

# Masterarbeit

Jan Strothmann

Distanzmodellierung für 3D-Audio zum Einsatz in  
binauralem Echtzeit-VR

Jan Strothmann

# Distanzmodellierung für 3D-Audio zum Einsatz in binauralem Echtzeit-VR

Masterarbeit eingereicht im Rahmen der Masterprüfung  
im Studiengang Master of Science Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Wolfgang Fohl  
Zweitgutachter: Prof. Dr. Andreas Meisel

Eingereicht am: 5. Oktober 2021

**Jan Strothmann**

**Thema der Arbeit**

Distanzmodellierung für 3D-Audio zum Einsatz in binauralem Echtzeit-VR

**Stichworte**

3D-Audio, Benutzerstudie, Distanzmodellierung, Spiegelschallquellen, Binaurale HRTF-Filterung, VR, Echtzeit, MATLAB, C, R

**Kurzzusammenfassung**

Ziel dieser Arbeit ist die Untersuchung einer echtzeitfähigen Distanzmodellierung für 3D-Audio. Das Einsatzgebiet ist der Bereich Virtual Reality, deshalb wird ein binauraler, kopfhörerbasierter Ansatz verwendet.

Das Modell ist vom Entwurf her echtzeitfähig und wird zuerst in MATLAB- und C-Code realisiert, um in einer Studie analysiert zu werden. Die Erkenntnisse aus der Studie können dann für eine VR-Simulation eingesetzt werden.

**Jan Strothmann**

**Title of Thesis**

Distancemodelling for realtime VR-3D-Audio using binaural HRTFs

**Keywords**

3D-Audio, User-Study, Distancemodelling, Mirror Sound Sources, Binaural HRTF-Filtering, VR, Realtime, MATLAB, C, R

**Abstract**

In this thesis a realtime distance modelling for 3D-audio is explored.

The model is meant to be used in virtual reality and for this reason a binaural, hedphone-based aprooch is used. The created model is designed to be capable of realtime but is first realised in MATLAB- and C-Code to use it in a study.

The results from the study then can be used for the improval of VR-Simulations.

# Inhaltsverzeichnis

Abbildungsverzeichnis	vi
Tabellenverzeichnis	ix
Abkürzungen	xi
<b>1 Einleitung</b>	<b>1</b>
<b>2 Auditive Wahrnehmung der Richtung und Distanz von Tonquellen</b>	<b>2</b>
2.1 Grundlegende Eigenschaften des menschlichen Gehörs . . . . .	2
2.1.1 Funktionsweise des menschlichen Richtungshörens . . . . .	4
2.1.2 Ausgewählte Aspekte und Eigenheiten des Entfernungshörens . . .	5
2.1.3 Raumschall als zentraler Aspekt des Entfernungshörens . . . . .	8
<b>3 Erstellung von 3D-Audio mit auditiver Distanz für VR-Systeme</b>	<b>9</b>
3.1 Eigenschaften und Besonderheiten von 3D-Audio in VR-Systemen . . . . .	9
3.1.1 Zu erfüllende Rahmenbedingungen für Echtzeit-VR . . . . .	11
3.2 Grundlagen der filterbasierten Audiotbearbeitung . . . . .	12
3.2.1 Systemtheorie und Faltung als Basis der Filterung . . . . .	12
3.2.2 Praktische Filtererzeugung durch Messung der Systemantwort . . .	14
3.3 Vor- und Nachteile ausgewählter Verfahren zur 3D-Audio-Erzeugung . . .	16
3.3.1 HRTF-Filterung für die Verwendung in binauralen Systemen . . .	16
3.3.2 Alternative und ergänzende Techniken im Bereich 3D-Audio . . . .	17
3.4 Künstliche Erzeugung von Raumschall für die Distanzmodellierung . . . .	22
3.4.1 Erzeugung von Raumschall durch Verwendung von Faltungshall . .	23
3.4.2 Das Spiegelschallquellenmodell als modellbasierter Ansatz . . . . .	23
<b>4 Zur Zeit verwendete Techniken im Bereich binaurales VR-Audio</b>	<b>26</b>
4.1 Anwendungsbeispiel: Immersive 360°-Videos mit Raumklang . . . . .	26
4.2 Hardwarebasierte binaurale Simulation von Raumklangsystemen . . . . .	27

4.3	Binaurale 3D-Audiosoftware: VST-Plugin und virtuelle Soundkarte . . . . .	29
4.4	Das Vorgängerprojekt "Augmented-Reality-Audio-SOC" als Ausgangspunkt	30
<b>5</b>	<b>Aufbau und Auswertung von Studien im Audiobereich</b>	<b>32</b>
5.1	Theoretische Grundlagen zur Studienrealisierung . . . . .	32
5.1.1	Statistische Auswertung einer Studie im Audiobereich . . . . .	36
<b>6</b>	<b>Durchführung einer Studie zur Bewertung der Distanzmodellierung</b>	<b>38</b>
6.1	Erstellung des 3D-Audio-Modells für die Studie . . . . .	38
6.1.1	Fähigkeiten der entwickelten Softwareraumschallsimulation . . . . .	39
6.1.2	Aufbau der Software für die Raumschallsimulation . . . . .	42
6.1.3	Anfertigung benötigter Filter mit Hilfe von Kunstkopfmessungen . . . . .	44
6.2	Untersuchungsgegenstände und Aufbau der Studie . . . . .	47
6.2.1	Auftretende Einschränkungen durch die Pandemiesituation . . . . .	47
6.2.2	Ablaufplan der videobasierten Studie . . . . .	49
6.2.3	Ausführliche Beschreibung der durchlaufenen Testfälle . . . . .	54
6.2.4	Aufbau des für die Studie verwendeten PDF-Fragebogens . . . . .	61
<b>7</b>	<b>Auswertung der Studie zur auditiven Distanz</b>	<b>63</b>
7.1	Übersicht über die Proband_Innen und deren jeweilige Versuchsumgebungen	63
7.2	Detailauswertung der einzelnen Teilversuche . . . . .	65
7.2.1	Versuch 1: Überblick und Vergleich der Geräuschquellen . . . . .	65
7.2.2	Versuch 2: Richtungshören mit optischen Markierungen . . . . .	68
7.2.3	Versuch 3: Bewegung, Natürlichkeit und Externalisierung . . . . .	70
7.2.4	Versuch 4: Bestimmung des Auflösungsvermögens . . . . .	73
7.2.5	Versuch 5: Tests mit akustischen Markierungen . . . . .	74
7.3	Methodenvergleich und Nebenbetrachtungen . . . . .	76
7.3.1	Azimet-Fehler bei akustischen und optischen Markierungen . . . . .	76
7.3.2	Auswirkung der Körpergröße auf die Richtungswahrnehmung . . . . .	77
7.3.3	Effekt von Expertise, Monitor, Audiokonsum und Videospiele . . . . .	78
7.4	Zusammenfassung der Ergebnisse und Methodenkritik . . . . .	80
<b>8</b>	<b>Verwendung der Erkenntnisse über das erzeugte VR-3D-Audio-Modell</b>	<b>82</b>
8.1	Mögliche Funktionserweiterungen des Modells . . . . .	82
8.2	Optimierung und Ausbau der Nutzerstudie . . . . .	83
8.3	Anwendungsszenarien für das erstellte Modell . . . . .	83

<b>9</b>	<b>Fazit</b>	<b>84</b>
<b>A</b>	<b>Anhang</b>	<b>89</b>
A.1	Gesprochener Begrüßungstext . . . . .	89
A.2	Mittelwert, Standardabweichung und Median optisches Richtungshören . .	90
A.3	Externalisierung und Natürlichkeit . . . . .	91
A.4	Box-Whisker-Plots rein akustische Versuche . . . . .	92
A.5	Distanzeinschätzung bezogen auf die Körpergröße . . . . .	93
	<b>Selbstständigkeitserklärung</b>	<b>102</b>

# Abbildungsverzeichnis

2.1	Azimut (links) und Elevation (rechts) . . . . .	2
2.2	Logarithmische Darstellung der Hörfläche aus [16] . . . . .	3
2.3	Bereich des Cone-of-Confusions (blau) . . . . .	4
2.4	Wirkung der Distanzveränderung auf den Einfallswinkel des Schalls . . . . .	6
2.5	Zeitliche Abfolge bei der Wahrnehmung eines Schallereignisses . . . . .	8
3.1	VR-Headset mit HMD und integriertem Kopfhörer (Foto: Jan Strothmann) . . . . .	10
3.2	Zusammenhang von Umgebung, System und Ein-/Ausgangssignal . . . . .	12
3.3	MLS-Schieberegister mit Abgriffen, Initialisierung und Ausgabe . . . . .	15
3.4	Zeitliche Ausbreitung der Elementarwellen bei der Wellenfeldsynthese . . . . .	18
3.5	Das Ambisonics-Mikrophon SoundField SPS200 . . . . .	19
3.6	Zusammenhang von Hörerposition, Lautsprecher und beteiligten Winkeln . . . . .	20
3.7	Umwandlung aus dem A-Format in binaurale Quadrophonie . . . . .	20
3.8	Virtuelle Quadrophonie zur Verwendung von Ambisonics . . . . .	21
3.9	Verteilung der Spiegelschallquellen im Raum . . . . .	24
3.10	Zeitlicher Verlauf des durch die Spiegelschallquellen erzeugten Musters . . . . .	25
4.1	Physikalischer Aufbau des ARA-Systems . . . . .	30
4.2	Komponentenkommunikation . . . . .	31
6.1	Direkte Filterung von Direktschall und Spiegelschallquellen . . . . .	39
6.2	Panningbasierte Wiedergabe der Spiegelschallquellen . . . . .	40
6.3	Überlappung der Teilstücke bei Bewegung der virtuellen Schallquelle . . . . .	43
6.4	Aufbauvarianten für die Messung der HRTFs im Schallmessraum . . . . .	44
6.5	Winkel der drei Höhenstufen bei der Kunstkopfmessung . . . . .	45
6.6	Stereospur der aneinandergereihten Impulsantworten der HRTF-Messung . . . . .	45
6.7	Darstellung der für die beiden Messungen verwendeten Geräte . . . . .	46
6.8	Logische Verbindungen der eingesetzten Messgeräte . . . . .	46
6.9	Ursprünglich geplanter Versuchsaufbau an der HAW. . . . .	48

6.10	FOV für die Versuchsdurchführung . . . . .	48
6.11	Positionierungsschema für die Tonquellen . . . . .	51
6.12	Grundrisse der mit dem Raumschallmodell simulierten Räume . . . . .	52
6.13	Aufbau für die Aufnahme der Geräusche in der Studie . . . . .	53
6.14	Monitorbild der Richtungs- (links) und Entfernungsfragen (rechts) . . . . .	55
6.15	Raster für die Angabe des Winkels der Schallquelle . . . . .	56
6.16	Bewegungsmuster für den Versuch mit bewegten Tonquellen . . . . .	57
6.17	Angezeigte Bilder: Saal (links) und kleines Zimmer (rechts) . . . . .	57
6.18	Richtungshören mit akustischen Referenzpunkten . . . . .	59
A.1	Mittelwert des Azimut-Fehlers beim optischen Richtungshören . . . . .	90
A.2	Standardabweichung des Azimut-Fehlers beim optischen Richtungshören . . . . .	90
A.3	Median des Azimut-Fehlers beim optischen Richtungshören . . . . .	90
A.4	Externalisierung und Natürlichkeit Direktschall . . . . .	91
A.5	Externalisierung und Natürlichkeit sparsame Modellierung . . . . .	91
A.6	Externalisierung und Natürlichkeit genaue Modellierung . . . . .	91
A.7	Box-Whisker-Plot akustisches Richtungshören . . . . .	92
A.8	Box-Whisker-Plot akustisches Distanzhören . . . . .	92
A.9	Fehler bei der Distanzeinschätzung bezogen auf die Körpergröße . . . . .	93

