den tieferen Frequenzen neue Teiltöne hinzugefügt worden. Auch Koinzidenzen waren hier nicht mehr auszumachen.

Insgesamt wurden die Ziele bei *Klang D* erreicht. Vom ursprünglichen Klangcharakter wurde in Maßen abgewichen und der neue Klang entsprach der Gestaltungsvorgabe.

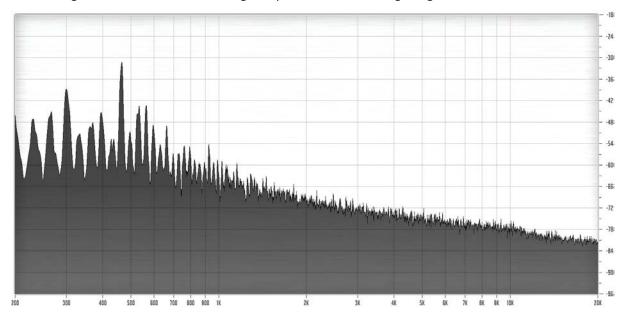


Abbildung 4: Spektrum von Klang D<sup>96</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

#### 7.3.4 Klang E: Layering

Beim Layering wird das Originalsignal mit anderen Aufnahmen gemischt.

Hierfür wurde aus dem Geräuschearchiv eine Aufnahme eines Düsentriebwerks ausgewählt; ihr kontinuierlicher, rauschhafter Klang unterstützt den mechanischen Charakter der Drohne. Sie wurde Klang A hinzugemischt. Das Spektrum wurde dadurch bis auf 17 kHz erweitert.

Bei den Spitzen und Kerben unterhalb von 1 kHz ließ sich beobachten: Die charakteristischen Spitzen des Ausgangsklanges sind noch deutlich vorhanden, die Kerben jedoch mit weiteren Spitzen teilweise aufgefüllt worden.

<sup>&</sup>lt;sup>96</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

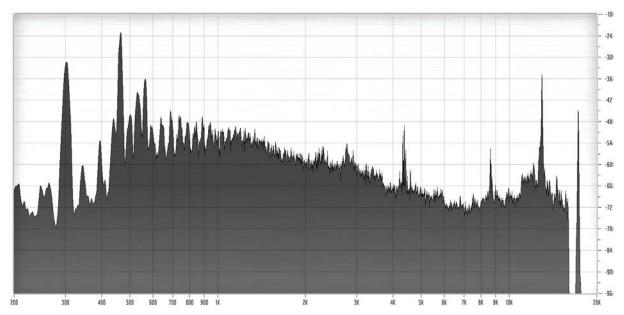


Abbildung 5: Spektrum von Klang E<sup>97</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

Darüber sind die Spitzen bei 4,3, 8,6, 13 kHz und 17kHz auffällig. Sie sind nicht ungewöhnlich bei Aufnahmen von Turbinen; filtert man sie heraus, bleibt fast nur noch Rauschen übrig. Es lässt sich also sagen, sie bilden den Klangcharakter; somit sind sie unverzichtbar. In Abschnitt 4 wurde jedoch festgestellt, dass mikroskopische Eigenschaften des Spektrums nur unterhalb von etwa 2 kHz relevant werden, werden Ausprägungen oberhalb von etwa 2 kHz Oktavbreite aufweisen müssten. Deshalb geht die vorliegende Arbeit davon aus, dass diese Spitzen zu schmal sind, um Auswirkungen zu haben.

Klang E erfüllt die Ziele weitestgehend: Die Lücke wurde gefüllt, das Spektrum stark erweitert. Der Klangeindruck wirkt mechanisch und weist auf einen (Flug-)Antrieb hin, was durchaus passend zum Gestaltungsziel wirkt.

#### 7.3.5 Klang F: Pitch-Layering

Pitch-Layering bedeutet, das ursprüngliche Signal mit demselben oder anderem Material, das in der Tonhöhe verändert wurde (Pitching), zu mischen.

In dieser Arbeit wurde dafür das Open-Source-Programm *Paul's Extreme Sound Stretch* genutzt. Es nutzt die FFT und ist eigentlich dazu gedacht, Time-Stretching um ungewöhnlich hohe Faktoren zu ermöglichen; es bietet jedoch auch eine Funktion namens *Octave Mixer*, mit der sich das Material mit transponierten Kopien seiner selbst mischen lässt. Dadurch, dass das Programm in der Frequenzdomäne arbeitet, werden zeitliche Aspekte – in diesem Fall die bereits vorhandene Modulation – unverändert übernommen, was einen einheitlich wirkenden Klang ermöglicht.

Klang A wurde mit Transpositionen um eine, anderthalb (entsprechend der dritten Harmonischen) und zwei Oktaven gemischt. Der so erzeugte Klang F hat trotz großer Veränderung im Frequenzspektrum noch große Ähnlichkeit mit dem ursprünglichen Signal, der metallisch-gläserne Charakter blieb erhalten.

Das Frequenzspektrum (Abbildung 6) wurde nach oben hin erweitert, wenn es auch dabei abfällt, sodass nur bis etwa 3-6 kHz noch nennenswerte Signalanteile vorhanden sind – gerade die für die Elevation wichtigen Anteile sind also schwach bis nicht vorhanden.

<sup>&</sup>lt;sup>97</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

Es wurde deshalb versucht, die höheren Frequenzen mit weiteren Pitchlayerings aufzufüllen; alternativ wurde auch hochpassgefiltertes Rauschen verwendet. Beides führte jedoch zu klanglich wenig überzeugenden Ergebnissen, sodass davon abgesehen wurde.

Auch gibt es viele ausgeprägte Spitzen und Kerben. Somit handelt es sich bei *Klang F* weiterhin um ein problematisches Signal, wenn auch der Klangcharakter passend ist und dem ursprünglichen, gläsernen Klang nahekommt.

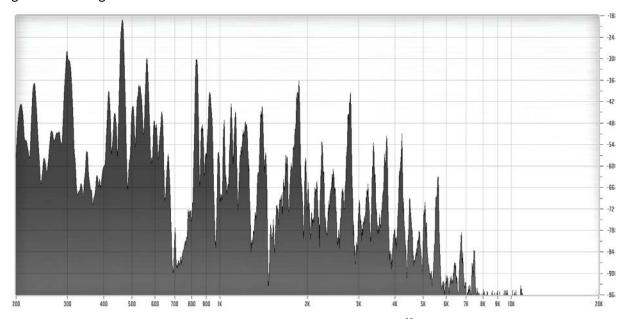


Abbildung 6: Spektrum von Klang F<sup>98</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

## 7.4 Amplitudenmodulation

In diesem Abschnitt soll Problem 4, der Dynamikverlauf des Ausgangssamples, adressiert werden. Im Frequenzbereich der ITD-HF sollten ausgeprägte Transienten mit adäquaten Abständen entstehen. Auch der Frequenzbereich der ITD-LF profitiert von Transienten, die jedoch eine möglichst steile Anstiegsphase haben sollten.

#### 7.4.1 Klang G: Layering

Es wurde passendes Material recherchiert, das Transienten mit den gewünschten Eigenschaften beinhaltet. Dabei wurde das Rotorgeräusch eines Hubschraubers als passend eingeschätzt, da es mit seiner Regelmäßigkeit den maschinenhaften Charakter unterstützt. Das Sample wies jedoch Transienten mit einem Abstand von nur etwa 90 ms auf; deshalb wurde es durch Verlangsamung nach unten transponiert, bis der Abstand 166 ms betrug. Das hatte drei Nebeneffekte: Erstens, dass die obere Grenzfrequenz von etwa 20 kHz mit heruntertransponiert wurde und so der Frequenzgang nur noch bis etwa 11 kHz reichte. Zweitens, dass der Sound etwas verfremdet wurde, was jedoch passend klang. Drittens, dass das Sample jetzt langsamer abgespielt wurde.

<sup>98</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

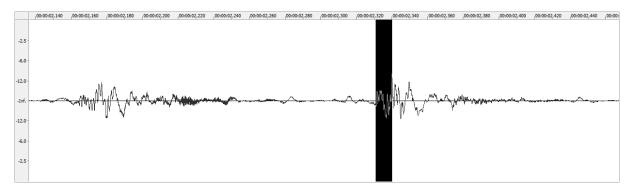


Abbildung 7: Transienten des Helikopter-Samples<sup>99</sup> Schwarz markierter Bereich: 9 ms Horizontale Achse: Zeit (s), vertikale: Pegel (dBFS)

Die Transienten hatten einen Attack von 8 bis 15 ms (siehe Abbildung 7) und wirkten im Bereich von etwa 200 Hz bis 11 kHz.

Das Sample war ein Gemisch aus einer kontinuierlichen Turbine und einem dynamischen Rotor. Damit die Transienten des Rotors sich deutlicher von der Turbine und auch vom Ausgangsklang absetzen, wurde das Sample mit einem Transient Designer bearbeitet; hierfür wurde das Plug-In *Transient* von *Audio Assault* eingesetzt.

Zuletzt wurde noch eine Spitze im Frequenzgang bei etwa 5 kHz entdeckt. Im Gegensatz zu vergleichbaren Spitzen beim Sample in Abschnitt 7.3.4 klang diese jedoch penetrant, und wurde deshalb mit einem EQ deutlich abgeschwächt. Zum Einsatz kam dafür der *ReaEQ* mit einer Pegelreduktion von 18 dB bei einer Frequenz von 5050 Hz und einer Bandbreite von 0,1 Oktaven.

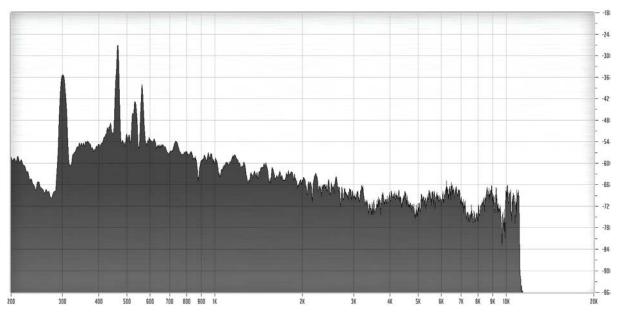


Abbildung 8: Spektrum von Klang G<sup>100</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

Das so bearbeitete Sample wurde abschließend mit Klang A gemischt. In der Spektralanalyse (Abbildung 8) zeigt der so entstandene *Klang G* einen relativ regelmäßigen Frequenzgang, mit

<sup>99</sup> Eigene Grafik, angelehnt an Screenshot von Sony Soundforge

 $<sup>^{100}</sup>$  Eigene Grafik, angelehnt an Screenshot von Voxengo Span

Signalanteilen bis zu 11 kHz. Es sind weiterhin Spitzen und Kerben vorhanden, wenn auch nicht mehr so ausgeprägt.

Dass *Klang G* im tieferen Bereich des Spektrums deutlich herausragende Transienten mit scharfen Attacks besitzt (Abbildung 9), soll bei der Lokalisation über ITD-LF in komplexen Umgebungen helfen, wenn auch die Attackzeit nur bei 8 bis 15 ms statt den idealen 2,5 ms liegt.

Weiterhin sind die Abstände zwischen den Transienten von 166 ms hilfreich bei der Lokalisation in reflexionsarmen Umgebungen über ITD-HF.