# Diagrammverzeichnis

1	GNU-Octave-Plot der Normalverteilung . . . . .	37
2	Impuls ohne, mit niedriger und mit hoher Spiegelschallquellenordnung . . . . .	41
3	FFT-Analyse der Frequenzmaxima der verwendeten Geräusche . . . . .	53
4	Auf 5 Jahre gerundetes Alter der Proband_Innen . . . . .	64
5	Auf 10 cm gerundete Körpergröße der Proband_Innen . . . . .	64
6	Brummen mit und ohne Raumschall bei der Richtungsauswahl . . . . .	65
7	Vergleich verschiedener Geräusche bei der Richtungsauswahl . . . . .	66
8	Vergleich verschiedener Geräusche bei der Distanzauswahl . . . . .	67
9	Azimut-Fehler beim Richtungshören mit optischen Markierungen . . . . .	68
10	Hinten-Vorne-Fehler beim Richtungshören mit optischen Markierungen . . . . .	69
11	Bewegungserkennung nach verwendetem Raum und Modell . . . . .	70
12	Bewegungserkennung nach gefragter Bewegungsrichtung . . . . .	70
13	Übersicht Externalisierung (links) und Natürlichkeit (rechts) . . . . .	71
14	Auswirkung der gezeigten Bilder auf die Natürlichkeit . . . . .	72
15	Distanzauflösungsvermögen verschiedener Modellierungsarten und Räume . . . . .	73
16	Mittelwert des Azimut-Fehlers beim akustischen Richtungshören . . . . .	74
17	Standardabw. des Azimut-Fehlers beim akustischen Richtungshören . . . . .	74
18	Mittelwert des Fehlers bei der akustischen Distanzeinschätzung . . . . .	75
19	Standardabw. des Fehlers bei der akustischen Distanzeinschätzung . . . . .	75
20	Vergleich der optischen und akustischen Richtungsgenauigkeit . . . . .	76
21	Azimut-Fehler bezogen auf die Körpergröße . . . . .	77
22	Mittelwert des Azimut-Fehlers bezogen auf Gruppen von Körpergrößen . . . . .	78

# Tabellenverzeichnis

6.1	Absorptionsgrad verschiedener Materialien im Frequenzbereich 1-4 kHz . . .	52
7.1	Median des Azimut-Fehlers nach Gruppen . . . . .	79
7.2	Abschließender Vergleich der drei Modellierungsvarianten . . . . .	80

# Abkürzungen

**AC97** Audio Codec '97.

**AR** Argumented Reality.

**ASIO** Audio Stream Input/Output.

**ASM** Algorithmic State Machine.

**BRIR** Binaural Room Impulse Response.

**COVID-19** Coronavirus Disease 2019.

**CPU** Central Processing Unit.

**DAW** Digital Audio Workstation.

**dB** Dezibel.

**DSGVO** Datenschutz-Grundverordnung.

**DSP** Digitaler Signalprozessor.

**FFT** Fast Fourier Transformation.

**FIR-Filter** Finite Impulse Response Filter.

**FOV** Field of View.

**FPGA** Field Programmable Gate Array.

**GIMP** GNU Image Manipulation Program.

**GPL** General Public License.

**GUI** Graphical User Interface.

**HAW** Hochschule für Angewandte Wissenschaften.

**HMD** Head Mounted Display.

**HRTF** Head-Related Transfer Function.

**IIR-Filter** Infinite Impulse Response Filter.

**ILD** Interaural Level Difference.

**ITD** Interaural Time Difference.

**JACK** JACK Audio Connection Kit.

**LZI-System** lineares zeitinvariantes System.

**MLS** Maximum Length Sequence.

**OSC-Protocol** Open Sound Control Protocol.

**RIR** Room Impulse Response.

**SRV** Signal-Rausch-Verhältnis.

**VBAP** Vector Base Amplitude Panning.

**VPD** Virtual Product Development.

**VR** Virtual Reality.

**VST** Virtual Studio Technology.

**WDM** Windows Driver Model.

**WFS** Wellenfeldsynthese.

# 1 Einleitung

Das Themengebiet 3D-Audio ist in der Forschung bereits etabliert, dennoch bietet es weiter Raum für neue Anwendungen und Ansätze. Im Alltag finden sich verschiedene Methoden wieder, um ein räumliches Hörerlebnis zu reproduzieren, z.B. der Raumklang in Kinosälen. Durch die Entwicklungen im Bereich Virtual Reality (VR), z.B. durch die Oculus-Rift<sup>1</sup>, ist das kopfhörerbasierte 3D-Audio vermehrt in den Fokus getreten. Kopfhörer können mit dem in der Brille verwendeten Head Mounted Display (HMD) und Headtracking kombiniert werden, um ein immersives VR-Erlebnis zu erzeugen. Mit binauralem 3D-Audio wird über Kopfhörer ein Richtungshören ermöglicht. Ein Ansatz hierfür ist die Filterung mit einer Head-Related Transfer Function (HRTF), um so trocken aufgenommene Signale direkt aus bestimmten Richtungen erscheinen zu lassen.

Ziel der Arbeit ist es festzustellen, ob eine erstellte, potenziell echtzeitfähige Distanzmodellierung für den Audiobereich dieser Systeme verwendet werden kann und welche Auswirkungen dies auf die anderen Aspekte der auditiven Wahrnehmung hat.

Das dazu benötigte Grundwissen über die beteiligten physikalischen und psychoakustischen Effekte aus den Bereichen Akustik, Hören und Raumschall, bei dem Richtungs- und Distanzhören wird deshalb zu Beginn des Grundlagenkapitels vermittelt. Anschließend werden die Anforderungen von VR-Systemen thematisiert und Verfahren aus dem Gebiet 3D-Audio verglichen. Es wird vertieft auf die verwendete Methode eingegangen und die Auswahl begründet. Außerdem werden die Möglichkeiten zur Erzeugung eines Raum-/ Distanzeindrucks vorgestellt. Dem folgen aktuelle Anwendungs- und Fallbeispiele. Das Grundlagenkapitel endet mit Informationen zur Durchführung von Studien im Audiobereich. Den Hauptteil nimmt die Benutzerstudie ein. Zuerst wird detailliert auf das verwendete Modell und die für dessen Erstellung nötigen Messungen eingegangen. Anschließend wird das gewählte Vorgehen beschrieben und die Auswahl der Testfälle begründet. Darauf folgt die Auswertung der Studie inklusive der Nebenbetrachtungen. Den Abschluss bilden Nutzungs- und Erweiterungsszenarien des erzeugten Modells und das abschließende Fazit.

---

<sup>1</sup> Informationen abrufbar unter: <https://www.oculus.com/blog/oculus-ces-2015/>

## 2 Auditive Wahrnehmung der Richtung und Distanz von Tonquellen

In diesem Kapitel wird das notwendige Grundwissen über das menschliche Gehör und dessen Fähigkeiten bereitgestellt. Der Schwerpunkt liegt auf der Wahrnehmung von Distanz und Richtung. Im folgenden Kapitel kann darauf aufbauend eine 3D-Audio-Technik für die Verwendung in VR-Systemen ausgewählt werden. In diesem Kapitel werden zuerst die Informationen zu der auditiven Wahrnehmung als solche und wichtige Begriffe aus dem Audibereich gegeben. Anschließend wird zuerst das Richtungshören und hiernach das Entfernungshören beschrieben. Abschließend wird der für die Distanzwahrnehmung ausschlaggebende Raumschall dargestellt.

### 2.1 Grundlegende Eigenschaften des menschlichen Gehörs

Das Sehvermögen ist der dominante menschliche Sinn, doch auch das Gehör leistet wichtige Funktionen wie die frühzeitige Erkennung von Gefahren und verbale Kommunikation. Es ist weder auf das Sichtfeld beschränkt, noch von optischer Verdeckung betroffen.

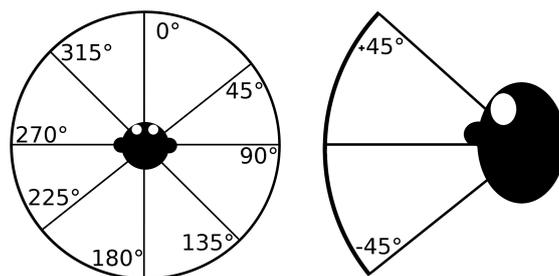


Abbildung 2.1: Azimut (links) und Elevation (rechts)

Zur Beschreibung der Richtung ist das Kugelkoordinatensystem geeignet. Der horizontale Winkel wird Azimut  $\varphi$ , der Höhenwinkel wird Elevation  $\theta$  genannt (Abb. 2.1).

Die Reichweite des Gehörs ist von der Lautstärke abhängig und schwankt dementsprechend stark. Bei dem wahrgenommenen Schall, also den Tönen und Geräuschen, handelt es sich um Schwingungen des Luftdrucks im dreidimensionalen Raum. Durch die Umgebungsluft bewegen sich diese Wellen mit der Schallgeschwindigkeit von 343,2 m/s. Die Luftdruckschwankungen erzeugen dann eine Schwingung des Trommelfelds in beiden Ohren. Wie bei einem Messgerät, kann diese Hörleistung des Menschen angegeben werden (Abb. 2.2).

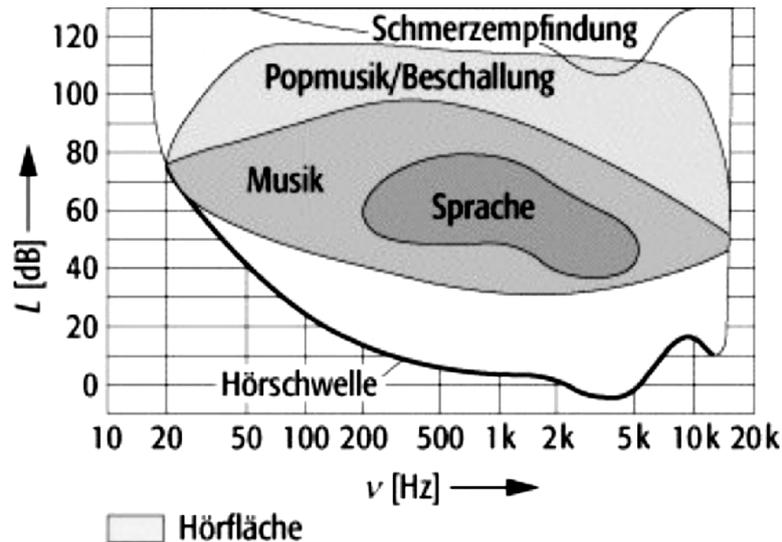


Abbildung 2.2: Logarithmische Darstellung der Hörfläche aus [16]

Den Nullpunkt der Skala bildet die Hörschwelle, ab der ein Ereignis wahrgenommen wird. Der Schalldruckpegel wird in Dezibel (dB) angegeben. Bei ca. 130 dB befindet sich die obere Grenze des Hörvermögens. Lautere Geräusche führen zu Schmerzen und können das Gehör beeinträchtigen.

Das menschliche Gehör ist in der Empfindlichkeit frequenzabhängig. Um dies zu korrigieren, wird Bewertungsfilter A eingesetzt und man erhält die übliche Angabe in dB(A). Für die Beschreibung der empfundenen Lautstärke ist diese Skala nicht sinnvoll, da diese in Sone angegeben wird. Eine Verdoppelung des Zahlenwertes in Sone bedeutet eine Verdoppelung der empfundenen Lautstärke.

### 2.1.1 Funktionsweise des menschlichen Richtungshörens

Mit dem Gehör kann die Richtung von Geräuschquellen bestimmt werden. Wichtig ist das binaurale Hören, welches durch zwei Ohren ermöglicht wird. Es geht über bloße Lautstärkeunterschiede hinaus. Die gleichzeitige Wahrnehmung von Schall an zwei verschiedenen aber bekannten Positionen ist der entscheidende Faktor [1]. Die beschriebenen Aspekte des Richtungshörens werden unbewusst genutzt.