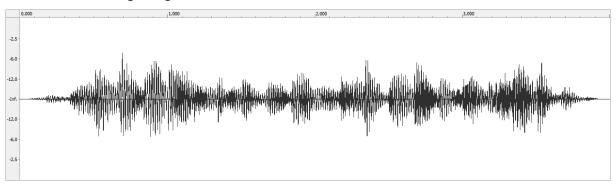


Abbildung 9: Wellenform von Klang G<sup>101</sup> Horizontale Achse: Zeit (s), vertikale: Pegel (dBFS)

Der Klangeindruck des Ergebnisses war jedoch etwas separiert, die beiden Samples verschmolzen nur teilweise zu einem als Einheit wahrgenommenen Klang.

#### 7.4.2 Klang H: Envelope Follower

Deshalb sollte eine andere Herangehensweise versucht werden: Layering mit einem Sample, das zwar keine Dynamik aufweist, jedoch klanglich gut mit dem Ausgangsklang verschmilzt; und ein nachträgliches Hinzufügen der Dynamik.

Dafür wurde ein Klang aus einem der vorigen Abschnitte ausgewählt: *Klang D* (Distortion), da er den besten Frequenzgang aufwies.

Zur Dynamikgenerierung diente ein Envelope Follower, der einem Signal den Pegelverlauf eines anderen Signals aufmoduliert. Das VST-Plug-In *mda Envelope* übernahm diese Aufgabe.

Die Modulationsquelle war ein Bassdrum-Sample mit einer Attackzeit von 10 ms, das alle 166 ms wiederholt wurde. Als Modulationsziel diente Klang D. Als Attack- und Release-Zeit wurden die kürzesten Zeiten, jeweils 1 ms, gewählt, sodass der Follower dem Quellsignal unmittelbar folgte. Der Hub der Modulation wurde mit 9 dB eingestellt.

Rückblickend bestand *Klang D* aus einer Mischung von Klang A und dessen Verzerrung. Nach der Modulation war der aus *Klang A* bestehende Anteil jedoch kaum noch wahrnehmbar. Deshalb wurde hier nachträglich mehr von *Klang A* (unmoduliert) hinzugemischt.

Die Wellenform (Abbildung 10) von *Klang H* zeigt deutliche Transienten; mit dem Envelope Follower ließ sich der Hub frei einstellen, was die Transienten hier deutlicher hervor kommen lässt, als bei *Klang G*. Die Abstände sind identisch, die Attacks vergleichbar mit denen von *Klang G*.

<sup>&</sup>lt;sup>101</sup> Eigene Grafik, angelehnt an Screenshot von Sony Soundforge

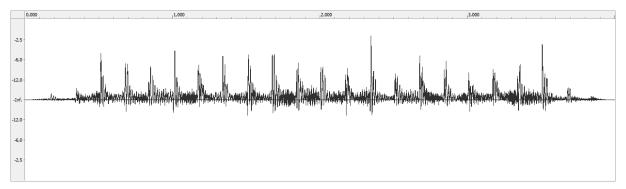


Abbildung 10: Wellenform von Klang H<sup>102</sup> Horizontale Achse: Zeit (s), vertikale: Pegel (dBFS)

Das Spektrum (Abbildung 11) ist, wie zu erwarten war, ähnlich dem des unmodulierten Samples *Klang D* (vgl. Abbildung 4). Lediglich die prominenten Spitzen von 300 bis 600 Herz traten wieder deutlicher hervor, da mehr vom Ausgangsklang hinzugemischt wurde.

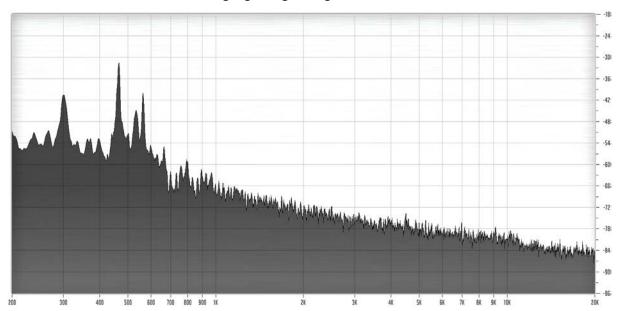


Abbildung 11: Spektrum von Klang H<sup>103</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

Insgesamt sollte es sich bei *Klang H* also um gut lokalisierbares Material handeln. Klanglich ist es passend zum Gestaltungsziel; der bedrohliche Charakter hat durch die Modulation aber etwas abgenommen.

### 7.4.3 Klang I: Envelope Follower

Dasselbe Verfahren sollte noch auf einem anderen Klang als Modulationsziel angewendet werden. Klang F (Pitch-Layering) wurde gewählt, weil er besonders gut mit dem Ausgangsklang verschmolz und klanglich am überzeugendsten war.

Es wurde dasselbe Bassdrum-Sample als Modulationsquelle sowie dasselbe Plug-In genutzt; dabei erwiesen sich die gleichen Einstellungen als adäquat.

35

<sup>&</sup>lt;sup>102</sup> Eigene Grafik, angelehnt an Screenshot von Sony Soundforge

<sup>&</sup>lt;sup>103</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

Klang F wies oberhalb von 3-6 kHz wenig bis keinen Inhalt mehr auf, und der Versuch, dies durch Zumischen von Rauschen zu ändern, scheiterte. Bei der Modulation wurde dies jedoch erneut ausprobiert: Das Rauschen wurde dem Modulationsziel hinzugefügt, also mitmoduliert. Dieses Mal funktionierte dies auch klanglich gut.

In der Spektralanalyse (Abbildung 12) zeigen sich bei *Klang I* somit spektrale Inhalte bis 20 kHz; geblieben sind jedoch die vom Klang F bekannten starken Spitzen & Kerben (vergleiche Abbildung 6).

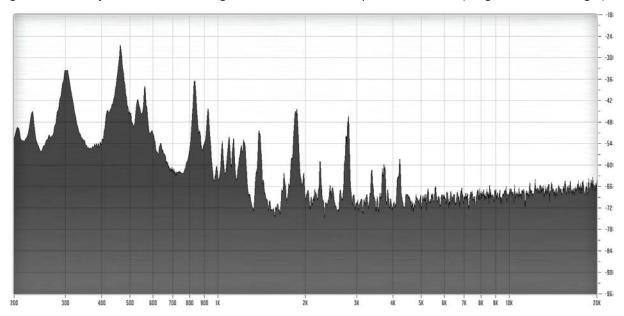


Abbildung 12: Spektrum von Klang I<sup>104</sup> Horizontale Achse: Frequenz (Hz), vertikale: Pegel (dBFS)

Die Wellenform (Abbildung 13) zeigte ausgeprägte Transienten mit ansonsten ähnlichen Werten für Attacks (6-15 ms) und Transienten-Abstand (166 ms).

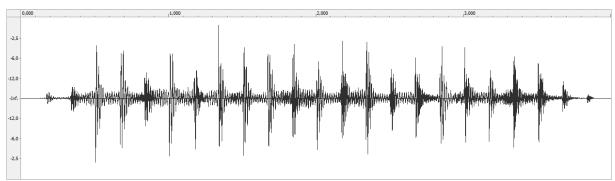


Abbildung 13: Wellenform von Klang I<sup>105</sup> Horizontale Achse: Zeit (s), vertikale: Pegel (dBFS)

Das klangliche Ergebnis ist sehr zufriedenstellend. Abschließend lässt sich sagen, die Dynamik scheint bei *Klang I* ideal, die Spitzen im Frequenzgang jedoch problematisch.

-

<sup>&</sup>lt;sup>104</sup> Eigene Grafik, angelehnt an Screenshot von Voxengo Span

 $<sup>^{105}</sup>$  Eigene Grafik, angelehnt an Screenshot von Sony Soundforge

## 7.5 Zusammenfassung

Es wurde eine Reihe von Klangbearbeitungen vorgenommen. Die beim ursprünglichen Klang A festgestellten Probleme konnten dadurch teilweise oder ganz gelöst werden.

Das Problem der Bandbreite ließ sich besonders durch Distortion gut lösen. Ebenso durch Layering, auch wenn es stark davon abhängt, wie geeignet das Sample ist. Pitch-Layering konnte nur in Maßen helfen: Der Frequenzgang ließ sich nur um 2 Oktaven erweitern.

Das Spektrum von Klang A war insofern ein Problem gewesen, dass sich der wesentliche Signalanteil mit dem Vorne-Band, sowie eine Lücke mit dem Hinten-Band gedeckt hatten. Diese Lücke ließ sich erfolgreich füllen, sei es mit nichtlinearer Verzerrung, Layering oder Pitch-Layering. Die Deckung mit dem Vorne-Band konnte jedoch höchstens abgemildert werden: Solange Klang A Hauptbestandteil des bearbeiteten Klanges war, blieb auch die Betonung des Bereiches von 250 bis 600 Hz vorhanden und damit die Gefahr, dass der Klang fälschlicherweise vorne lokalisiert wird. Der Spektralanalyse zufolge hat Klang D noch am erfolgreichsten abgemildert; das überrascht nicht, da die Distortion den Klang am stärksten verfremdet hat.

Gleiches gilt für die Spitzen & Kerben im Frequenzgang unterhalb von 2 kHz, die sich zwar abmildern, aber nicht beheben ließen. Sie wurden in einigen Verfahren im Rauschanteil anderer Signale "versteckt" oder mit erzeugten Obertönen gemischt.

Den Dynamikverlauf zu gestalten war wiederum sehr erfolgreich, sowohl durch Layering als auch mit dem Envelope Follower. Dabei gab das Layering natürlichere Ergebnisse, während es mit dem Envelope Follower einfacher war, für starken Hub der Transienten vom Rest des Signals oder für besonders scharfe Attacks zu sorgen.

Insgesamt war es ein schwieriges Unterfangen, zwei verschiedenen Aspekten gerecht zu werden: Den Wunsch, Klänge stimmig zu gestalten, und ihnen lokalisationsfördernde Eigenschaften zu verleihen. Hier muss im Gestaltungsprozess immer wieder neu abgewägt werden. Wenn auch eine stimmige Gestaltung für Sounddesigner stets Vorrang haben dürfte, schlossen sich in einigen Fällen beide Aspekte nicht aus, sodass die Berücksichtigung der Erkenntnisse der vorliegenden Arbeit Vorteile zeigte.

## 8 Erhebung

In vorherigen Abschnitt wurde untersucht, inwiefern sich Klänge unter Berücksichtigung der Besonderheiten der Richtungslokalisation gestalten lassen. Eine Erhebung sollte nun prüfen, inwieweit diese gestalterischen Maßnahmen ihr Ziel erreicht haben. Bringen die so erarbeiteten Klänge einen Unterschied in der Richtungslokalisation in Hörsituationen, wie sie in aktuellen VR-Games vorzufinden sind?

## 8.1 Eigenschaften der Erhebung

Um dieser Frage angemessen nachzugehen, wurden einige Anforderungen an die Erhebung wie folgt identifiziert:

- Die Erhebung sollte in VR stattfinden, um in einer Umgebung zu erheben, die dem beschriebenen Szenario nahekommt.
- Die Eingabe durch die VP sollte intuitiv sein und sich ebenfalls am Szenario orientieren. Bevorzugt wurde dabei in Erwägung gezogen, dass die VP die HER eingeben, in dem sie ihren Kopf in diese Richtung orientieren. Diese Methode wurde von Makous & Middlebrooks vorgeschlagen. 106 Sie wurde von Gilkey mit zwei anderen Methoden (Zeigen auf einer Kugel, mündlicher Bericht) verglichen und ist diesen in Genauigkeit und Geschwindigkeit überlegen. 107
- Zum Einsatz sollte eine Binauralsynthese-Software kommen, die auch in einem VR-Spiel eingesetzt würde.
- Nichtakustische Cues sollten ausgeschlossen sein.
- Es sollte, neben der HER, auch erhoben werden, bei welchen Klängen und Richtungen Phänomene wie IKL und Vertauschungen auftreten.
- Die Gruppe der VP sollte nicht nur aus Studenten bestehen; auch sollten Menschen ohne professionellem Schwerpunkt auf Ton oder Games dabei sein; nicht zuletzt sollten die Menschen maximal geringe Hörschäden aufweisen.
- Der Ablauf der Erhebung sollte sich zeitlich nach den Bedürfnissen der VP richten, um sie nicht unnötig unter Druck zu setzen. Sie sollten also nicht zur Eingabe innerhalb einer bestimmten Zeitspanne gezwungen werden.

#### 8.1.1 Dauer der Erhebung

Einerseits sollten möglichst viele VP mittels vieler Stimuli aus vielen unterschiedlichen Richtungen untersucht werden, um ein möglichst differenziertes Bild zu erhalten. Andererseits sollte der Zeitaufwand, sowohl für die VP als auch für den Autor, in einem sinnvollen Rahmen bleiben. Deshalb wurde in folgenden Aspekten abgewägt:

#### Dauer eines Erhebungsdurchlaufs

Bech und Zacharov weisen darauf hin, dass VP besonders in repetitiven Aufgabenstellungen Ermüdung und Langeweile unterliegen, und empfehlen deshalb für Hörtests eine Dauer von 20 Minuten, halten aber auch Längen von bis zu 40 Minuten für akzeptabel. 108

<sup>108</sup> vgl. Bech & Zacharov 2006, S. 302

<sup>&</sup>lt;sup>106</sup> vgl. Makous & Middlebrooks 1990, S. 2197–2198

<sup>&</sup>lt;sup>107</sup> vgl. Gilkey et al. 1995, S. 9–10

Im Vortest war der Hörtest daraufhin zunächst auf 30 Minuten angelegt. Dabei hat sich jedoch gezeigt, dass die meisten Vortester dies bereits als erschöpfend empfanden. Sie gaben die körperliche Aktivität beim Eingabevorgang und die Belastung, die längeres Tragen des HMDs verursacht, als Gründe an. Für die Erhebung wurde deshalb versucht, die Dauer des Hörtests auf etwa 20 Minuten zu senken.

Für die Begrüßung und Verabschiedung, Ausfüllen des Fragebogens, Lesen der Anleitung und Aufsetzen der VR-Ausrüstung wurde eine ungefähre Dauer von 15 Minuten ermittelt. Außerdem wurde ein Sicherheitspuffer von 10 Minuten festgelegt.

Daraus resultierte als Ziel eine Taktung der Termine im Abstand von 45 Minuten.

#### 8.1.1.2 Dauer der Stimuli

Längere Stimuli geben mehr Information über ihre Richtung als kurze. Dennoch sollte sich die Länge der Stimuli an der im Szenario erörterten Spielsituation orientieren: In actionreicheren VR-Games ist damit zu rechnen, dass der Spieler nicht besonders viel Zeit hat, ein HE zu lokalisieren. Also sollte diese Erhebung eher kurze Stimuli einsetzen.

Deshalb, und weil die Dauer der Erhebungsdurchlaufs kurz gehalten werden sollte, wurde die Länge der Stimuli auf 4 Sekunden festgelegt.

#### 8.1.1.3 Anzahl der Richtungen

Es sollten die Vertikalebene und mindestens noch eine weitere Ebene berücksichtigt werden, um sowohl die Performance über monaurale, als auch interaurale Merkmale betrachten zu können. Als zweite Ebene neben der Medianebene wurde die Horizontalebene ausgewählt, weil in dieser Ebene der Großteil des Spielgeschehens vorkommt, und weil sich die Mehrzahl der Forschungsarbeiten zum Thema interaurale Merkmale auf diese Ebene beziehen.

Es wurde entschieden, dass die Unten-Position ausgelassen werden kann, weil nicht damit zu rechnen ist, dass ein HE genau senkrecht unten lokalisiert wird, wo der Körper des Probanden im Weg wäre.

In den Ebenen ist noch die Anzahl der diskreten Schallquellenpositionen relevant. Eine hohe Anzahl würde ein feiner aufgelöstes Resultat geben, würde die Dauer der Erhebung jedoch auch steigen lassen. Wählt man sie zu gering, besteht die Gefahr, dass die VP merken, dass die SE immer wieder an den gleichen Stellen passieren; ihre Eingaben könnten deshalb in diese Richtungen tendieren.

Es wurde entschieden, dass ein Abstand von 45° noch ausreichend wäre. Die sechs VP des Vortests wurden am Ende befragt, ob sie den Verdacht gehabt hatten, die Signale wären aus diskreten Richtungen gekommen; alle verneinten dies. Deshalb wurde davon ausgegangen, dass der befürchtete Effekt nicht oder nur im geringen Maße auftreten würde.

In Abbildung 14 sind die sieben im Versuchsaufbau verwendeten Schallquellenpositionen in der Medianebene veranschaulicht, und in Abbildung 15 die acht Schallquellenpositionen in der Horizontalebene.

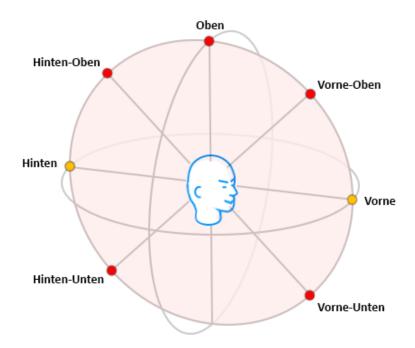


Abbildung 14: Versuchsanordnung der virtuellen Schallquellen in der Medianebene <sup>109</sup> Die Vorne- und die Hinten-Positionen sind farblich anders (orange) gekennzeichnet, da sie an den Schnittpunkten mit der Horizontalebene sitzen und so in beiden Ebenen vorkommen.

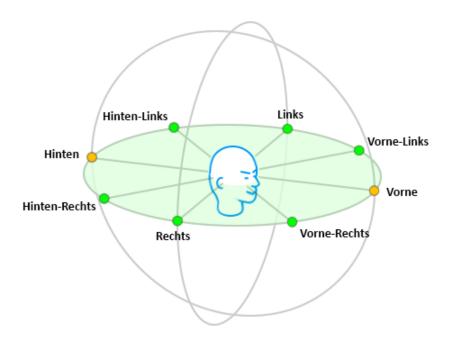


Abbildung 15: Versuchsanordnung der virtuellen Schallquellen in der Horizontalebene 110

Es wurde festgestellt, dass sich die Horizontalebene entlang der Vorne-Hinten-Achse auf eine Hälfte "zusammenklappen" lässt, also die Ergebnisse der linken und rechten Hälfte zusammenführen lassen. Das beinhaltet, den Kopf vereinfachend als ideal symmetrisch zu betrachten und damit die Ergebnisse für unabhängig davon, auf welcher der Hälften sie gemessen werden. Da dieses Vorgehen eine doppelte Datenmenge gegenüber den anderen Schallquellenpositionen ergeben würde, wurde

\_

<sup>&</sup>lt;sup>109</sup> Eigene Grafik, angelehnt an Blauert 1974, S. 14

<sup>&</sup>lt;sup>110</sup> Eigene Grafik, angelehnt an ebd., S. 14

entschieden, dass es ausreichte, diese Positionen nur mit halber Häufigkeit einzusetzen. In der späteren Datenverarbeitung und -visualisierung wurden die erhobenen Daten gemeinsam unter den Positionsbeschreibungen Vorne-Seitlich, Seitlich, und Hinten-Seitlich gehandhabt.

Somit liegt die Zahl der Schallquellenpositionen bei zehn; davon sieben in der Medianebene, und drei in der Horizontalebene.

#### 8.1.2 Größe der Versuchsgruppe

Zunächst wurde als Ziel festgelegt, mindestens 20, idealerweise aber 40 VP zu untersuchen. Der Takt von 45 Minuten und das Ziel, den Zeitaufwand bei etwa zwei bis drei Tagen zu halten, gab jedoch eher eine Gruppengröße von 20 bis 30 VP vor. Letztendlich meldeten sich 24 Personen für den Test; davon wurde eine Person aufgrund eines Hörschadens auf einem Ohr ausgeschlossen, und drei erschienen nicht zu ihrem Termin.

Die Versuchsgruppe umfasste somit 20 Personen, was noch innerhalb der Zielgröße lag.

#### 8.1.3 Bewegungsfreiheit

Weil Kopfbewegungen beim Richtungshören helfen und Headtracking in VR-Games Standard ist, wäre es naheliegend, freie Kopfbewegungen in der Erhebung zu erlauben. Jedoch entsteht dabei ein Dilemma: Soll z.B. mit einem von vorne abgespielten Stimulus das Richtungshören in der Vorne-Richtung gemessen werden, und dreht die VP während dem Abspielen den Kopf, sodass der Stimulus z.B. rechts von ihm liegt, so würde nun das Richtungshören nach rechts gemessen.

Brimijoin et al. haben Kopfbewegungen beim Richtungshören gemessen; durchschnittlich bewegten VP ihren Kopf um 29°, wobei die Standardabweichung bei ±12° lag. 111 Bei einer Kopfbewegung um 29° wäre in dem Versuchsaufbau dieser Arbeit bereits eine benachbarte Schallquelle näher an der Blickrichtung als die ursprüngliche in Vorwärtsrichtung.

Der Kompromiss lautet, Kopfbewegungen zuzulassen, aber soweit einzuschränken, dass die Richtungen überwiegend dieselben bleiben. Als Maßstab wurde hier der Abstandswinkel der diskreten Schallquellenpositionen untereinander gewählt. Da er 45° beträgt, soll der Blickwinkel maximal um die Hälfte, also um 22,5° abweichen – So ist gewährleistet, dass er nie näher an einer anderen diskreten Schallquellenposition ist als an der Vorne-Richtung.

#### 8.1.4 Reihenfolge & Wiederholungen

Die erwähnten Lerneffekte können verursachen, dass Stimuli, die später in der Erhebung bzw. zum wiederholten Mal gespielt werden, besser abschneiden, als andere. Um eine Verzerrung der Ergebnisse zu verhindern, sollten die Stimuli und Richtungen in zufälliger Reihenfolge abgespielt werden.

Aus gleichem Grund sollte wiederholtes Abspielen der Stimuli auf Wunsch der VP nicht zugelassen werden.

-

<sup>&</sup>lt;sup>111</sup> vgl. Brimijoin et al. 2013, S. 3

## 8.2 Vorbereitung

#### 8.2.1 Auswahl der Stimuli

Aus Abschnitt 7 haben sich neun Klänge ergeben, sie sollten getestet werden. Klang B wurde jedoch herausgenommen, da die Bearbeitung wenig erfolgreich und der Unterschied zum Klang A zu gering war. Es wurden also Klang A, sowie Klänge B bis I als Stimuli eingesetzt.