Die zwei Begriffe Interaural Level Difference (ILD) und Interaural Time Difference (ITD) sind für die horizontale Einordnung relevant. Die ILD beschreibt die frequenzabhängige Abschwächung des Signals durch die Dämpfungswirkung des Kopfes im Weg des Schalls und die zwischen den Ohren zurückgelegte Strecke. Höhere Frequenzen werden hiervon stärker beeinflusst und es liegt eine Abhängigkeit vom eintreffenden Winkel vor.

Die ITD ist hingegen bei der Lokalisierung von tieferen Frequenzen ausschlaggebend. Mit der ITD wird eine durch die Laufzeit zwischen den Ohren verzögerte Wahrnehmung eines Signals beschrieben. Im einfachen Beispiel wird eine Schallquelle rechts neben dem Ohr zuerst dort und nach einer durch Schallgeschwindigkeit und Ohrenabstand bestimmten Zeit am linken Ohr wahrgenommen. Mit diesen beiden Hilfsmitteln bleibt noch ein Bereich übrig, in dem sich das Geräusch befinden kann: Der Cone-of-Confusion (Abb. 2.3).

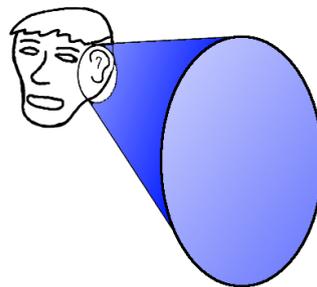


Abbildung 2.3: Bereich des Cone-of-Confusions (blau)

Die Position der Schallquelle wird mit Hilfe der HRTF weiter aufgelöst. Unter HRTF versteht man die Filterwirkung, welche die Form des äußeren Ohres, die Pinna, und der Oberkörper auf den eingehenden Schall hat. Sie ermöglicht eine genaue Bestimmung der Elevation. Durch eine Kopfbewegung kann die genaue Position anschließend aufgelöst werden [21]. Die HRTF ist für jeden Menschen individuell. Sie kann mit In-Ear-Mikrofonen direkt gemessen werden oder auch für einen Kunstkopf bestimmt werden. Bei diesen Messungen wird die ITD und ILD gleichzeitig erhalten. Die so gemessenen HRTFs können für die Erzeugung von binauralem 3D-Audio verwendet werden.

### 2.1.2 Ausgewählte Aspekte und Eigenheiten des Entfernungshörens

Neben der Fähigkeit, die Richtung eines Geräusches zu bestimmen, kann mit dem Gehör auch dessen Entfernung eingeschätzt werden. Durch die Kombination dieser Fähigkeiten kann eine Schallquelle lokalisiert werden. Ähnlich des Richtungshörens gibt es auch hier wieder verschiedene Teilaspekte, die zusammen die Distanzeinschätzung ermöglichen. Je nach Situation werden verschiedene Distanz-Hinweise vom Gehör wahrgenommen, welche unterschiedlich stark die Wahrnehmung beeinflussen [32]. Es gibt die Gruppen der statischen, dynamischen und der multisensorischen Hinweise. Letztere beruhen auf einer Kombination des Gehörs mit anderen Sinnen. Die meisten Hinweise sind dabei statischer Natur, d.h. ohne Bewegung von Quelle oder hörender Person.

Der einflussreichste der Hinweise ist das Verhältnis von Direktschall zu Raumschall. Also dem Verhältnis von direkt von der Schallquelle stammenden zu zuerst von der Umgebung reflektiertem Schall. Je höher der Anteil des indirekten Schalls ist, desto größer ist die Entfernung der Tonquelle von der Hörerposition.

Zusätzlich zu der Entfernung der Quelle enthält der eintreffende Raumschall auch Informationen über die Umgebung. Je nach Umgebung ist der Raumschall unterschiedlich stark ausgeprägt. Physikalisch gesehen ist der Raumschall von der Geometrie der Umgebungsobjekte, deren Oberflächen und des dabei durchdrungenen Mediums abhängig. In dem Abschnitt 2.1.3 wird dieses Thema ausführlich erläutert.

Die Lautstärke des Geräusches nimmt mit der Entfernung ab. Die Luftschwingungen, die als Schall wahrgenommen werden, breiten sich dreidimensional aus und verteilen dabei ihre Energie auf eine größere Oberfläche. Die Schallleistung  $P_{ak}$  nimmt deshalb mit  $\frac{1}{r^2}$  ab, dabei ist der Radius  $r$  der Abstand von der Tonquelle. Da die in dieser Entfernung messtechnisch erfasste Amplitude proportional zur Wurzel der Schallleistung ist, nimmt die Amplitude mit  $\frac{1}{r}$  ab. Der mit der Amplitude zusammenhängende Schalldruck ergibt dann die psychoakustische Größe der Lautstärke, wenn man diesen auf die frequenzabhängige Hörschwelle bezieht. Als weiterer Einfluss existiert die dämpfende Wirkung der Umgebungsluft. Durch dieses Medium bewegt sich der Schall. Wie bei einem Tiefpassfilter werden hohe Frequenzen stärker reduziert als niedrige. Die Auswirkung nimmt mit der Entfernung zu, wird aber erst ab 15 Metern relevant. Für Innenräume kann dies also meist vernachlässigt werden. Bei den bisher betrachteten Punkten handelt es sich um statische Distanzhinweise.

Ein dynamischer Hinweis ist der Dopplereffekt. Der Dopplereffekt beschreibt eine Frequenzverschiebung von Schwingungen, welche von einer Quelle ausgesandt werden, die sich relativ zur Beobachterposition bewegt. Nähert sich die Quelle der Beobachterpositi-

on, dann werden die Wellen zusammengestaucht, die Wellenlänge nimmt ab und damit erhöht sich die Frequenz. Andersherum werden bei einer sich entfernenden Quelle die Wellen gestreckt, die Wellenlänge nimmt zu und damit wird die Frequenz verringert.

Ein Beispiel aus dem alltäglichen Leben ist die wahrnehmbare Veränderung der Tonlage des Martinshorns eines vorbeifahrenden Einsatzfahrzeuges: Wenn sich dieses nähert, dann ist die Tonlage des Martinshorns höher. Entfernt sich das Fahrzeug wieder, dann kehrt sich der Effekt um und die Tonlage ist tiefer. Für die Entfernungseinschätzung bedeutet das Auftreten des Effekts, dass diese verbessert wird. Ebenfalls in die Gruppe der dynamischen Hinweise gehört die Kopfbewegung. Mit dieser können mit HRTF, ILD und ITD Entfernungsinformationen extrahiert werden.

Ein multisensorischer Hinweis ist die Kombination von Gehör und Sehsinn. Dadurch verbessert sich die Entfernungseinschätzung, wobei der ausgeprägtere Sehsinn dominiert. Ebenfalls von Bedeutung ist, ob die Person das Geräusch bereits kennt. Die Fähigkeit zur Entfernungseinschätzung nimmt bei bekannten Geräuschen zu [5]. Bei Entwurf und Auswertung von Versuchen muss dies berücksichtigt werden.

Die Entfernung zur Quelle hat auch Auswirkungen auf die HRTF im Nahbereich. Der Effekt ist ausgeprägter, je näher eine Quelle ist. Eine, absolut gesehen, kleine Positionsveränderung führt zu einer relativ großen Änderung des Einfallswinkels (Abb. 2.4).

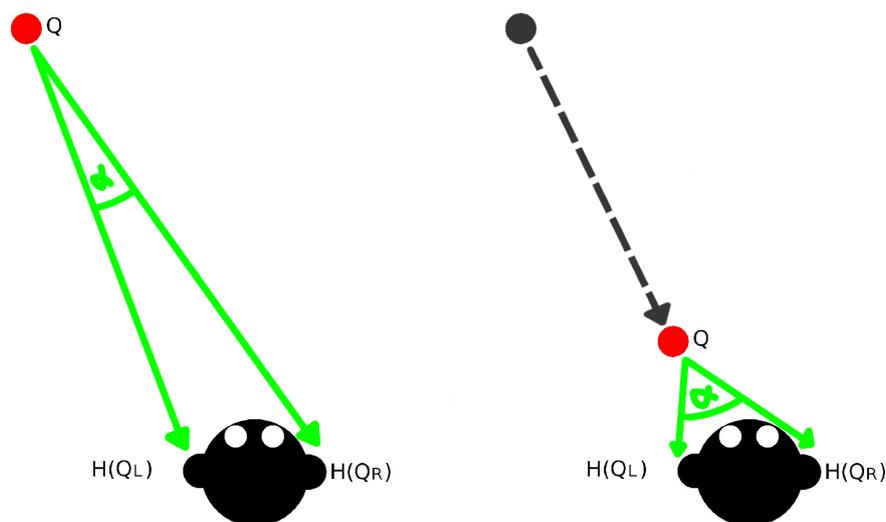


Abbildung 2.4: Wirkung der Distanzveränderung auf den Einfallswinkel des Schalls

Die Quelle  $Q$  bewegt sich auf die Hörerposition zu. Trotz gleicher Richtung ändert diese Entfernungsänderung für beide Ohren den Einfallswinkel  $\alpha$  und so die HRTFs  $H(Q_L)$  und  $H(Q_R)$ .

### Abschließender Vergleich zur Auswahl der Entfernungsmodellierung

Aus den vorgestellten Aspekten des Entfernungshörens müssen nun die passenden für eine Modellierung ausgewählt werden. Für die Glaubwürdigkeit der angestrebten Modellierung ist es wichtiger die richtigen Hinweise umzusetzen, als möglichst viele [19],[9]. Die vorgestellte Auswahl zeigt die Bedeutung von Raumschall.

Mit der wissenschaftlichen Arbeit *Direct comparison of the impact of the head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source* von D. Begault[6] können noch zusätzliche Aspekte in die Entscheidung einbezogen werden. Es werden die Auswirkungen von personalisierten HRTFs, Raumschall und Headtracking verglichen. Als personalisierte HRTF werden Übertragungsfunktionen verstanden, welche direkt an Proband\_Innen für den Versuch gemessen wurden. Der Raumschall wird in zwei Ausprägungen modelliert:

In beiden Fällen sind per Software *CATT-Acoustic*<sup>1</sup> erzeugte frühe Reflexionen vorhanden, welche mit HRTFs gefiltert werden. In einer Variante wird diffuser Nachhall hinzugefügt. Untersuchungsgegenstand sind die Punkte Azimut-Fehler in der zweidimensionalen Ebene, Elevationsgenauigkeit, Hinten-Vorne-Fehler und schlussendlich Externalisierung. Die Begriffe werden in Abschnitt 5.1 erläutert.

Werden vom Raumschall nur die frühen Reflexionen simuliert, dann reduziert sich der Azimut-Fehler. Wird auch der diffuse Nachhall hinzugefügt, nimmt die Genauigkeit ab, bleibt aber höher als in der Version ohne Raumschall. Die wahrgenommene Externalisierung wird in allen Fällen erzeugt.

Das Headtracking verstärkt die Auswirkung auf die Untersuchungsaspekte meist und die personalisierten HRTFs hatten die geringsten Auswirkungen.

In der Studie *Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap* von Tianshu Qu et al. [24] wird die isolierte Auswirkung von HRTFs mit Distanzinformationen untersucht. In einem reflexionsfreien Raum wurde eine Tonquelle aus verschiedenen Entfernungen aufgenommen und so die Testfälle erzeugt. Es wird gezeigt, dass die Verwendung von so erzeugten HRTFs nur für Tonquellen, welche sich sehr nah an der Hörerposition befinden, Auswirkungen hat.

Für das zu erstellende Distanzmodell bedeutet dies, dass Raumschall als wichtigster Aspekt realisiert wird. Wegen der Einfachheit bei der Implementierung wird zusätzlich noch die Lautstärke modelliert. Wegen der geringen Auswirkungen und des hohen Aufwandes werden die distanzabhängigen HRTFs nicht eingesetzt.