Es handelt sich um kontinuierliche Signale, die im Szenario dauerhaft (in Schleife) abspielen würden. Für diese Erhebung ist das jedoch nicht praktikabel. Deshalb wurden sie jeweils auf eine Länge von 4 Sekunden gekürzt.

Onsets ermöglichen die Lokalisation in komplexen Umgebungen. Sie sind für diese Erhebung jedoch kontraproduktiv, da kontinuierliche Signale untersucht werden sollen, und sie nur deshalb einen Onset haben, da sie für diese Erhebung gekürzt wurden. Deshalb wurden ihre Anfänge und Enden mit einer Fade-Zeit von 500 ms ein- bzw. ausgeblendet.

Um den Einfluss von Lautstärkeunterschieden zu minimieren, wurden die Klänge nach der EBU-Norm R128<sup>112</sup> auf -23 LUFS gepegelt.

#### 8.2.2 Software

Das VR-Labor wurde in der Unreal Engine erstellt und ausgeführt, in einer *Custom Version*, basierend auf der Version 4.12.5. Sie ist gegenüber der Originalversion nur in organisatorischen Aspekten angepasst, was diese Erhebung nicht berührt.

Als Audio Engine kam FMOD Studio zum Einsatz, da es für den Sounddesigner bequeme Verwaltung der Sound Events ermöglichte, und als Host für das Plug-In 3Dception der Firma TwoBigEars diente, das die Binauralsynthese durchführte.

#### 8.2.3 Hardware

#### 8.2.3.1 VR-System

Dem Gameslab der HdM stehen HMDs von zwei Herstellern zur Verfügung: Die HTC Vive und die Oculus Rift. Für diese Erhebung wurde die Vive in der Version "Consumer Edition" aufgrund folgender Eigenschaften ausgewählt:

- Während die Oculus aufliegende (supraaurale) Kopfhörer mit unbekannten Spezifikationen fest verbaut hat, bringt die Vive keine eigenen Kopfhörer mit, womit die Wahl der Kopfhörer für die Erhebung frei ist.
- Ihr Trackingsystem ist dem der Oculus überlegen. Das der Oculus funktioniert in die Vorne-Richtung besser, während es zu ungenauen, hängenden oder springenden Trackinginformationen führen kann, wenn der Benutzer sich nach hinten orientiert. Das Trackingsystem der Vive hat sich bewährt, solange man reflektierende Flächen wie z.B. Glasflächen vermeidet.

Das HMD der Vive ist mittels eines Kabelstrangs, der es mit HDMI-Bildsignal, analogem Tonsignal und Stromversorgung versorgt, an einen PC verbunden.

#### 8.2.3.2 Kopfhörer

Der für die Erhebung eingesetzte Kopfhörer sollte folgende Anforderungen erfüllen:

\_

<sup>&</sup>lt;sup>112</sup> EBU Europäische Rundfunkunion 2014

- Offene Bauweise: So können die VP mit dem Versuchsleiter kommunizieren, ohne den Kopfhörer abnehmen zu müssen. Außerdem trägt die offene Bauweise durch weniger Hitzestau zum Tragekomfort bei.
- Ohrumschließende (circumaurale) Bauweise
- Diffusfeldentzerrung: Laut Theile wird der Frequenzgang eines Signals im Diffusfeld nicht abhängig von der Klangeinfallsrichtung verzerrt.<sup>113</sup>
- Gute Klangeigenschaften und neutraler Frequenzgang
- Allgemein hoher Tragekomfort

Da er diese Anforderungen erfüllt, wurde der DT 990 Pro von Beyerdynamic gewählt.

#### 8.2.3.3 DA-Wandler

Als Audio-Interface diente das eingebaute USB-Interface der Vive, das den Vorteil bietet, dass der Benutzer nur einen Kabelstrang mit sich umhertragen muss.

#### 8.2.4 Die Labore

Das Labor, in dem die Erhebung stattfinden sollte, muss als zwei Räume betrachtet werden: Dem realen und dem virtuellen.

#### 8.2.4.1 Das reale Labor

Als reales Labor diente das Tonstudio des Gameslab der HdM. Dank Akustikelementen an der Decke, Molton-Vorhängen an zwei von vier Wänden, und Teppichboden ist die Akustik akzeptabel (wenn auch sie nicht professionellen Ansprüchen genügt), da die Stimuli ja nicht im realen Raum abgespielt wurden; es war lediglich das Ziel, einen eher "trockeneren" Raum auszuwählen, damit der Unterschied in der Raumakustik zwischen realem und VR-Labor nicht eklatant und damit eventuell irritierend auf die VP wirken könnte.

Auch die Isolation ist nicht perfekt, aber gut genug, dass eine Beeinflussung der Erhebung durch Störgeräusche von außen unwahrscheinlich war.

Als Störquelle kam der Computer in Frage, der zwei leistungsfähige Grafikkarten und dementsprechend dimensionierte Lüfter hat. Da das VR-Labor dank seiner grafischen Schlichtheit für eine VR-Anwendung verhältnismäßig wenig Grafikleistung benötigte, wurden die Drehzahlen der Lüfter relativ weit heruntergeregelt und somit war nur ein leichtes, gleichmäßiges Rauschen vernehmbar.

Für den Versuchsleiter wurde außerhalb des Trackingbereichs ein Platz mit Zugriff auf den PC eingerichtet. Auf dem Bildschirm waren dabei zu keinem Zeitpunkt Daten zu sehen.

Die Basisstationen des Trackingsystems wurden auf einander gegenüberliegenden Schränken montiert.

<sup>&</sup>lt;sup>113</sup> vgl. Theile 1981, S. 169



Abbildung 16: Das reale Labor mit VP und Arbeitsplatz des Versuchsleiters<sup>114</sup>

#### 8.2.4.2 Das VR-Labor

Das VR-Labor und der Programmcode der Erhebung wurde, nach Vorgaben des Autors der vorliegenden Arbeit, von Kasimir Blust in der Unreal Engine erstellt. Dabei wurde teilweise auch auf bereits vorhandene Grafikelemente und Code aus der HdM-Studioproduktion *Elena* zurückgegriffen.

Der VR-Raum wurde als Würfel mit einer Kantenlänge von 3,30 Metern gestaltet. Diese Größe entspricht in etwa dem Trackingbereich, der im realen Labor maximal möglich war, und lässt den Mittelpunkt des Systems aus Kopf und HMD bei einer durchschnittlich großen Person in etwa in der Raummitte sein – so hat ein Mensch mit einer Körpergröße von 1,75 Metern diesen Mittelpunkt etwa auf einer Höhe von 1,65 Metern.

In die Ecken des Raumes wurden ein paar Gegenstände platziert, die keine weitere Funktion hatten, aber der Orientierung im Raum dienen sollten, indem sie verhindern, dass jede Ecke gleich aussieht. Dabei sollten Gegenstände verwendet werden, die man als klanglos kennt, damit man sie nicht irrtümlich für den Ursprung eines Klangs halten könnte. Somit wurden Pflanzen und ein Mülleimer ausgewählt. Außerdem kam ein Basketball dazu, dessen Funktion später besprochen wird.

Die Hauptfunktionalität des VR-Labors bietet ein Set aus "Indikatoren", die dazu dienen, den Bewegungsspielraum der VP während dem Abspielen der Stimuli einzugrenzen. Sie nehmen entweder die Farbe Rot oder Grün an.

Der Indikator "Blickrichtung" registriert, ob der Kopf der VP in die richtige Richtung zeigt. Befindet sich der "Cursor", ein blau leuchtender Punkt, der stets in der Mitte des Sichtfeldes des Probanden eingeblendet ist, deckend mit dem Indikator, ist die Blickrichtung innerhalb der Grenzen, die in Abschnitt 8.1.3 bestimmt wurden, und die Farbe des Indikators ist grün. Zeigt der Cursor auf einen Punkt außerhalb des Indikators, wird dieser Rot.

Ist dieser Indikator für die Rotation des Kopfes, so ist der andere verantwortlich für die Translateration, also die Richtungen vor und zurück, links und rechts. Gemessen wird diese

-

<sup>114</sup> Eigenes Foto

Abweichung der X- und Y-Koordinate der VP von der Idealposition, die im Bild mit Standpunkt markiert ist. Hier soll die VP stehen, und sich nicht zu weit davon entfernen. Tut sie es doch, wechselt die Farbe des Indikators von Grün nach Rot.

Auch die Standpunktmarkierung wechselt dann ihre Farbe, sie und der Indikator zeigen also immer dieselbe Farbe an. Im Vortest hatte sich gezeigt, dass die VP meist nach unten gucken, wenn sie die Position das erste Mal einnehmen. Wenn sie während der laufenden Erhebung jedoch nach vorne guckten, sollten sie auch Feedback bekommen, ob sie noch richtig standen, ohne dass sie jedes Mal nach unten gucken müssten. Deshalb wurde dieses System aus zwei synchronisierten Anzeigen gewählt.

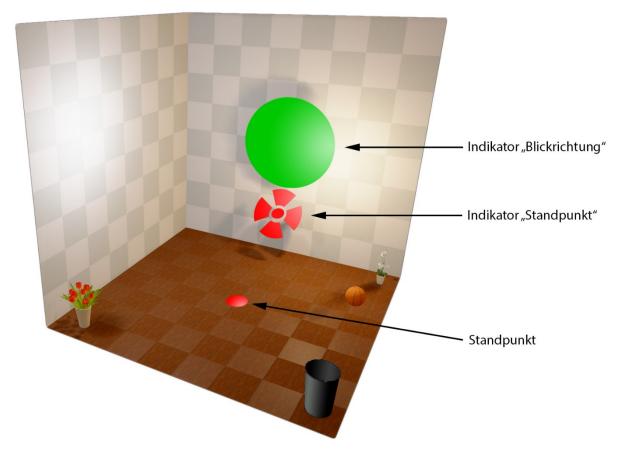


Abbildung 17: Das VR-Labor 115

Der Spielraum in Bewegung und Blickwinkel, den dieses System noch zuließ, sollte so gering wie möglich sein, aber gerade noch so groß, dass er sich einhalten lässt, ohne dass die VP "versteifen". So wurde der Spielraum des Standpunktes zunächst auf 5cm Radius festgelegt; nach dem Vortest (vergleiche Abschnitt 8.2.10) wurde er auf 15cm vergrößert.

Die Größe des Blickwinkel-Indikators wurde so gewählt, dass sie selbst genau die Zone darstellt, in der sich der Cursor des HMD bewegen darf, will die VP die Abweichung ihres Blickwinkels unter den bereits festgelegten 22,5° halten. In anderen Worten: Geht der blaue Punkt über den Rand des Indikators hinaus, ist er mehr als 22,5° vom Mittelpunkt entfernt, und der Indikator wird Rot.

Solange nicht alle Indikatoren grün sind, werden keine Stimuli gestartet.

<sup>&</sup>lt;sup>115</sup> Eigene Grafik, angelehnt an Screenshot der Unreal Engine

#### 8.2.5 Akustik des VR-Labors

Die Raumakustik des VR-Labors wurde mittels des Room-Modelling-Features von 3Dception erzeugt. Dabei wurden, basierend auf dem "Living Room"-Preset, folgende Parameter eingestellt:

- Diffusion: 1,0

Reflexion hoher Frequenzen: 0,8Reflexionsfaktor für jede Wand: 0,78

Der simulierte Raum hatte eine Kantenlänge von 165 cm.