---

<sup>1</sup> Informationen Online abrufbar: <http://www.catt.se/>

### 2.1.3 Raumschall als zentraler Aspekt des Entfernungshörens

Raumschall ist der Anteil des Gesamtschalls, der nicht auf direktem Wege von der Schallquelle zur Hörerposition gelangt, diese aber trotzdem erreicht. Der Raumschall wird mindestens einmal, aber meist häufiger an den Wänden und Objekten reflektiert. Dadurch ist die zurückgelegte Strecke länger und dieser Teil des Schalls trifft verzögert ein. Bei den auftretenden Reflexionen wird ein materialabhängiger Anteil der Schallenergie absorbiert, wobei der Grad der Absorption frequenzabhängig ist. Die Begriffe Direktschall, Raumschall, Nachhall, Echo und Hallmaß [10] werden im zeitlichen Verlauf beschrieben.

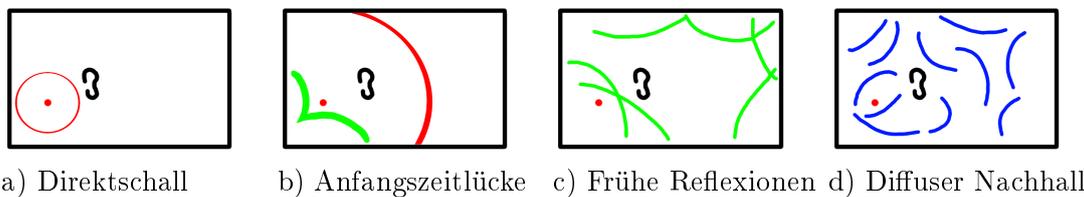


Abbildung 2.5: Zeitliche Abfolge bei der Wahrnehmung eines Schallereignisses

Dies ist für einen rechteckigen Raum mit einem schwarzen Ohr als Hörerposition dargestellt (Abb. 2.5). Die Ausbreitung des Schalls beginnt bei der Geräuschquelle und ist abhängig von deren Abstrahlcharakteristik. Gerichtete Quellen strahlen nur in eine Richtung ab. Die als roter Punkt dargestellte Punktschallquelle strahlt in alle Richtungen gleichmäßig ab. Der Direktschall (rot) *a*) ist der Anteil des Gehörten, der auf direktem Wege und ohne Reflexionen eintrifft und nur vom durchlaufenen Medium beeinflusst wird. Hiernach breitet sich der Schall weiter aus, wird von den Wänden zurückgeworfen und trifft als frühe Reflexionen (grün) *c*) wieder bei der Hörerposition ein. Die Zeit zwischen dem Eintreffen der frühen Reflexionen und des Direktschalls wird Anfangszeitlücke *b*) genannt. Wenn diese größer als 50 ms ist, dann wird dies Echo genannt und es werden zwei getrennte Ereignisse wahrgenommen. Wird diese Zeit unterschritten, gilt *das Gesetz der ersten Wellenfront* oder auch Haas-Effekt [15]: Es wird ein einziges Schallereignis aus Richtung der ersten eintreffenden Wellenfront wahrgenommen. Anschließend wird der Schall in dem Raum immer wieder reflektiert und dabei absorbiert. An der Hörerposition wird dies als diffuser Nachhall (blau) *d*) wahrgenommen. Das Verhältnis von Direkt- zu Raumschall wird als Hallmaß bezeichnet und ist für die Distanzbestimmung ausschlaggebend. Der Bereich, in dem Direkt- und Raumschall die gleiche Amplitude aufweisen, wird als Hallradius  $r_H$  bezeichnet. Zusätzlich treten stehende Wellen auf, sogenannte Raummoden, welche Töne unterhalb einer vom Nachhall abhängigen Grenzfrequenz stören.

## 3 Erstellung von 3D-Audio mit auditiver Distanz für VR-Systeme

Auf Basis der beschriebenen Grundlagen zur auditiven Wahrnehmung kann nun das zu verwendende Modell für die Erzeugung von 3D-Audio ausgewählt werden. Als erstes wird verdeutlicht, in welchem Umfeld die beschriebenen 3D-Audio-Techniken eingesetzt werden sollen. Dazu wird das Themengebiet VR angerissen und die Anforderungen an den Audiopart dargestellt. So können die Entwurfsentscheidungen nachvollzogen werden. Anschließend werden die Grundlagen der Audiosignalverarbeitung präsentiert.

Hiernach werden verschiedene Möglichkeiten diskutiert, künstlich 3D-Audio zu erzeugen. Nach der Auswahl eines geeigneten Verfahrens für die Richtungssimulation, wird ein Verfahren zur Raumschallsimulation ausgewählt.

Grundlegend ist der Wunsch, eine Situation vorgeben zu können und trocken aufgenommene Geräusche so zu bearbeiten, dass ein Raumeindruck entsteht. Mit Situation ist in diesem Zusammenhang gemeint, dass es eine Beobachterposition und eine oder mehrere Tonquellen in einer definierten Umgebung vorhanden sind. Ein trockenes Signal ist ein Tonsignal ohne Richtungs- und Rauminformationen.

### 3.1 Eigenschaften und Besonderheiten von 3D-Audio in VR-Systemen

Ziel ist die Erprobung einer auditiven Distanzdarstellung, welche für die Nutzung in VR-Systemen geeignet ist. Unter dem Begriff VR versteht man die Möglichkeit zur Betrachtung von oder die Interaktion mit einer simulierten Wirklichkeit. Die Computersimulation ahmt die Realität audiovisuell nach und soll das Eintauchen in diese, die Immersion, ermöglichen. Im Gegensatz zur Argumented Reality (AR) ersetzt die Simulation dabei die Wirklichkeit vollständig und erweitert diese nicht nur um Zusatzinformationen.

Hauptanwendungsgebiet ist die Simulation zu Trainingszwecken, die virtuelle Produktentwicklung und der Einsatz in der Unterhaltungsindustrie.

Von den fünf klassischen Sinnen des Menschen, also Sehen, Hören, Tasten, Riechen und Schmecken werden häufig nur die ersten beiden Sinne direkt angesprochen. Der Sehsinn durch ein HMD und das Gehör meist mit einem Kopfhörer (Abb. 3.1).



Abbildung 3.1: VR-Headset mit HMD und integriertem Kopfhörer (Foto: Jan Strothmann)

Ein HMD ist ein am Kopf befestigter Bildschirm. Durch die Aufteilung eines einzigen oder die Verwendung von zwei Bildschirmen wird ein dreidimensionales Bild ermöglicht. Von den übrigen Sinnen, also Gleichgewichtssinn, Temperatursinn und der Körperwahrnehmung werden einige Teilaspekte angesprochen. Eine hierfür genutzte Technik ist die Verfolgung der Position und der Bewegungen der nutzenden Person, was als Tracking bezeichnet wird. Es kann zwischen Inside-Out-Tracking und Outside-In-Tracking unterschieden werden. Eine Echtzeitbildanalyse ermöglicht Inside-Out-Tracking. Hierzu wird durch in das Gerät integrierte Sensorik die Umgebung ausgewertet und die Position im Raum bestimmt. Beim Outside-In-Tracking wird das zu verfolgende Objekt von außen beobachtet. Ein in Echtzeit durchgeführtes Tracking ermöglicht es, die Bewegungen des Kopfes zu verfolgen und damit die Bildinhalte an die aktuelle Ausrichtung anzupassen. Dadurch liefern Seh- und Gleichgewichtssinn keine unterschiedlichen Eindrücke und die Simulatorkrankheit wird vermieden. Auch wird es durch Tracking möglich, dass virtuelle Tonquellen trotz Kopfbewegung an ihrer Position verweilen und diese natürliche Lokalisationstechnik verwendet werden kann.

### 3.1.1 Zu erfüllende Rahmenbedingungen für Echtzeit-VR

Für alle beteiligten Komponenten, inklusive des betrachteten Audioparts, müssen für eine ausreichende Wiedergabetreue folgende Rahmenbedingungen eingehalten werden.

**Echtzeitfähigkeit mit hohem Durchsatz:** Die Zeit zwischen einer Aktion in der VR-Umgebung und der sichtbaren Reaktion durch die Umgebung muss zuverlässig niedrig sein. Üblich ist eine Bildwiederholrate der Anzeige im HMD von 90-120 Hz. So wird alle 12 ms ein neues Bild ausgegeben und stellt so die minimal erreichbare Latenz dar. Dazu kommt noch die Latenz der Verarbeitungskette, also des Tracking sowie Betriebssystem- und Treiberoverhead.

Gleichzeitig soll die dargestellte Umgebung detailliert sein. Durch die Nähe der Anzeigen zu den Augen und dem großen Bereich des Sichtfeldes, den diese einnehmen, ist eine hohe Auflösung pro Auge notwendig. Trotz allem muss die Aktualisierung der Simulationssituation, Grafik- und Soundberechnungen in dem Intervall abgeschlossen werden. Es handelt sich in diesem Zusammenhang um die minimal zu erreichende Bildrate und nicht den Durchschnittswert. Auf Grund der ansonsten einsetzenden Simulatorkrankheit sollte die Bildrate mit dem 99. Perzentil der Anzeigefrequenz von 90 Hz eingehalten werden. Für ein Audiomodell bedeutet dies, dass die Berechnungen möglichst ressourcenschonend ablaufen müssen, um den Durchsatz nicht zu gefährden. Gleichzeitig ist die Zeit zum Abschluss der Berechnungen durch die Latenzanforderungen begrenzt. Dies schränkt die Wahl der zu nutzenden Algorithmen ein.

**Interaktiv und Dynamisch:** Auch wenn inhaltlich ein vorgegebener Ablauf existiert, liegt durch die Möglichkeiten zur Orientierung und Bewegung in der VR-Umgebung eine interaktive Situation vor. Dadurch sind keine Vorberechnungen oder Vorproduktionen möglich. Das verwendete Audiomodell muss zur Laufzeit dynamisch an die Situationen anpassbar sein und auf die geänderte Orientierung reagieren können.

**Ergonomisch und Mobil:** Eine uneingeschränkte Bewegung in der VR-Umgebung muss möglich sein. Entweder muss die gesamte Ausrüstung so leicht und kompakt sein, dass diese am Körper getragen werden kann, oder sie muss ähnlich dem externen Outside-In-Tracking in den vorgegebenen Bereich zur Wiedergabe integriert sein. Die zur Zeit verwendeten HMDs werden häufig mit Kopfhörern kombiniert. Es sind aber auch lautsprecherbasierte Lösungen denkbar, bei der sich jedoch die Raumakustik des Wiedergaberaums auswirkt.

## 3.2 Grundlagen der filterbasierten Audiobearbeitung

In der Signalverarbeitung und in technischen Anwendungen werden Filter eingesetzt. Im Bereich der Audioverarbeitung werden diese genutzt, um Signalverläufe zu manipulieren. Beispiele dafür sind Hoch- und Tiefpassfilter oder Equalizer, welche das Spektrum des Ausgangssignals verändern. Es ist möglich, dass für die Filterung ein analoges Bauteil, eine Softwarelösung oder ein Digitaler Signalprozessor (DSP) eingesetzt wird.

### 3.2.1 Systemtheorie und Faltung als Basis der Filterung

Häufig soll ein Filter das Signalverhalten eines Systems reproduzieren oder kompensieren. Als System wird im Zusammenhang der Systemtheorie der Bestandteil bezeichnet, welcher das Eingangssignal zum Ausgangssignal wandelt. Das System wird als Blackbox verstanden und nur das Ein-/Ausgabe-Verhalten ist von Interesse. Es wird von der Umgebung umschlossen. Die Darstellung zeigt den Begriffszusammenhang (Abb. 3.2).

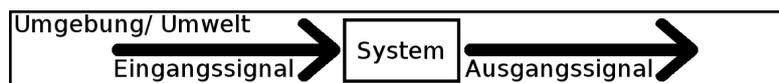


Abbildung 3.2: Zusammenhang von Umgebung, System und Ein-/Ausgangssignal

Durch Anregung des Systems und Messung der Impulsantwort kann das Verhalten bestimmt werden, sofern es sich um ein lineares zeitinvariantes System (LZI-System) handelt. Linear bedeutet in der Systemtheorie, dass das Superpositionsgesetz gilt: Wenn das Eingangssignal aus Überlagerungen von Teil-Eingangssignalen besteht, dann erhält man das Gesamt-Ausgangssignal durch Überlagerung der einzelnen Teil-Ausgangssignale. Zeit-invariant heißt, dass der Zeitpunkt der Messung keine Auswirkung auf das Ergebnis hat und nicht in dieses eingeht. Es ist möglich, die Messung auf der Zeitachse zu verschieben ohne dass die Werte dadurch angepasst werden müssten.

Das System kann alternativ über physikalische Zusammenhänge oder als Übertragungsfunktion beschrieben werden. Mit den Koeffizienten der Impulsantwort kann ein Signal gefiltert werden, um die Auswirkungen des Systems auf das Signal zu reproduzieren.

$$(g \circledast f)(x) = f(x) \circledast g(x) = \sum_{n \in \mathbb{Z}} f(x)g(x - n) \quad (3.1)$$

Die Gleichung 3.1 zeigt die diskrete Faltung ( $\circledast$ ) des Eingangssignals ( $f(x)$ ) mit der Impulsantwort ( $g(x)$ ), wodurch das Ausgangssignal ( $(g \circledast f)(x)$ ) entsteht.

Die Rechenvorschrift kann mit den Koeffizienten einer Impulsantwort genutzt werden, um ein Filter mit endlicher Impulsantwort, ein Finite Impulse Response Filter (FIR-Filter), zu erzeugen. FIR-Filter haben keine Rückkopplung und sind nur von den Eingangswerten abhängig. Deshalb klingen sie nach einer Zeit ab, d.h. sie sind stabil. Sie werden häufig linearphasig entworfen, sie weisen also keine Laufzeitunterschiede bei den Frequenzanteilen auf, weshalb es nicht zu hörbaren Phasenverzerrungen kommt. Durch die Verwendung der schnellen Fourier-Transformation mittels der Fast Fourier Transformation (FFT) kann der Rechenvorgang für längere Filterlängen beschleunigt werden. Es wird, statt im Zeitbereich im Frequenzbereich, gerechnet. Bei kürzeren Signalen kann eine erhöhte Latenz auftreten, da die Transformation zwischen Zeit- und Frequenzbereich nötig ist.

Filter mit unendlicher Impulsantwort, Infinite Impulse Response Filter (IIR-Filter), verfügen über eine Rückkopplung. Vorherige Ausgangswerte sind an der Erzeugung des neuen Ausgangssignals beteiligt. Dies kann genutzt werden, um nach einer einmaligen Anregung periodische Signale auszugeben. Es kann auch instabiles oder eskalierendes Verhalten mit selbstständiger Verstärkung auftreten. IIR-Filter ermöglichen es, den gleichen Betragsfrequenzgang mit einer kürzeren Filterlänge als ein FIR-Filter zu erreichen.

#### **Audiobearbeitung mit der Open Source Software GNU-Octave**

GNU-Octave ist die freie Implementierung der Sprachkonstrukte aus MATLAB [26] und ermöglicht die Erstellung scriptbasierter Anwendungen mit mathematischem Fokus. Die Software wird mit Toolboxen für bestimmte Anwendungszwecke erweitert. Die Signalverarbeitungstoolbox ermöglicht die Nutzung verschiedener Filter, das Einlesen und Erstellen von Audiodateien sowie die Frequenzanalyse. Auch ist die Generierung von Filtern, z.B. für einen gegebenen Frequenzgang, möglich. Grafiken können mit der MATLAB-Syntax über GNU-Plot ausgegeben werden. Im Gegensatz zur MATLAB-Lizenz ist die Verwendung nicht auf die Lehre beschränkt und damit auch für die Forschung zugänglich.

#### **Dekorrelation als Mittel zur Meidung störender Überlagerungen**

Eine störende Überlagerung bei der gleichzeitigen Wiedergabe mehrerer identischer Signale wird Korrelation genannt. Die Wechselwirkung zwischen den Signalen führt dazu, dass es Bereiche konstruktiver und destruktiver Überlagerungen gibt, welche von der Position im Raum abhängig sind. Dies äußert sich in einer Erhöhung oder Reduktion der

Lautstärke bis hin zur kompletten Auslöschung. Durch einen festen zeitlichen Versatz der Signale ändert sich der Ort, aber nicht das Auftreten dieser Störung.

Diesem Problem kann mit Dekorrelation entgegengewirkt werden. Dazu werden die beteiligten Signale an jedem Wiedergabezeitpunkt um einen zufälligen Betrag verschoben. Geschieht die Verschiebung unabhängig voneinander, dann wird die Korrelation statistisch verhindert und ist nicht wahrnehmbar. Eine derartige Verschiebung ist mit einem speziellen Allpass-FIR-Filter möglich, der eine Maximum Length Sequence (MLS) als Filterkoeffizienten verwendet. MLS werden in Abschnitt 3.2.2 vorgestellt. Allerdings kann durch diese Verschiebungen Phasing auftreten. Dabei handelt es sich um eine Phasenverschiebung, welche zu einem unnatürlichen Höreindruck führt. Bei einer Abtastrate von 44,1 kHz wird bei 2047 Koeffizienten ein natürlicher Eindruck wahrgenommen [31].

#### 3.2.2 Praktische Filtererzeugung durch Messung der Systemantwort

Ein etablierter Weg um das Verhalten eines unbekanntes oder physikalisch komplex zu beschreibenden Systems zu bestimmen, ist die Messung der Systemantwort. Zu diesem Zweck muss das System mit einem Eingangssignal angeregt werden, um das Ausgangssignal messen zu können [17]. Für die Anregung des Systems stehen verschiedene Signale zur Auswahl, welche sich in der Empfindlichkeit gegenüber Störungen und der Nachbearbeitung unterscheiden.

Ziel ist das Erreichen eines hohen Signal-Rausch-Abstandes. Der Signal-Rausch-Abstand, auch Signal-Rausch-Verhältnis (SRV) genannt, beschreibt das Verhältnis von Nutzsignal zum Hintergrundrauschen. Ist das SRV zu niedrig, kann die Aufnahme nicht verwendet werden, da das gewünschte Signal nicht enthalten ist oder die Störungen durch das Hintergrundrauschen zu stark sind.

**Impuls und Sprung:** Die Impulsfunktion ist das klassische Anregungssignal, bei dem direkt die Impulsantwort erhalten wird. In der Theorie muss ein Signal erzeugt werden, welches von dem Pegel Null auf den Maximalwert und wieder zurück wechselt. Dies ist in der Praxis nur näherungsweise möglich und die übertragene Energie und damit der Signalrauschabstand ist deshalb niedrig. Durch die kurze Energie-Spitze können zusätzliche Verzerrungen durch das Messequipment auftreten.

Die Sprungfunktion ähnelt der Impulsfunktion. Diese wechselt vom Nullpegel auf das Maximum und bleibt dort. Diese Methode ist einfacher zu realisieren, verhält sich sonst ähnlich. Zum Erhalt der Impulsantwort aus der Sprungantwort muss differenziert werden. Beide Methoden sind für unkorrelierte Störungen anfällig.

**Weißes Rauschen, rosa Rauschen und MLS:** Beim weißen Rauschen handelt es sich um ein breitbandiges, pseudozufälliges Signal ohne Wiederholungen. Rosa Rauschen wird aus dem weißen Rauschen erzeugt, indem ein Tiefpassfilter eingesetzt wird, welches das Signal um 3 dB pro Oktave abschwächt. Dies dient dem Schutz der für die Anregung eingesetzten Lautsprecher. Beide Signale ermöglichen ein höheres SRV als die Impulsfunktion. Unkorrelierte Signale, wie z.B. Störgeräusche, gehen nicht in die Messung ein. Die Impulsantwort wird bei weißem Rauschen mit einer Kreuzkorrelation von Anregungs- und Messsignal erzeugt. Bei rosa Rauschen muss der veränderte Frequenzgang beachtet werden. Die MLS ist ebenfalls ein pseudozufälliges Signal, im Gegensatz zu den vorher genannten enthält dieses aber nur den Minimal- und Maximalwert und wiederholt sich erst nach einer maximalen Länge [14]. Mit der MLS wird für die drei betrachteten Anregungssignale die maximale Energie und so der bestmögliche Signalrauschabstand erzielt. Sie kann mit einem Schieberegister mit Abgriffen erzeugt werden (Abb. 3.3).

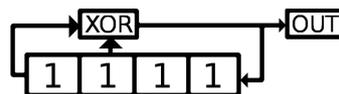


Abbildung 3.3: MLS-Schieberegister mit Abgriffen, Initialisierung und Ausgabe

Bei höherer Bitbreite des Schieberegisters sind durch Kombination und Positionierung der Abgriffe verschiedene MLS erzeugbar. Es kann bei der Messung nicht durch unkorrelierte Signale gestört werden. Die Impulsantwort wird wieder durch Kreuzkorrelation der gemessenen Systemantwort mit dem Anregungssignal erhalten.

**Sinus Sweep:** Eine weitere Möglichkeit zur Anregung eines Systems ist die Verwendung eines Sinus Sweeps. Dafür werden mit einem Wobbelgenerator im zeitlichen Verlauf logarithmisch ansteigende Sinusschwingungen erzeugt. Mit diesen Sinusschwingungen wird dann der gewünschte Frequenzbereich abgedeckt, z.B. der Bereich zwischen 20 Hz - 20 kHz. Die vorhergehenden Methoden haben den Nachteil, dass Lautsprecher nicht für die Wiedergabe von pseudozufälligen Werten ausgelegt sind. Die schwingende Membran lässt sich deutlich passender mit Sinusschwingungen ansteuern. Hierdurch ist ein höherer Anregepegel möglich, ohne den Lautsprecher zu beschädigen. Das Verfahren hat Vorteile in Situationen, in denen das SRV schlecht ist oder ein hoher Pegel benötigt wird. Zur Erzeugung der Impulsantwort aus dem gemessenen Signal wird mit Dekonvolution gearbeitet.

### 3.3 Vor- und Nachteile ausgewählter Verfahren zur 3D-Audio-Erzeugung

Mit den dargestellten Grundlagen und den Anforderungen durch das Einsatzszenario kann nun die eigentliche Methode zur dreidimensionalen Audioproduktion ausgewählt werden. Dazu werden die Verfahren und deren Vor- und Nachteile beleuchtet.

Grundlegend haben alle Methoden gemein, dass durch Berechnungen ein bearbeitetes Monosignal die gewünschte Räumlichkeit erzeugt. Das trockene Monosignal wird so zu einer virtuellen Schallquelle in der virtuellen Umgebung. Die in Abschnitt 3.1.1 definierten Rahmenbedingungen für die Eignung für VR müssen durch das System eingehalten werden. Die Auswahl erfolgt nach der Integrierbarkeit in das System, den Aufwand und der erreichten Richtungstreue.

Als Technik wird HRTF-Filterung eingesetzt, dies wird im ersten Abschnitt beleuchtet. Die alternativ möglichen Techniken werden ebenfalls dargestellt. Dies dient der Begründung der Auswahl und auch der Einführung notwendiger Begriffe.

#### 3.3.1 HRTF-Filterung für die Verwendung in binauralen Systemen

Viele VR-Systeme verwenden bereits Kopfhörer für die Wiedergabe des Audioanteils. Durch das Tracking des HMDs werden die Kopfbewegungen und damit auch das Audiowiedergabegerät im Raum erfasst. Für die Positionierung von virtuellen Tonquellen in der Umgebung wird deshalb ein Verfahren benötigt, welches diese relativ zur Ausrichtung der Kopfposition aus einer bestimmten Richtung erscheinen lässt. Auch muss dieses Verfahren sich schnell an die Veränderungen der Simulationssituation anpassen können. Dies ist mit der Wiedergabe von mit HRTFs gefalteten Monosignalen auf Stereokopfhörern möglich [34]. Durch die Filterung, mit einer für eine bestimmte Raumrichtung gemessenen HRTF, erscheinen die Tonquellen aus der gewünschten Richtung. Für eine glaubwürdige Simulation sind Filterkoeffizienten für alle Raumrichtungen und für verschiedene Höhenniveaus nötig. Auch müssen genügend Messungen vorhanden sein, um diese als Stützpunkte für eine Interpolation zu verwenden.