#### 8.2.6 Die virtuellen Schallquellen

Den festgelegten Schallquellenpositionen entsprechend wurden *Audio Sources* um den Kopf der VP angeordnet, alle im gleichen Abstand von 150 cm.

Dieses virtuelle Lautsprechersystem wurde zu Beginn jedes Erhebungsdurchlaufs auf die vertikale Position des HMDs, die aufgrund der unterschiedlichen Körpergrößen der VP variierte, einmalig kalibriert (siehe Abschnitt 8.4.2).

Das Lautsprechersystem war für die VP unsichtbar.

#### 8.2.7 Reihenfolge der Stimuli

Die Reihenfolge der Stimuli und aus welcher Richtung sie abgespielt wurden, war vorab mit Zufallszahlen von Random.org<sup>116</sup> festgelegt. Jede VP erhielt eine neue, zufällige Reihenfolge. Zu keinem Zeitpunkt waren dem VP oder dem Versuchsleiter die Reihenfolgen bekannt.

#### 8.2.8 Fragebogen

Mittels eines Fragebogens galt es, einige Informationen über die VP herauszufinden:

Damit der Versuchsleiter wusste, ob er den VP eine Einführung zu VR geben sollte (z.B. wie man unfreiwilliges Stoßen gegen reale Wände verhindert), wurde nach vorheriger Erfahrung mit VR gefragt.

Aufschluss über die Hörfähigkeiten der Gruppe der VP sollte die Frage geben, ob sie über ein geschultes Gehör verfügen. Außerdem, ob es durch ein Studium oder Beruf mit Schwerpunkt Ton, oder auf anderem Wege (Angabe in eigenen Worten, z.B. Gesangs- oder Instrumentalunterricht) geschult wurde.

Um Menschen mit einem erheblichen Hörschaden aus der Gruppe der VP auszuschließen, wurde danach gefragt.

Zuletzt belehrt der Fragebogen darüber, dass keine audiovisuellen Aufnahmen angefertigt werden und die Daten Name und E-Mail-Adresse nicht veröffentlicht werden, und dass die VP die Erhebung jederzeit unterbrechen darf, sollte ihr unwohl sein.

Der Fragebogen ist im Anhang dieser Arbeit zu finden.

#### 8.2.9 Anleitung

Die VP brauchten eine Anleitung zu den Mechanismen des VR-Labors und zum Ablauf der Erhebung.

Beim Vortest wurde diese Anleitung im persönlichen Gespräch mit dem Versuchsleiter erteilt. Dabei hat sich gezeigt, dass es die VP überfordert. Deshalb wurde dieser Schritt durch eine schriftliche

<sup>&</sup>lt;sup>116</sup> vgl. Randomness and Integrity Services Ltd. 2017

Anleitung ersetzt, die sich die VP vor der Erhebung in eigenem Tempo durchlesen konnten. Zahlreiche Bilder sollten dabei dem Verständnis helfen.

- Unter anderem wurden darin den VP die drei möglichen Kopfbewegungen aus bewusst gemacht, sie wurden über ihre förderliche Wirkung auf das Richtungshören aufgeklärt, und aufgefordert, sie auszuprobieren.
- Auch steht darin, dass die Stimuli aus allen Richtungen kommen können.
- Entgegen den Tatsachen wird konstatiert, dass es sich um kontinuierliche statt diskrete Richtungen handelt. Dies sollte dem Effekt vorbeugen, dass die VP zu einigen wenigen, diskreten Richtungen tendieren.

Die Anleitung ist ebenfalls im Anhang dieser Arbeit zu finden.

#### 8.2.10 Vortest

Es wurde ein Vortest mit 6 VP durchgeführt.

Neben den bereits genannten Erkenntnissen aus diesem Vortest, ist die bemerkenswerteste Angabe der VP die, dass

- sich ein Teil der HE zu bewegen schienen; sei es, dass sie sich kontinuierlich durch den Raum bewegten, oder zwischen zwei Positionen sprangen.
- sich einige wenige Klänge in zwei räumlich getrennte HE spalteten; so wurde z.B. beschrieben, dass die hohen Frequenzanteile aus einer Richtung, die tieferen aus einer anderen Richtung kämen, und dass der Klang dadurch als zwei separate HE wirkte.

Um diesen Beobachtungen nachzugehen, wurde der Versuchsaufbau so geändert, dass die VP auch die Möglichkeit bekamen, HE mit den Eingaben "bewegt sich" und "ist in mehreren Richtungen", zusätzlich zu den anderen möglichen Eingaben, zu markieren; auch dies geschah als mündliche Ansage an den Versuchsleiter.

Außerdem wurde im Vortest festgestellt:

- Durchschnittlich nahmen sich die VP 14,5 Sekunden Zeit, um ihre Eingabe zu machen (Dauer des Stimulus mitgerechnet). Bei 80 Eingaben führt das zu einer Dauer eines Erhebungsdurchlaufs von etwa 19 Minuten.
- Viele VP gaben an, dass es sehr anstrengend sei, sich für die Eingabe der hinten gelegenen HER umzudrehen. Das hing damit zusammen, dass die meisten VP sich in der Hüfte drehten, die Füße dabei aber in der gleichen Position ließen, und sich so sehr verrenken mussten; dies passierte wohlgemerkt, ohne dass sie dazu aufgefordert wurden, es so zu tun. Es wird vermutet, dass die Erklärungen zur Einschränkung der Bewegungsfreiheit dazu geführt hatten, dass die VP im Zweifelsfall eher möglichst wenig an ihrer Position verändern.
  - Um auszuschließen, dass die VP die Hinten-Richtungen meiden, und um Verletzungen vorzubeugen, wurde in der Anleitung ein Hinweis eingefügt, der explizit das Umdrehen mit Bewegung der Füße erlaubt.
- Zwei VP hatten Schwierigkeiten, sich so auf den Standpunkt zu stellen, dass dieser bei kleinen Bewegungen nicht ständig zwischen Rot und Grün hin- und hersprang. Daraufhin wurde der erlaubte Radius von 5 auf 15 cm vergrößert.
- Einige VP haben ihren Kopf während dem Abspielen der Stimuli stark bewegt, andere überhaupt nicht. Auf Nachfrage wurde angegeben, dass es einfach vergessen wurde. Auch wurde beobachtet, dass die Bewegungen fast ausschließlich aus Drehen und Nicken bestanden, während Kippen kaum vorkam. Deshalb wurde in die Anleitung ein Abschnitt eingefügt, der die drei Arten der Kopfrotation noch einmal mit Bildern vorstellte und darauf hinwies, dass sie beim Richtungshören helfen.

#### 8.2.11 Versuchspersonen

Aus der Auswertung der Fragebögen wissen wir:

Die VP sind im Durchschnitt 26 Jahre alt. Die jüngste VP war 19, die älteste 46 Jahre alt.

Die Körpergröße lag im Durchschnitt bei 1,72 m. Dabei war die kleinste VP 1,60 m, die größte 1,90 m groß.

11 der 20 VP hatten vorherige Erfahrung mit VR.

Insgesamt gaben 11 an, ein geschultes Gehör zu besitzen, davon 4 durch ein Studium oder Beruf mit Schwerpunkt Ton; weitere 5 hatten es durch eine Gesangs- oder Instrumentalausbildung geschult und 2 durch eine Tätigkeit beim Hochschulradio.

Zwei der VP gaben an, an einem leichten Hörschaden in Form eines Tinnitus zu leiden. Wegen geringer Ausgeprägtheit wurden sie dennoch zum Hörversuch zugelassen. (Die bereits erwähnte Person mit schwerem Hörschaden auf einem Ohr wurde von vornerein nicht zugelassen, hat deshalb keinen Fragebogen ausgefüllt und kommt somit in dieser Statistik nicht vor.)

## 8.3 Durchführung

Die Erhebung wurde an drei konsekutiven Tagen durchgeführt. Nach der Begrüßung sollten die VP den Fragebogen ausfüllen und die Anleitung lesen. Anschließend hat der Versuchsleiter nach Unklarheiten gefragt und ggf. beantwortet.

Wurde auf dem Fragebogen keine oder wenig Erfahrung mit VR angegeben, wurde Wert darauf gelegt, das HMD und die Controller zu erklären, auf die Stolpergefahr durch den Kabelstrang des HMDs hinzuweisen, und die Einblendung der Spielfeldgrenzen zu erklären, damit die Probanden nicht gegen reale Wände stoßen.

#### 8.3.1 Kalibrierung & Einhörphase in VR

Im VR-Labor wurde die VP gebeten, normal aufrecht zu stehen. Das diente der Kalibrierung des Lautsprechersystems auf die Körpergröße. Dazu hatte das Erhebungstool eine Funktion, bei der auf Knopfdruck die Z-Koordinate des HMDs ausgelesen wird und das Lautsprechersystem einmalig auf diese Höhe bewegt wird.

Anschließend wurde der VP knapp eine Minute Musik aus der Vorne-Position vorgespielt. Dies sollte ihr Gelegenheit geben, sich in den virtuellen Raum einzuhören. Vorher wurde sie dazu aufgefordert, während der Musik den Basketball zu finden und damit zu spielen. Da sie sich so viel bewegte, würde sie die Musik aus vielen unterschiedlichen Richtungen zu hören kriegen; außerdem konnte sie sich so mit dem Raum und Bewegungen in VR im Allgemeinen vertraut machen.

#### 8.3.2 Erhebung in VR

Die eigentliche Erhebung lief wie folgt ab:

- Die VP wurde aufgefordert, sich auf die Standpunktmarkierung zu stellen und ihren Kopf nach vorne auszurichten.
- Waren beide Indikatoren grün geworden, erschien auf dem Blickrichtungsindikator der Text: "Um Klang abzuspielen: Trigger (Zeigefinger) drücken."
- Dieser Schritt folgte nur, wenn die VP den Trigger drückte und beide Indikatoren nach wie vor grün waren:
  - Der Text verschwand; es verging noch eine halbe Sekunde, dann spielte der Stimulus ab. Wurde nach Beginn des Stimulus einer der Indikatoren wieder Rot, wurde der Stimulus dennoch nicht unterbrochen.

Somit wurde über Übertretungen der Grenzen des Spielraums hinweggesehen; die VP hatten jedoch stets direktes visuelles Feedback darüber.

- War der Stimulus zu Ende, verschwanden die Indikatoren. Die VP durften ihren Kopf wieder in alle Richtungen frei bewegen, und sollten ihre Eingabe machen.
- Für die Eingabe drehten die VP ihren Kopf so, dass der Cursor in die Richtung zeigte, in der sie den Klang gehört hatten, und drückten den Trigger.
- Anschließend drehten sie sich wieder nach vorne, wo die Indikatoren wieder zu sehen waren, und das Prozedere erneut begann.

Waren alle Stimuli absolviert, durfte die VP die VR-Ausrüstung ablegen.

#### 8.3.3 Aussagen der VP während der Erhebung

Manche VP machten während dem laufenden Hörtest Aussagen oder stellten Fragen:

- Zwei VP sagten aus, manche Klänge hätten eine Ausdehnung. Sie wurden gebeten, in diesem Fall die Mitte der Ausdehnung als Richtung einzugeben.
- Eine VP fand, dass Kopfbewegungen die Richtungslokalisation schwerer statt leichter machten, und fragte, ob sie den Kopf stillhalten dürfe. Sie wurde aufgefordert, es so zu machen, wie es sich für sie gut anfühlte.