Durch ein nachgeschaltetes, inverses Filter kann die Filterwirkung des eigenen Außenohres, auf dem der genutzte Kopfhörer bei der Wiedergabe aufliegt, ausgeglichen werden. Hierzu ist die Erstellung eines Kopfhörerkompensationsfilters notwendig.

### 3.3.2 Alternative und ergänzende Techniken im Bereich 3D-Audio

Um die Einordnung des ausgewählten Verfahrens zu ermöglichen, wird dieses noch mit weiteren Verfahren aus dem Bereich 3D-Audio verglichen. Teilweise lassen sich auch verschiedene Verfahren kombinieren und so Teilaspekte in der Arbeit verwenden.

#### Panning zur Erzeugung von positionierbaren Phantomschallquellen

Die Nutzung von Phantomschallquellen und das zu deren Erzeugung verwendete Panning ist das Standardverfahren zur richtungsgebenden Audiowiedergabe [30]. Es handelt sich um ein psychoakustisches Verfahren mit einem kanalbasierten Ansatz. Es wird ausgenutzt, dass bei dem Hören des gleichen Signals aus zwei verschiedenen Richtungen eine einzige Tonquelle zwischen den Wiedergabekanälen wahrgenommen wird. Das Ausgangssignal wird durch Lautstärkeunterschiede der beteiligten Ausgabekanäle positioniert. Dies funktioniert nur für einen kleinen Bereich, den sogenannten Sweetspot, und schränkt den nutzbaren Bereich für die VR-Umgebung ein. Es setzt mindestens zwei Audiokanäle voraus, ist aber auch bei Mehrkanalsystemen anwendbar.

Mit bekannten Raumklangsystemen, wie z.B. Quadrophonie und 5.1-Wiedergabe, ist die Wiedergabe aus allen Richtungen möglich. Der optimale Wiedergabebereich wird gegenüber dem Stereopanning weiter eingeschränkt. Für Quadrophonie werden vier Lautsprecher an den Ecken eines um die Hörerposition gedachten Quadrats aufgestellt. Bei 5.1-Systemen stehen vier Hochtöner an diesen Positionen und ein Lautsprecher wird mittig positioniert. Kombiniert wird dies mit einem relativ frei platziertem Tieftöner. Auch in dem Bereich, in dem das Panning optimal funktioniert, ist die Richtungsbestimmung einer Lösung auf Basis von HRTFs unterlegen [18]. Der verwendete Algorithmus kann vor der Wiedergabe in einem Tonstudio oder in Echtzeit durchgeführt werden.

#### Wellenfeldsynthese als Möglichkeit zur direkten Schallfeldreproduktion

Die Wellenfeldsynthese (WFS) ist ein Verfahren der Holophonie, welches nicht auf psychoakustischen Effekten basiert, sondern die Schallwellen direkt reproduziert [2]. Dies ermöglicht eine genauere Richtungsbestimmung und die Wiedergabe ist nicht von einem Sweetspot abhängig. Die Technik basiert auf dem Huygensschen Prinzip, der Überlagerung von Elementarwellen. Aus vielen Elementarwellen wird die gewünschte Schallwelle geformt. Hierzu wird eine große Anzahl an Lautsprechern verwendet.

Diese sind zu Arrays zusammengefasst. Meist werden die Lautsprecherarrays auf einer horizontalen Ebene um den Wiedergabebereich aufgestellt. Dies schränkt die Wiedergabe auf eine Höhenstufe ein. Theoretisch ist auch die Reproduktion dreidimensionaler Wellen möglich. Die dargestellten Elementarwellen sind farblich den erzeugenden Lautsprechern zugeordnet. Die virtuelle Schallquelle und die durch Überlagerung erzeugte Gesamtschallwelle ist violett gefärbt (Abb. 3.4).

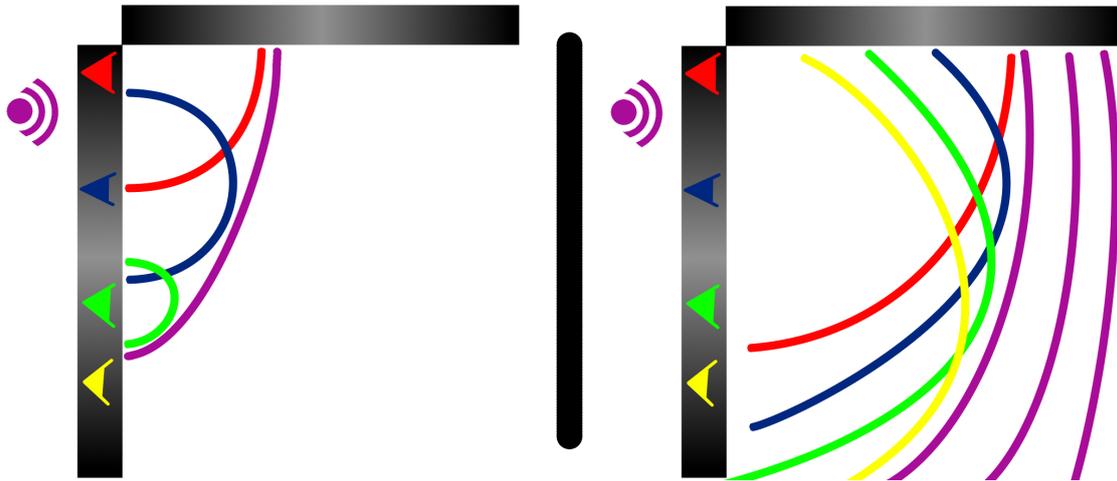


Abbildung 3.4: Zeitliche Ausbreitung der Elementarwellen bei der Wellenfeldsynthese

Die Ausgabesignale der beteiligten Lautsprecher müssen einzeln berechnet werden. Hierbei kann der verwendete Algorithmus sowohl in Software als auch in Hardware, z.B. auf einem Field Programmable Gate Array (FPGA), umgesetzt werden [29].

Alle Personen im Wiedergabebereich erhalten denselben Raumeindruck, um die Schallquellen lokalisieren können. Dies ermöglicht die Teilnahme mehrerer Personen an derselben Simulation. Nachteilig sind die Kosten und der benötigte Platz.

#### **Ambisonics als Format zur Aufnahme und Bearbeitung des Klangfeldes**

Schon seit langem wird im Bereich des dreidimensionalen Audios mit dem Ambisonics-Audioformat gearbeitet [12]. Es ermöglicht die Aufnahme, Speicherung und Verarbeitung mit etablierten Techniken und Methoden. Bei der Aufnahme bleiben die Richtungsinformationen inklusive des Höhenwinkels der Geräusche erhalten.

Im ersten Schritt müssen mit einer Messung die Rohdaten gewonnen werden. Dazu wird z.B. ein Soundfield-Mikrophon mit vier Kapseln in tetraederförmiger Anordnung verwendet (Abb. 3.5). Eine Mikrofonkapsel ist ein Aufnahmekanal eines Mikrophons.

Durch die Ausrichtung der Kapseln wird das Schallfeld aus allen Richtungen gleichzeitig aufgenommen und eine Aufnahme für Ambisonics der ersten Ordnung erstellt.



Abbildung 3.5: Das Ambisonics-Mikrofon SoundField SPS200

Für höhere Ordnungen werden mehr Mikrofonkapseln benötigt, dafür nimmt die Genauigkeit der Rauminformationen zu. In der Theorie werden alle Kapseln in einem Punkt liegend angenommen. Dies ist physikalisch nicht möglich und muss bei der Verarbeitung kompensiert werden. Die so erzeugte Aufnahme wird als A-Format bezeichnet und kann nicht direkt verwendet werden. Mit der Gleichung 3.2, welche die räumliche Verteilung der Kapseln nicht kompensiert, muss diese in das B-Format umgewandelt werden.

$$\begin{aligned} W &= LF + RF + LB + RB \\ X &= LF + RF - LB - RB \\ Y &= LF - RF + LB - RB \\ Z &= LF - RF - LB + RB \end{aligned} \tag{3.2}$$

Die Namen in der Gleichung entsprechen den Kanälen des Mikrophons SoundField SPS200 in folgender Weise: Left Front ( $LF$ ): Kanal 1, Right Front ( $RF$ ): Kanal 2, Left Back ( $LB$ ): Kanal 3 und Right Back ( $RB$ ): Kanal 4. Das hieraus erzeugte B-Format erster Ordnung hat vier Kanäle. Es besteht aus dem ungerichteten Schalldruck  $W$  und den drei Kanälen für die Raumrichtungen  $X$ ,  $Y$  und  $Z$ .

Um die Aufnahme mit einem oder mehreren Lautsprechern auszugeben, muss diese für die Wiedergabe dekodiert werden. Darin liegt einer der Vorteile dieses Formates. Es ist nicht kanalbasiert und auf keine Lautsprecheranordnung oder sogar Wiedergabegeräte beschränkt. Dadurch ist eine freie Positionierung der Lautsprecher und so die optimale Anpassung an den Wiedergaberaum möglich.

Die Signale eines einzelnen Ausgabekanals können nach der für  $N \geq 4$  Wiedergabekanäle gültigen Gleichung 3.3 berechnet werden [13].

$$\begin{aligned}
 w_i &= \frac{\sqrt{2}}{N} \\
 x_i &= \frac{3 \cdot \cos \phi \cdot \cos \delta}{N} \\
 y_i &= \frac{3 \cdot \sin \phi \cdot \cos \delta}{N} \\
 z_i &= \frac{3 \cdot \sin \delta}{N}
 \end{aligned} \tag{3.3}$$

$$\text{Lautsprecher}(\phi, \delta) = w_i \cdot W + x_i \cdot X + y_i \cdot Y + z_i \cdot Z$$

Ziel ist es, das Schallfeld und so das aufgenommene Hörerlebnis wiederzugeben. Hierzu wird der Horizontalwinkel  $\phi$  und der Höhenwinkel  $\delta$  relativ zur Hörerposition benötigt. Dieser Zusammenhang ist mit einem Lautsprecher und einem Hörer und den beteiligten Winkeln dargestellt (Abb. 3.6).

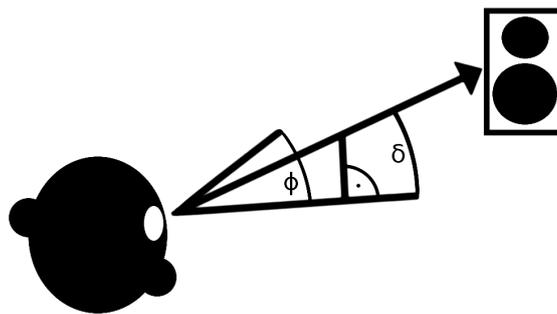


Abbildung 3.6: Zusammenhang von Hörerposition, Lautsprecher und beteiligten Winkeln

Das Verfahren lässt sich mit HRTF-Filterung kombinieren, um mit binauraler Wiedergabe eine spezifische Lautsprecheranordnung zu simulieren. Zuerst wird die Aufnahme im A-Format erstellt, anschließend wird dieses in das B-Format gewandelt. Im nächsten Schritt werden die vier Kanäle des Quadrophonieaufbaus dekodiert und abschließend, mit der passenden HRTF-Filterung versehen, auf einem Kopfhörer wiedergegeben (Abb. 3.7).



Abbildung 3.7: Umwandlung aus dem A-Format in binaurale Quadrophonie

Das Verfahren wurde testweise implementiert, um dessen Verwendbarkeit zu prüfen.