#### 8.3.4 Aussagen der VP am Ende

Zum Abschluss wurde die VP noch gefragt, ob sie Anmerkungen machen wollten.

Darauf äußerten einige ihre Beobachtung, dass manche Klänge Assoziationen mit bestimmten Schallquellen (z.B. Hubschrauber) hervorriefen. Und dass sie Gedanken gehabt hätten, ob sie das Hörereignis z.B. oben lokalisierten, weil es für sie "wirklich von da" kam, oder ob sie sich die Oben-Richtung nur "einbildeten", weil sich Hubschrauber normalerweise eher über einem befinden.

In ähnlicher Weise äußerten sich 6 VP wie folgt:

- 6 VP bezeichneten Stimuli als Hubschrauber und dass diese eher oben zu verorten seien.
- Eine VP assoziierten mit einem Flugzeug und deshalb mit der Oben-Richtung.
- Eine VP assoziierte ebenfalls mit einem Flugzeug, aber einem, das bereits gelandet sei, und deshalb eher der Seitenrichtung zuzuordnen sei.
- Eine VP nannte einen Staubsauger, den sie deshalb eher am Boden vermutete
- Eine VP hörte Kirchturmglocken, die typischerweise eher von oben kämen.
- Eine VP sagte, "dunkle Sounds" würde er eher unten vermuten. Auf Nachfrage wurde das Attribut "dunkel" mit "arm an hohen Frequenzanteilen" spezifiziert.
- Eine VP beschrieb ein "röhrendes Geräusch wie von einem Lüftungsschacht", das sie eher im "Mittelpunkt der Wände oder der Decke, aber nicht in den Ecken oder am Boden" vermutete.

Weiterhin sagten zwei VP aus, dass sich HE stärker bewegten oder sprängen, wenn sie den Kopf stärker bewegten.

Eine VP merkte an, senkrecht nach oben zu gucken sei sehr mühsam.

## 8.4 Datenverarbeitung

Folgende Daten wurden aufgezeichnet und in Microsoft Excel verarbeitet:

- Nummer der Versuchsperson
- Stimulus (Klang A, C bis I)
- Vektor der Schallereignisrichtung (SER)
- Eingegebener Vektor der Hörereignisrichtung (HER)
- Eingabe "Ich weiß es nicht", "Im Kopf", "Mehrere Richtungen", "Bewegt sich"

Ob die VP die Grenzen ihrer Bewegung (Blickwinkel, Standpunkt) eingehalten hat

#### 8.4.1 Gültigkeit der Daten

Insgesamt wurden den 20 VP jeweils 8 Stimuli aus 10 Richtungen vorgespielt. Daraus resultierten 1600 Datensätze.

Davon wurden 192 Datensätze als ungültig markiert und von der weiteren Berücksichtigung ausgeschlossen, weil sich die VP während des Abspielens des Stimulus nicht mehr innerhalb der vorgegebenen Grenzen befanden, indem sie sich zu weit weg von der Standpunktmarkierung bewegt hatten und/oder ihre Blickrichtung vom Blickwinkelindikator abschweifen ließen.

#### 8.4.2 Lokalisationsfehler

Für jede Eingabe wurde jeweils der Vektor der HER, also der Eingabe der VP, und der SER aufgezeichnet. Der Winkel zwischen den beiden Vektoren wurde berechnet, er stellt den Lokalisationsfehler dar.

272 Datensätze waren jedoch ohne Vektoren, da die Optionen "Ich weiß nicht" oder IKL eingegeben wurden, die keine Richtungsinformation zuließen.

#### 8.4.3 Vertauschungen

Es sollte die Häufigkeit der Vorne-Hinten- und Oben-Unten-Vertauschungen pro Klang festgestellt werden.

Wenn diese Phänomene auftreten, liegt das Hörereignis in einer anderen Hemisphäre als das SE.

Nicht bei allen Schallereignispositionen ist eine Vertauschung möglich: Liegt sie z.B. auf der Horizontalebene, kann nicht von einer Oben-Unten-Vertauschung gesprochen werden.

Als Positionen kommen für die Vorne-Hinten-Vertauschung 8 Positionen<sup>117</sup> in Frage, und 5 Positionen<sup>118</sup> für die Oben-Unten-Vertauschung. Die Datensätze mit SE in diesen Positionen wurden als Vertauschungskandidaten bezeichnet.

Lag ein Hörereignis zwar in einer anderen Hemisphäre als das SE, aber sehr nahe an der die Hemisphären trennenden Ebene, wurde dies als noch auf der Ebene liegend betrachtet und nicht als Vertauschung.

Demnach wurde die Vertauschung für die Datenverarbeitung wie folgt definiert:

- Das SE befindet sich in der einen, das Hörereignis in der gegenüberliegenden Hemisphäre.
- Das SE befindet sich nicht auf der Ebene, die diese Hemisphären voneinander trennt.
- Der Winkel zwischen dieser Ebene und dem Hörereignis beträgt mindestens 22,5°.

Diesen Kriterien nach wurde die Zahl der Vertauschungen für jeden Stimulus gezählt. Dabei wurde noch einmal unterschieden, in welche Richtung die Vertauschung stattfand: Es wurde also getrennt gezählt für Vertauschungen

- von vorne nach hinten
- von hinten nach vorne
- von oben nach unten
- von unten nach oben

Es wurde auch der prozentuale Anteil der Vertauschungen an Vertauschungskandidaten berechnet.

\_

<sup>&</sup>lt;sup>117</sup> Vorne-Unten, Vorne, Vorne-Oben, Hinten-Oben, Hinten, Hinten-Unten, Vorne-Seitlich und Hinten-Seitlich.

<sup>&</sup>lt;sup>118</sup> Vorne-Unten, Vorne-Oben, Oben, Hinten-Oben, Hinten-Unten.

#### 8.4.4 Hörereignisse mit mehrdeutiger oder ohne Richtung

Es wurden für jeden Klang folgende Angaben gezählt:

- Diffuslokalisation (Option "Ich weiß nicht")
- IKI
- Bewegung (Option: Hörereignis "hat sich bewegt")
- Spaltung (Option: Hörereignis liegt "in mehreren Richtungen")

## 8.5 Ergebnisse & Interpretation

Die jeweiligen Ergebnisse für die bearbeiteten Klänge lassen sich nur begrenzt miteinander zwecks Interpretation vergleichen, da bei komplexen Signalen immer mehrere Parameter gleichzeitig geändert werden; man spricht dabei von Konfundierung.

Dennoch lässt sich jeder Klang jeweils mit dem Ausgangsklang vergleichen, und damit die zentrale Fragestellung der Erhebung beantworten: Inwiefern die Bearbeitungen der Klänge die Richtungslokalisation beeinflusst haben.

#### 8.5.1 Lokalisationsfehler

Abbildung 1 zeigt die Abweichungen der HER von der SER als jeweils für die Klänge getrennte Boxplots. Die Werte umfassen dabei fast den gesamten Wertebereich, bis hinauf zum höchstmöglichen Wert von 180°, wenn also das Hörereignis dem SE genau gegenüber liegt – dies war zu erwarten, da das Problem der Vertauschungen in extremen Winkeln der Fehllokalisation resultiert.

Weiterhin ist auffällig, dass die Richtungslokalisation für alle Klänge insgesamt lediglich passabel funktioniert, wie sich an den Medianen zwischen 46° und 62° ablesen lässt.

#### 8.5.1.1 Klang A

Die Basis aller weiteren Bearbeitungen, Klang A, zeigt dabei das schlechteste Ergebnis. Die Hälfte der Lokalisationsfehler befindet sich zwischen 44° und 92°, mit dem Median bei 62°. Betrachtet man dies im Zusammenhang mit dem Szenario, in dem ein Fehler bis 30° als Erfolg und bis 55° als Teilerfolg gewertet wird, ist dies ein unbefriedigendes Ergebnis: Gerade mal 16% der Lokalisationen sind mit weniger als 30° Abweichung erfolgt, und nur 45% mit weniger als 55° (Abbildung 2). Bei mehr als der Hälfte der Versuche würde sich der Spieler also der falschen Richtung zuwenden, in der er keinen Gegner in seinem Sichtfeld vorfände.

#### 8.5.1.2 Bearbeitete Klänge (C bis I)

Der Median von Klang F hat sich mit 62° gegenüber dem von Klang A gar nicht verbessert. Lediglich die Quartile sind leicht besser, und der Anteil von Lokalisationsfehlern unter 30° liegt bei etwas besseren 20%. Der Anteil der Fehler unter 55° hat sich mit 43% sogar etwas verschlechtert. Insgesamt sind diese Unterschiede aber gering, womit Klang F kaum eine nennenswerte Verbesserung darstellt.

Die Klänge C bis E und G bis I weisen relativ ähnliche Ergebnisse auf: Die Lokalisation hat sich bei ihnen im Vergleich zu Klang A verbessert, mit Medianen von 46° bis 53°, immerhin eine Verbesserung um 9 bis 16°.

Das schlägt sich auch in Hinblick auf das Szenario nieder: Alle Klänge steigern den Anteil der Lokalisationen mit Fehlern unterhalb von 30°, mit Ausnahme von Klang I. Gleiches gilt für die Lokalisationen, bei denen der Fehler höchstens 55° betrug, nur hier mit der Ausnahme von Klang F.

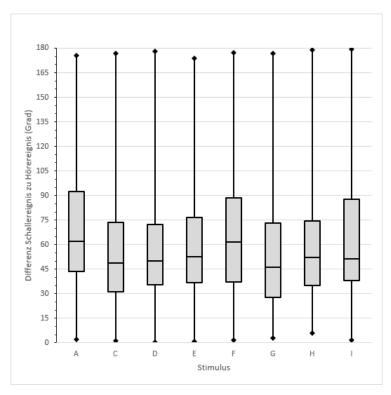
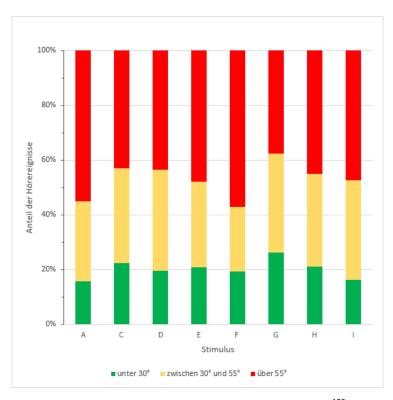


Abbildung 18: Fehllokalisationen<sup>119</sup>
Der Boxplot zeigt die Quartile 1 und 3 als Enden des grauen Kastens, und den Median als waagerechter Strich im Kasten. Die schwarzen Rauten am Ende der Linien zeigen das Minimum und Maximum.



 $Abbildung\ 19: Szenario-relevante\ Fehllokalisationen^{120}$ 

<sup>120</sup> Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel

 $<sup>^{119}</sup>$  Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel

#### 8.5.2 Vertauschungen

Die Vertauschung stellt ein ungünstiges Phänomen dar: Tritt sie auf, ist der Lokalisationsfehler stets erheblich; im besten Fall 67,5°, im schlechtesten 180°. In allen Fällen wäre er tief im roten Bereich von Abbildung 2, und würde für das Szenario als gescheitert gelten.

#### 8.5.2.1 Vorne-Hinten-Vertauschungen

Betrachtet man die Ergebnisse für von vorne nach hinten und von hinten nach vorne in Abbildung 3 zusammen, hat sich, gegenüber Klang A, die Situation bei den meisten Klängen verbessert.