Mit dem Mikrofon SoundField SPS200 wurde eine A-Format-Aufnahme erstellt. Eine dauerhaft sprechende Person umrundete das Mikrofon. Das GNU-Octave-Script *ambisonics.m* implementiert die Formeln 3.2 und 3.3 und kombiniert dies mit HRTF-Filterung und erzeugt so virtuelle Lautsprecher für eine Quadrophonieanordnung (Abb. 3.8).

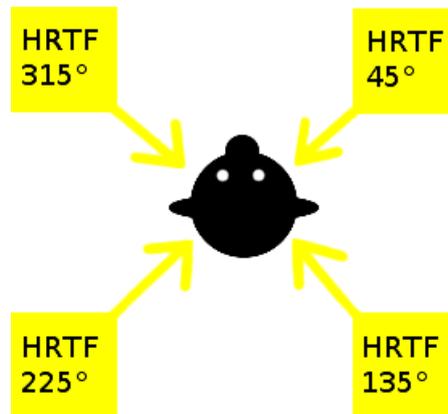


Abbildung 3.8: Virtuelle Quadrophonie zur Verwendung von Ambisonics

Der bei Wiedergabe mit einem Stereokopfhörer erzeugte Eindruck war zufriedenstellend. Das Ergebnis wirkte natürlich und die Bewegung der sprechenden Person konnte nachvollzogen werden. Es handelt sich nur um einen subjektiven Eindruck, da für diesen Versuch keine Befragung durchgeführt wurde. Für lineare VR-Inhalte ist das Verfahren geeignet. Die direkte Aufnahme bei der Inhaltserstellung und die etablierte Werkzeugkette sprechen für dieses Format. Auch die situationsabhängige Wahl der Wiedergabemöglichkeit ist für eine breite Verwendung geeignet.

Durch die interaktiven Handlungsmöglichkeiten bei Simulationen können allerdings keine um die Person zentrierte, fertig abgemischten Audioumgebungen verwendet werden. Hier ist ein objektorientierter Ansatz sinnvoller, welcher die einzelnen Tonquellen als Audio-Objekte abbildet. Diese können so auch direkt per HRTF gefiltert und damit genauer positioniert werden. Die Richtungstreu bei Ambisonics erster Ordnung ist der von HRTF-Filterung unterlegen und erst bei höheren Ordnungen gleichwertig<sup>1</sup>.

Die Möglichkeit alle Raumrichtungen mit einer mehrkanaligen Aufnahme gleichzeitig zu messen und der bei diesem Verfahren übliche Einsatz von virtuellen Lautsprechern, ist aber für die Realisierung von künstlichem Raumschall von Interesse.

---

<sup>1</sup>Balke, Stefan: Wahrnehmungsbasierte Analyse von Higher-Order-Ambisonics-Signalen bei binauraler Synthese, Leibniz Universität Hannover, Diplomarbeit, 2013, S. 45-46

### 3.4 Künstliche Erzeugung von Raumschall für die Distanzmodellierung

Die bisher vorgestellten Lösungen behandeln die Erzeugung eines Richtungseindrucks. Für die Distanzmodellierung ist die Realisierung des in Abschnitt 2.1.3 ausgewählten Raumschalls nötig. Das Verhältnis von Direkt- zu Raumschall hat den größten Einfluss auf den Distanzeindruck. Ausgangspunkt ist wieder ein trockenes Signal. Durch dessen Bearbeitung soll ein Raumeindruck und so die Distanzwahrnehmung ermöglicht werden. Für die computergestützte Modellierung von Raumschall existiert bereits Software, die z.B. im Bereich Schallschutz verwendet wird. Bei der Planung von Gebäuden, der akustischen Optimierung von Innenräumen, dem Entwurf technischer Geräte oder Schallschutzmaßnahmen, wie z.B. Schallschutzwänden an Autobahnen, müssen Berechnungen zur Schallausbreitung und Reflexion durchgeführt werden. Für Geräte bis hin zur Windkraftanlage wird z.B. die Software ANSYS eingesetzt und die Schallausbreitung und Erzeugung physikalisch für diese simuliert.

Mit einer ähnlichen Methodik ist theoretisch die künstliche Erzeugung von Raumschall für eine VR-Umgebung möglich. Die Räume, Landschaften und Gegenstände liegen bereits als 3D-Modelle vor. Mit der Festlegung von Materialeigenschaften für die Oberflächen und der Abstrahlcharakteristik der Tonquellen, wären die nötigen Informationen vorhanden. In diesem modellbasierten Ansatz müssten die physikalischen Effekte beachtet werden. Das beinhaltet die Schallausbreitung, die winkelhängige Reflexion unter Berücksichtigung der materialspezifischen Absorption und die Überlagerung der Schallwellen in der virtuellen Umgebung.

Obwohl diese Schritte einfachen Regeln folgen, ist die Berechnung durch die Kombination aller beteiligten Aspekte komplex. Bis sich die Reflexionen durch Absorption soweit abgeschwächt haben, dass der Einfluss auf das Gesamtschallbild zu vernachlässigen ist, müssen viele Rechenschritte durchgeführt werden.

Für die Analyse eines Modells, z.B. eines Gebäudes ist eine einmalige Berechnung möglich. Für eine Echtzeitanwendung ist das Vorgehen nicht geeignet. Der erzielte Qualitätsgewinn einer korrekten Simulation gegenüber ressourcensparenden Lösungen rechtfertigt nicht den Berechnungsaufwand.

### 3.4.1 Erzeugung von Raumschall durch Verwendung von Faltungshall

Ein intuitiver Ansatz ist die Verwendung von Faltungshall. Bei Faltungshall handelt es sich um die Faltung einer gemessenen Room Impulse Response (RIR) mit einem trockenen Signal. Die LZI-System-Eigenschaft gilt näherungsweise für das Raumschallsystem, weshalb eine Messung möglich ist. Es wird der Raumschall der gemessenen Umgebung direkt reproduziert und so ein hoher Grad an Realismus erreicht. Genaugenommen wird der Raumschall für eine der möglichen Kombinationen aus Sender und Empfänger gemessen. Für eine realistische Simulation muss eine ausreichende Abdeckung der dabei möglichen Kombinationen vorhanden sein. Dann kann durch Interpolieren für alle Positionen im Raum ein realitätsnaher Raumschall erzeugt werden. Ein Problem ist, dass mit einer üblichen Messung nur für eine Raumrichtung Daten gewonnen werden. Die frühen Reflexionen sind allerdings richtungsabhängig. Hier müsste ein Soundfieldmikrofon eingesetzt werden, um mit einer Messung alle Raumrichtungen abzudecken. Aus dem Ambisonicsformat könnten dann mehrere Kanäle um die Hörerposition herum erzeugt und diese per HRTF-Filterung mit einem Richtungseindruck versehen werden.

Denkbar wäre es auch, direkt mit einem Kunstkopf die Messpositionen abzuarbeiten. Das wäre eine kombinierte Messung von RIR und HRTF, sodass damit eine Binaural Room Impulse Response (BRIR) erhalten wird. Durch die Richtungsabhängigkeit bei der Aufnahme mit einem Kunstkopf müsste für jede der Messpositionen auch eine vollständige Umdrehung des Kunstkopfes durchgeführt werden.

Die beschriebenen Techniken sind für lineare VR-Inhalte, in denen lediglich um einen festen Punkt rotiert wird, verwendbar. Für interaktive Inhalte ist durch die benötigte Anzahl an Messungen für einen einzigen Raum eine andere Lösung zu finden. Hinzu kommt noch ein Problem, welches durch die Nachhallzeit entsteht. Um diese korrekt abzubilden, wird eine entsprechende Filterlänge benötigt. Dies führt entweder zu einem Anstieg der Rechenzeit oder erhöht bei der Verwendung von FFTs die Latenz.

### 3.4.2 Das Spiegelschallquellenmodell als modellbasierter Ansatz

Für die Verwendung in einem interaktiven VR-System wird eine an Situationsänderungen anpassbare und gleichzeitig performante Lösung benötigt. Das Spiegelschallquellenmodell als modellbasierter Ansatz leistet dies. Es basiert auf der zeitlichen Verzögerung und Dämpfung von Kopien des Ausgangssignals [33]. Es lässt sich damit sowohl in Software als auch in Hardware effizient realisieren.

Die mehrfache Spiegelung des zu simulierenden Raums, inklusive der darin enthaltenen Tonquellen, bildet die Grundlage. Das Modell ist deshalb auf rechteckige Räume beschränkt. Bei einer Änderung von Hörer- oder Quellposition muss die Situation aktualisiert und neu berechnet werden. Die Genauigkeit und Vollständigkeit der Raumschallsimulation ist von der Anzahl der durchgeführten Spiegelungen abhängig. Pro durchgeführter Spiegelung aller vorhandenen Räume wird von einer Ordnung gesprochen. Ein Raum mit nur der Direktschallquelle würde einem Spiegelschallquellenmodell der *Ordnung null* entsprechen, durch die Spiegelung würden vier zusätzliche Räume und ein Modell der *ersten Ordnung* entstehen.

Das Muster für die Erzeugung lautet wie folgt: An den Spiegelschallquellräumen der *Ordnung n* werden durch Spiegelung an den vier Kanten die Räume der *Ordnung (n+1)* erzeugt. Die neuen Räume dürfen sich nicht mit bereits vorhanden Räumen überlagern.

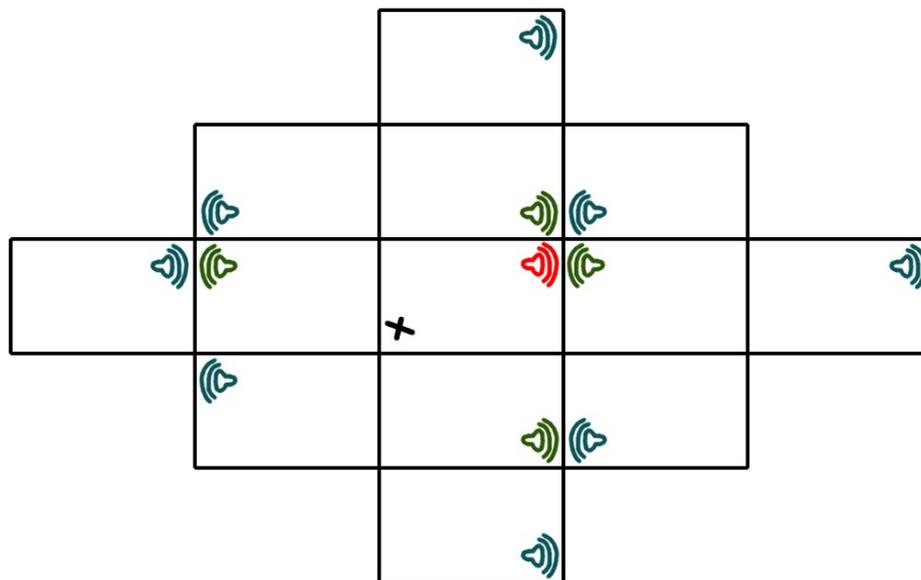


Abbildung 3.9: Verteilung der Spiegelschallquellen im Raum

Das dargestellte Beispiel (Abb. 3.9) zeigt die Situation für die zweite Ordnung. Die Direktschallquelle ist in Rot dargestellt, die Hörerposition wird durch ein  $X$  markiert. Die Spiegelschallquellen der ersten Ordnung sind grün gefärbt und die der zweiten Ordnung haben eine blaue Farbe. An der Position der Spiegelschallquellen im Raum ist die Spiegelung entlang der Kanten erkennbar. Für die erste Spiegelung werden alle vier Räume erzeugt, bei der nächsten Spiegelung wird jeweils ein Raum nicht erzeugt, da an dessen Position der Ursprungsraum liegt.