Lediglich Klang F und I schneiden ähnlich schlecht ab wie Klang A. Sie bestehen aus weitgehend demselben Klangmaterial mit vielen Spitzen und Kerben, nur dass Klang I um ein Rauschsignal erweitert und moduliert wurde. Die Modulation sollte bei dem rein spektralen Vorgang der Vorne-Hinten-Unterscheidung keinen Unterschied machen.

Bei Klang A, das fast nur Inhalte in einem Vorwärtsband besitzt, wäre zu vermuten gewesen, dass deutlich mehr Vertauschungen von hinten nach vorne stattfinden; jedenfalls sind sie hier noch am häufigsten, mit 21%. Bei Klang F betragen sie 18% und bei Klang I noch 10%.

Dass die Vertauschung von vorne nach hinten bei Klang F und I so besonders ausgeprägt ist, lässt sich vermutlich auf die Spitzen und Kerben zurückführen, die diese - im Gegensatz zu Klang A – über weitere Teile des Spektrums zeigen.

Auffällig ist ansonsten die gute Performance von Klang C, wobei es schwerfällt, einen Grund dafür auszumachen.

#### 8.5.2.2 Oben-Unten-Vertauschungen

Die Vertauschung von Oben nach Unten kam praktisch kaum vor, mit Werten von 0 bis 7,5%. Umso mehr umgekehrt: Die Lokalisation oben, wenn das SE unten war. Von dem Trio, das schon bei der Vorne-Hinten-Vertauschung schlecht abschnitt, tun sich auch hier wieder die Klänge A mit 46% und ganz besonders I mit 59% hervor. Klang F dagegen verhält sich, genau wie Klänge C und H, noch relativ moderat.

Es sei noch auf die Aussagen mancher VP verwiesen, die meinten, Klänge, die wie Hubschrauber (Klang G) oder Flugzeuge (Klang E) klingen, würden bevorzugt am Himmel lokalisiert: Die Daten lassen auf einen Zusammenhang nicht deuten, für diese Klänge ist die Vertauschung von unten nach oben besonders niedrig.

Insgesamt stellen die Ergebnisse bei den Vertauschungen eine deutliche Verbesserung dar für alle Klänge außer I und (teilweise) F.

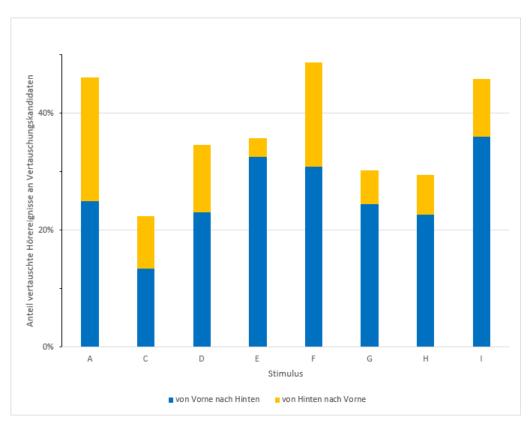


Abbildung 20: Vorne-Hinten-Vertauschungen<sup>121</sup>

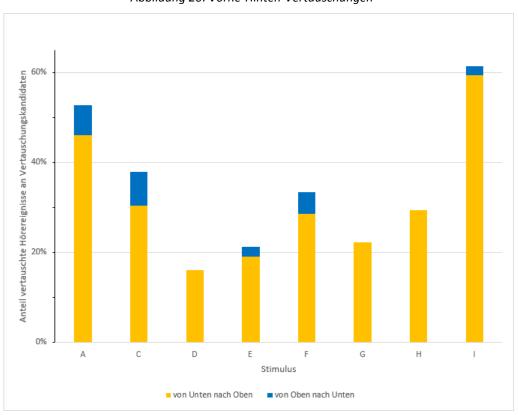


Abbildung 21: Oben-Unten-Vertauschungen 122

54

 $<sup>^{121}</sup>$  Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel  $^{122}$  Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel

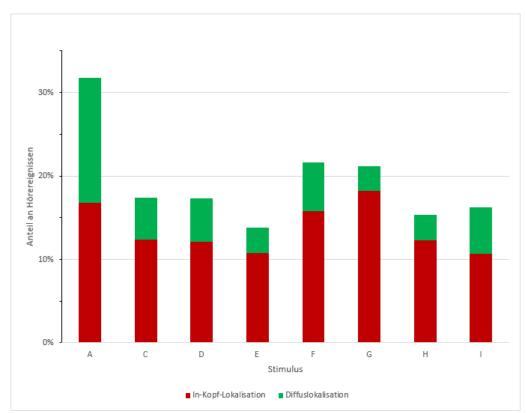


Abbildung 22:Hörereignisse ohne Richtung 123

#### 8.5.3 Ohne Richtung

Diffuslokalisation tritt ein, wenn die VP gar keine Richtungsinformation auswerten konnte. Sie ist, wie man Abbildung 1 entnehmen kann, bei Klang A mit 15% recht ausgeprägt, während hier sämtliche Klangbearbeitungen zu deutlichen Verbesserungen führten, mit Werten von 1,6% bis 5,6%. Da Klang A mit seinem begrenzten Frequenzgang und Dynamik am wenigsten Information beinhaltet, überrascht nicht, dass die VP bisweilen ratlos wahren. Die bearbeiteten Klänge unterscheiden sich stark, ihre Ergebnisse jedoch wenig. Eines haben sie aber alle gemeinsam: Dass sie ein deutlich breiteres Spektrum haben als Klang A. Daraus könnte man den Schluss ziehen, dass es schon genügt, dass ein Klang breitbandig ist, um einen guten Teil der Diffuslokalisationen zu verringern.

Die IKL dagegen liegt bei Klang A bei 17%. Bei ähnlichen Werten liegen Klang F (16%) und Klang G (18%). Die übrigen Klänge haben sich auf Werte zwischen 11 und 13% verbessert. IKL wird verursacht durch widersprüchliche Cues oder mangelnde Distanzinformation. Letztere war über die Erhebung konstant, da die Raumparameter und die Distanz der Schallereignisse zur VP konstant waren.

Zumindest bei Klang F zeigt das Spektrum (Abbildung 6) einen möglichen Grund für widersprüchliche Cues: Die bereits als problematisch besprochenen Spitzen und Kerben und die Koinzidenzen mit den Richtungsbestimmenden Bändern. Auch die schlechten Werte von Klang F bei der Vorne-Hinten-Vertauschung sprechen dafür, dass die VP hier einfach widersprüchliche Cues wahrnahm, sie sich mal in IKL oder in Vertauschung ausdrücken.

\_

<sup>&</sup>lt;sup>123</sup> Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel

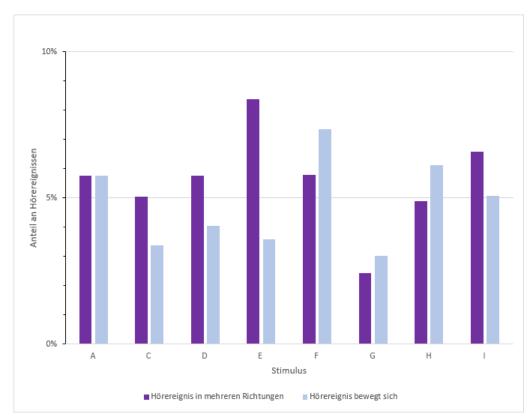


Abbildung 23: Hörereignisse mit mehrdeutiger Richtung<sup>124</sup>

#### 8.5.4 Mehrdeutige Richtungen

Die Werte für die Angabe, ein Hörereignis sei in zwei Richtungen gespalten, sind relativ konstant um etwa 5%. Lediglich Klang E schneidet mit 8,4% schlechter ab, und Klang G mit 2,4% besser. Die Häufigkeit der Aussage, dass sich die HER während des Abspielens bewegte, schwankte von 3% bis 8,6%, wobei Klang G am besten und Klang F am schlechtesten abschnitt.

Die Phänomene der gespaltenen oder sich bewegenden HE waren ursprünglich nicht Teil der Erkenntnisgewinnung der vorliegenden Arbeit. Als sie beim Vortest von den VP bemerkt wurden, wurden sie der Vollständigkeit halber in die Erhebung aufgenommen, und um eine Übersicht über die Größe des Problems zu bieten. Ohne theoretischen Bezug in vorherigen Teilen der Arbeit ist es kaum möglich, die Ergebnisse zu interpretieren.

Auch ob die Eigenschaften der Klänge oder z.B. der verwendete Binauralsynthese-Algorithmus Auslöser für die Phänomene sind, ließe sich höchstens mutmaßen. Es lässt sich jedoch sagen, dass es zwar, verglichen mit den anderen Ergebnissen, um ein seltener vorkommendes Problem handelt, es aber auch, wenn es auftrat, besonders irritierend auf die VP wirkte und das VR-Erlebnis maßgeblich störte. Daher wäre weitere Forschung in dieser Hinsicht wünschenswert.

#### 8.6 Fazit

Zu einem großen Teil waren die Klangbearbeitungen in Abschnitt 7 erfolgreich: Die bearbeiteten Klänge wurden in der Regel genauer lokalisiert.

<sup>&</sup>lt;sup>124</sup> Eigene Grafik, angelehnt an Screenshot aus Microsoft Excel

Ausnahmen waren hier die Klänge F und I, die sich kaum verbesserten oder gar schlechter abschnitten. Ein möglicher Grund wäre, dass beide hohe tonale Anteile und damit von Spitzen und Kerben dominierte Frequenzgänge besaßen.

Auch wurden die bearbeiteten Klänge seltener vertauscht, wieder mit Ausnahmen F und I.

Der Fall, dass VP ganz ohne Richtungseindruck waren, kam bei allen bearbeiteten Klängen deutlich seltener vor als beim Ausgangsklang. Es besteht die Vermutung, dass die Bandbreite des Spektrums dabei ausschlaggebend war.

Die IKL zeigt ein wenig eindeutiges Bild: Einige Klänge schnitten besser ab, andere nicht. Ein Zusammenhang mit widersprüchlichen Cues könnte man höchstens bei Klang F beobachten.

Um die teilweise problematischen Ergebnissen bei den Klängen F und I zu erklären, kann man die beiden Gemeinsamkeiten betrachten: Beide sind durch Pitch-Layering bearbeitet worden und beide haben einen hohen tonalen Anteil. Letzteres liegt daran, dass der Ausgangsklang selbst einen hohen tonalen Anteil hat, und das Pitch-Layering die einzige der verwendeten Methoden ist, die nicht den Rauschanteil stark erhöht. Insofern kann Pitch-Layering mit anderem Material erfolgreicher sein; es liegt nahe, das die schlechten Ergebnisse eher am kritischen Ausgangsmaterial als an der Bearbeitungsmethode liegen.

Das Ziel, das im Szenario als erfolgreiche Lokalisation definiert wurde, wurde in den meisten Fällen auch häufiger erreicht. Trotz der Verbesserung befindet sich die Quote auch weiterhin einem inakzeptablen Bereich: Etwa die Hälfte der Klänge wurden nicht erfolgreich lokalisiert, was für ein VR-Game, dessen Spielmechanik auf die Lokalisation angewiesen ist, ein schlechtes Ergebnis darstellt.

Die Aussagen der VP während der Erhebung, die semantische Bedeutung der Klänge könnte die Lokalisation beeinflussen, hat sich nicht bestätigt. Die von den VP berichteten Phänomene der gespaltenen oder bewegenden Hörereignisse erscheinen im Kontext von VR-Games interessant und bedürfen weiterer Forschung.

## 9 Zusammenfassung

Die vorliegende Arbeit stellte die Frage, was Sound Designer beachten können, damit ihre Klänge in VR-Games besser lokalisiert werden.

Zunächst wurden Grundlagen zum räumlichen Hören und zur Binauralsynthese vorgestellt. Ein Szenario, bei dem die vorgestellten Wirkmechanismen der VR-Games wie Presence und Interaktion besonders vom Ton beeinflusst werden, wurde skizziert, und als Vorbild für ein exemplarisches Sound Design genutzt.

Die Methoden der Hüllkurvenbearbeitung waren erfolgreich, wie auch die der nichtlinearen Verzerrung und des Layering. Lediglich die Signale, die einen hohen tonalen Anteil vorwiesen, bereiteten erhebliche Probleme.

Auch zeigten die Klangbearbeitungen teilweise die Schwierigkeit, zwischen dem Ziel der Lokalisationsverbesserung und gestalterischen Aspekten abzuwägen. In Fällen, in denen die beiden einander ausschließen, dürfte die Gestaltung in der Praxis Vorrang haben. Gerade weil sie jedoch stark davon abhängig ist, welches Material verwendet und welches VR-Game vertont wird, kann der besprochene Sounddesign-Prozess keine Blaupause sein. Stattdessen sollte er Möglichkeiten und Schwierigkeiten aufzeigen, und ein Bewusstsein für die Macht des Sound Designers über die Lokalisation schaffen.

Die Erhebung hat gezeigt, dass die Klangbearbeitung in ihren Zielen zu großen Teilen erfolgreich war.

## Literaturverzeichnis

Algazi, V.; Duda, Richard (2011): Headphone-Based Spatial Sound. In: *IEEE Signal Process. Mag.* 28 (1), S. 33–42. DOI: 10.1109/MSP.2010.938756.

Anish Chandak (2014): 3D Audio for Virtual Reality. East Coast Games Conference 2014. Online verfügbar unter https://www.impulsonic.com.

Asano, Futoshi; Suzuki, Yoiti; Sone, Toshio (1990): Role of spectral cues in median plane localization. In: *The Journal of the Acoustical Society of America* 88 (1), S. 159–168. DOI: 10.1121/1.399963.

Bech, Søren; Zacharov, Nick (2006): Perceptual audio evaluation. Theory, method and application. Chichester, England, Hoboken, NJ: John Wiley & Sons. Online verfügbar unter http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10300625.

Blauert, Jens (1969): Untersuchungen zum Richtungshören in der Medianebene bei fixiertem Kopf.

Blauert, Jens (1974): Räumliches Hören. - [Hauptbd.]. Stuttgart: Hirzel.

Blauert, Jens; Braasch, Jonas (2008): Räumliches Hören. In: Stefan Weinzierl (Hg.): Handbuch der Audiotechnik. 1. Aufl. s.l.: Springer-Verlag (VDI-Buch), S. 87–122.

Brimijoin, W. Owen; Boyd, Alan W.; Akeroyd, Michael A. (2013): The Contribution of Head Movement to the Externalization and Internalization of Sounds. In: *PLoS ONE* 8 (12). Online verfügbar unter http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083068, zuletzt geprüft am 30.01.2017.

Brungart, Douglas; Kordik, Alex J.; Simpson, Brian D. (2006): Effects of Headtracker Latency in Virtual Audio Displays. In: *J. Audio Eng. Soc* 54 (1/2), S. 32–44. Online verfügbar unter http://www.aes.org/e-lib/browse.cfm?elib=13665.

Carlile, Simon; Pralong, Danièle (1994): The location-dependent nature of perceptually salient features of the human head-related transfer functions. In: *The Journal of the Acoustical Society of America* 95 (6), S. 3445–3459. DOI: 10.1121/1.409965.

Coleman, Paul D. (1962): Failure to Localize the Source Distance of an Unfamiliar Sound. In: *The Journal of the Acoustical Society of America* 34 (3), S. 345–346. DOI: 10.1121/1.1928121.

Dickreiter, Michael (Hrsg.); Dittel, Volker (Hrsg.); Hoeg, Wolfgang (Hrsg.); Wöhr, Martin (Hrsg.) (2014): Handbuch der Tonstudiotechnik. 8. Aufl. Berlin [u.a.]: de Gruyter.

Dietz, Mathias; Marquardt, Torsten; Salminen, Nelli H.; McAlpine, David (2013): Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (37), S. 15151–15156. DOI: 10.1073/pnas.1309712110.

Dörner, Ralf; Broll, Wolfgang; Grimm, Paul; Jung, Bernhard (2013): Virtual und Augmented Reality (VR / AR). Berlin, Heidelberg: Springer Berlin Heidelberg.

EBU Europäische Rundfunkunion (Hg.) (2014): R128. Loudness Normalisation and Permitted Maximum Level of Audio Signals. Online verfügbar unter https://tech.ebu.ch/docs/r/r128.pdf, zuletzt geprüft am 26.03.2017.

Gilkey, Robert H.; Good, Michael D.; Ericson, Mark A.; Brinkman, John; Stewart, John M. (1995): A pointing technique for rapidly collecting localization responses in auditory research. In: *Behavior Research Methods, Instruments, & Computers* 27 (1), S. 1–11. DOI: 10.3758/BF03203614.

Han, H. L. (Hg.) (1992): On the Relation Between Directional Bands and Head Movements. 92nd AES Convention. Wien, 24.03.1992. Delft University of Technology, Netherlands. Online verfügbar unter http://www.aes.org/e-lib/browse.cfm?elib=6840.

Klein-Hennig, Martin; Dietz, Mathias; Hohmann, Volker; Ewert, Stephan D. (2011): The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. In: *The Journal of the Acoustical Society of America* 129 (6), S. 3856–3872. DOI: 10.1121/1.3585847.

Klimmt, Christoph (2010): Computerspielen als Handlung. Dimensionen und Determinanten des Erlebens interaktiver Unterhaltungsangebote. Zugl.: Hannover, Hochsch. für Musik und Theater, Diss., 2004. Unveränd. Nachdr. Köln: von Halem (Unterhaltungsforschung, 2).

Kreylos, Oliver (2016): Lighthouse tracking examined. Internet-Blog "Doc-Ok.org". Online verfügbar unter http://doc-ok.org/?p=1478, zuletzt geprüft am 01.03.2017.

Macpherson, Ewan A.; Middlebrooks, John C. (2003): Vertical-plane sound localization probed with ripple-spectrum noise. In: *The Journal of the Acoustical Society of America* 114 (1), S. 430–445. DOI: 10.1121/1.1582174.

Makous, James C.; Middlebrooks, John C. (1990): Two-dimensional sound localization by human listeners. In: *The Journal of the Acoustical Society of America* 87 (5), S. 2188–2200. DOI: 10.1121/1.399186.

Møller, Henrik; Sørensen, Michael Friis; Jensen, Clemen Boje; Hammershøi, Dorte (1996): Binaural Technique: Do We Need Individual Recordings? In: *J. Audio Eng. Soc* 44 (6), S. 451–469. Online verfügbar unter http://www.aes.org/e-lib/browse.cfm?elib=7897.

Nicol, Rozenn (2010): Binaural technology. New York: Audio Engineering Society (AES monograph).

Oliver Kreylos (2016): Optical Properties of Current VR HMDs. Online verfügbar unter http://doc-ok.org/?p=1414, zuletzt geprüft am 02.02.2017.

Randomness and Integrity Services Ltd. (2017): Random.org. True Random Number Service. Online verfügbar unter https://www.random.org/integer-sets/, zuletzt geprüft am 10.02.2017.

Roffler, Suzanne K.; Butler, Robert A. (1968): Factors That Influence the Localization of Sound in the Vertical Plane. In: *The Journal of the Acoustical Society of America* 43 (6), S. 1255–1259. DOI: 10.1121/1.1910976.

Santala, Olli; Takanen, Marko; Pulkki, Ville (Hg.) (2013): Effect of Target Signal Envelope on Direction Discrimination in Spatially Complex Sound Scenarios. Convention Paper 8820. 134th AES Convention. Rome, Italy. Online verfügbar unter http://www.aes.org/e-lib/browse.cfm?elib=16721.

Sean Buckley (2015): This Is How Valve's Amazing Lighthouse Tracking Technology Works. Hg. v. Gizmodo.com. Online verfügbar unter http://gizmodo.com/this-is-how-valve-s-amazing-lighthouse-tracking-technol-1705356768, zuletzt aktualisiert am 19.05.2015, zuletzt geprüft am 15.02.2017.

Sherman, William R.; Craig, Alan B. (2003): Understanding virtual reality. Interface, application, and design. San Francisco, CA: Morgan Kaufmann (Morgan Kaufmann series in computer graphics and geometric modeling). Online verfügbar unter

http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=249304.

Statista GmbH Hamburg (Hg.) (2016): Virtual Reality. Statista Dossier. Online verfügbar unter http://de.statista.com/download/MTQ2ODc4NzY3OSMjNjQ2NDAjlzI5MzQ0IyMxIyNwZGYjI1N0dWR5 , zuletzt geprüft am 17.07.2016.

Stitt, Peter; Hendrickx, Etienne; Messonnier, Jean-Christophe; Katz, Brian (Hg.) (2016): The Influence of Head Tracking Latency on Binaural Rendering in Simple and Complex Sound Scenes.

Thaner, Mathias (1999): Einfluß von Kopfbewegungen auf die Lokalisation in der Medianebene. Diplomarbeit. Fachhochschule München, München. Institut für Rundfunktechnik.

Theile, Günther (1981): Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer. In: *Rundfunktechnische Mitteilungen* 25 (4), S. 155–170.

VRNerds Blog (2016): VR Brillen Vergleich. Hg. v. VRNerds.de. Online verfügbar unter http://www.vrnerds.de/vr-brillen-vergleich/, zuletzt geprüft am 12.02.2017.

Wenzel, Elizabeth M. u.a. (1993): Localization using nonindividualized head-related transfer functions. In: *Journal of the Acoustical Society of America* 94 (1), S. 111–123.

Wightman, Frederic L.; Kistler, Doris J. (1989): Headphone simulation of free-field listening. I: Stimulus synthesis. In: *Journal of the Acoustical Society of America* 85 (2), S. 866.

Wikipedia (2017): HTC Vive. Hg. v. Die freie Enzyklopädie Wikipedia. Online verfügbar unter https://de.wikipedia.org/w/index.php?title=HTC\_Vive&oldid=161319262, zuletzt aktualisiert am 05.01.2017, zuletzt geprüft am 07.02.2017.

Witmer, Bob G.; Singer, Michael J. (1998): Measuring Presence in Virtual Environments. A Presence Questionnaire. In: *Presence: Teleoperators and Virtual Environments* 7 (3), S. 225–240. DOI: 10.1162/105474698565686.

Xie, B.; Ning Xiang, R.P.I.; Blauert, J. (2013): Head-Related Transfer Function and Virtual Auditory Display: Second Edition: J. Ross Publishing, Incorporated. Online verfügbar unter https://books.google.de/books?id=fvDLCgAAQBAJ.

# Anhang / Datenträger

An dieser Stelle ist eine CD-ROM beigelegt.