Anschließend können die Entfernungen der Schallquelle zu der Hörerposition berechnet werden. Von der Distanz sind die Laufzeit und die Lautstärke abhängig. Aus der Laufzeit und der Schallgeschwindigkeit kann für jede Spiegelschallquelle eine Verzögerung berechnet werden. Die Lautstärke ergibt sich aus der entfernungsbedingten Lautstärkeabnahme und der Ordnung der Spiegelschallquelle. Jeder Spiegelschritt entspricht einer Reflexion, weshalb pro Ordnung mit einem Absorptionsfaktor gerechnet wird.

Die Funktionsweise wird an einem Beispiel verdeutlicht. Für dieses werden alle beteiligten Schallquellen als in alle Richtungen abstrahlend angenommen. Diese senden gleichzeitig eine Kopie des Ursprungssignals aus, in diesem Fall handelt es sich um einen Impuls. Bei der Hörerposition ergibt sich dadurch das dargestellte Muster (Abb. 3.10).

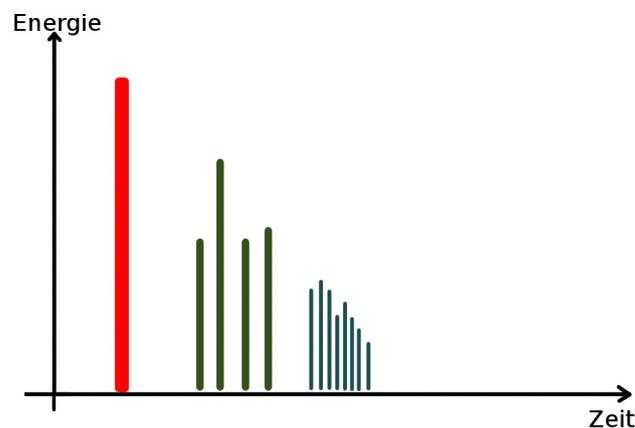


Abbildung 3.10: Zeitlicher Verlauf des durch die Spiegelschallquellen erzeugten Musters

Durch die Entfernung zur Hörerposition treffen die Kopien des Signals verzögert ein. Zuerst trifft der Direktschall (rot) ein, darauf folgen nach der Anfangszeitlücke die frühen Reflexionen (grün) und abschließend der diffuse Nachhall (blau).

Durch diese unkomplizierte Technik werden die für die Distanzeinschätzung wichtigen frühen Reflexionen hinreichend genau modelliert. Inklusive des Nachhalls ist auch das Verhältnis von direktem zu indirektem Schall korrekt. Durch die Modellierung mit gespiegelten Räumen liegen auch die Richtungen der Schallquellen vor. Dies kann mit der HRTF-Filterung kombiniert werden, um den Spiegelschallquellen einen Richtungseindruck zu verleihen. Der Rechenaufwand ist deutlich geringer als bei analytischen Modellen und das Ergebnis ist präziser als z.B. die Ausgabe verbreiteter Effektgeräte [34]. Aus diesem Grund wird für die Distanzmodellierung das Spiegelschallquellenmodell verwendet.

## 4 Zur Zeit verwendete Techniken im Bereich binaurales VR-Audio

In den vorhergehenden Kapiteln wird die kombinierte Nutzung einer HRTF-Filterung mit dem Spiegelschallquellenverfahren zur Erzeugung von VR-Audio begründet. In diesem Kapitel wird der aktuelle Stand bei vergleichbaren Systemen, die Anwendungsmöglichkeiten der Technik und die Fähigkeiten des Vorgängersystems<sup>1</sup> kurz angerissen. Die vorgestellten Lösungen enthalten entweder den Audioanteil von Echtzeit-VR oder sind dafür geeignet als Bestandteil in einem solchen System genutzt zu werden.

Auch in Videospielen, wie z.B. *Hellblade: Senua's Sacrifice*<sup>2</sup>, wird binaurales 3D-Audio eingesetzt. Nicht bei allen VR- oder AR-Anwendungsszenarien spielt der Audioanteil eine Rolle. Bei dem Virtual Product Development (VPD), einer Technik der Produktentwicklung, welche auf virtuellen Modellen basiert, dienen die Anwendungen, wie z.B. Vuforia View, der dreidimensionalen Repräsentation des Objektes.

Zuerst wird ein Anwendungsgebiet aus der Unterhaltungsindustrie mit linearen Inhalten vorgestellt. Als nächstes werden Lösungen betrachtet, die eine Möglichkeit besitzen binaural Raumklang zu erzeugen, aber der Inhaltserstellung dienen. Das Vorgängerprojekt, auf dem viele Überlegungen basieren, wird ebenfalls dargestellt.

### 4.1 Anwendungsbeispiel: Immersive 360°-Videos mit Raumklang

Ein Anwendungsfeld für lineare VR-Inhalte sind 360° und 180° Videos, welche z.B. mit 360°-Kameras erstellt werden. In diesen ist die Beobachterposition fest vorgegeben, der Bildausschnitt kann aber innerhalb des Betrachtungsbereichs variiert werden.

---

<sup>1</sup> Hemmer, David: Eine kopfhörerbasierte, virtuelle Audioumgebung zur räumlichen Positionierung mehrerer Schallquellen mit einer mobilen SoC-Plattform, HAW-Hamburg, Masterthesis, 2013

<sup>2</sup> Informationen Online abrufbar: <https://www.youtube.com/watch?v=gFdPXCzxMg8/>

Die Plattform YouTube VR<sup>3</sup> bietet die Möglichkeit diese Videos mit verschiedenen Endgeräten zu betrachten und stellt ein Format für deren Erstellung bereit.

Diese Spanne der verwendbaren Geräte reicht von normalen Desktop Computern, über Smartphones, bis hin zu den aktuellen VR-Headsets. Während auf dem Desktop die Kameraperspektive über die Maus verändert wird, geschieht dies bei den Smartphones und VR-Headsets über die Sensorik, respektive das Tracking. Durch einen Adapter kann auch ein Smartphone wie ein HMD getragen werden. Da lediglich der ausgewählte Ausschnitt eines Videos wiedergegeben wird, sind die Anforderungen an das Wiedergabegerät deutlich geringer als bei interaktiven Simulationen oder Spielen.

Durch die verschiedenen möglichen Endgeräte ergibt sich auch eine Anzahl möglicher Ausgabegeräte für den Audioanteil. Zusätzlich muss durch die freie Wahl des Bildausschnitts für alle Raumrichtungen eine Geräuschkulisse vorliegen. Hierfür wird Ambisonics der ersten Ordnung oder Ambisonics der ersten Ordnung in der Variante mit Head-Locked Stereo verwendet<sup>4</sup>. Dabei werden die zwei Vorzüge von Ambisonics genutzt:

Erstens ist es von den genutzten Wiedergabegeräten unabhängig, da die Ausgabekanäle berechnet werden und nicht kanalbasiert vorliegen. Zweitens kann auf eine Tonspur mit Informationen über alle Raumrichtungen zurückgegriffen werden.

Es gibt natürlich noch weitere Implementierungen und die dargestellte Lösung des Tochterunternehmens von Google stellt nur ein Beispiel dar. Durch den Aufwand wird deutlich, dass selbst für lineare VR-Inhalte der Audioanteil für relevant genug gehalten wird. Allerdings ist auch zu erkennen, dass die hier gewählte Herangehensweise nicht direkt auf interaktive Inhalte übertragen werden kann.

## 4.2 Hardwarebasierte binaurale Simulation von Raumklangsystemen

Es existiert die Möglichkeit, Raumklangsysteme über Kopfhörer zu erzeugen. Manche Hersteller verkaufen Produkte, bei denen in die Ohrmuscheln der Kopfhörer mehrere Treiber integriert sind. Durch ihre räumliche Anordnung sollen die verschiedenen Kanäle eines 5.1 Systems räumlich korrekt wiedergegeben werden. Dies ist fragwürdig, da bei einem aufgesetztem Kopfhörer das eigene Richtungshören ausgeschaltet ist. Es existieren aber auch Systeme, deren Konzepte wissenschaftlich fundiert und nicht widerlegt sind.

---

<sup>3</sup> Material abrufbar unter: <https://www.youtube.com/vr>

<sup>4</sup> Informationen Online abrufbar: <https://support.google.com/youtube/answer/6395969?hl=de>

Die Bearbeitungsbox Smyth Realiser A8<sup>5</sup> kann ein virtuelles Surroundsystem simulieren. Das System dient der Reproduktion des Eindrucks eines realen Lautsprecheraufbaus in dessen üblicher Umgebung. Es kann z.B. in beengten Studios für die Überprüfung eines Höreindrucks verwendet werden. An die Eingänge wird ein klassisches Surround-Format, wie z.B. 5.1-Surround angeschlossen, mit den Ausgängen wird ein Stereokopfhörer verbunden. Die Eingangssignale für die Lautsprecherpositionen werden mit HRTFs gefiltert, um diese bei der Wiedergabe aus der korrekten Richtung erscheinen zu lassen. Gleichzeitig wird Raumakustik hinzugefügt.

Für die Inbetriebnahme wird das Wiedergabesetup, welches reproduziert werden soll, mit In-Ear-Mikrofonen eingemessen. Hierbei werden HRTF und Raumakustik gleichzeitig gemessen. Da der Kopfhörer frei gewählt werden kann, wird auch eine Messung mit aufgesetztem Kopfhörer durchgeführt, um einen Kopfhörerkompensationsfilter zu erzeugen. Um die Tonquellen bei Kopfbewegungen in einem bestimmten Bereich stabil zu halten, wird Headtracking verwendet. Dieses ist für die sitzende Verwendung mit der Ausrichtung auf einen Monitor dimensioniert und der nutzbare Bereich ist nicht ausreichend für raumausfüllende Interaktion.

Die Kombination aus binauraler Wiedergabe, Headtracking und der Verwendung von HRTF inklusive Raumakustik macht das System erwähnenswert. Für VR ist es nicht geeignet, es ist aber ohnehin für den Studiobetrieb ausgerichtet. Creative Labs bietet eine ähnliche Produktreihe für den HiFi-Enthusiasten-Markt an. Es reproduziert ebenfalls mit HRTFs ein Surroundsetup nutzt allerdings kein Headtracking. Statt mit einer Messung durch In-Ear-Mikrophone werden die HRTFs grafisch mit Aufnahmen des Ohres konfiguriert. Die Technik wird in externe Soundkarten, sowie direkt in Kopfhörer integriert, die z.B. via Bluetooth mit Smartphones verbunden werden können.

In beiden Fällen wird HRTF-Filterung lediglich für ein Panning mit virtuellen Lautsprechern verwendet. Auf diese Weise kann keine hohe Richtungsgenauigkeit oder die Repräsentation von Höhenstufen ermöglicht werden. Die Raumakustik wird zwar in einem der Systeme mitausgegeben, diese wird aber nicht für die Distanzmodellierung verwendet. Beide Systeme sind nicht für die direkte Verwendung geeignet, sie zeigen aber, dass hardwarebeschleunigte HRTF-Filterung in kommerziellen Produkten erfolgreich eingesetzt werden kann.

---

<sup>5</sup> Spezifikationen abrufbar unter: <https://smyth-research.com/a8/>

### 4.3 Binaurale 3D-Audiosoftware: VST-Plugin und virtuelle Soundkarte

Im Bereich der professionellen Audioproduktion für Musik, Filme und (VR-)Videospiele/Inhalte wird Spezialsoftware verwendet, welche als Digital Audio Workstation (DAW) bezeichnet wird. Der Funktionsumfang dieser Softwarelösungen zur Audiotbearbeitung kann über VST-Plugin erweitert werden. Bei Virtual Studio Technology (VST) handelt es sich um eine proprietäre Programmierschnittstelle des Hamburger Unternehmens Steinberg Media Technologies GmbH. Mit dieser Schnittstelle können in die DAWs neben Effektgeräten auch komplexere Schritte zur Audiotbearbeitung integriert werden.

Bei dem VST-Plugin *New Audio Technology Spatial Audio Designer*<sup>6</sup> der Firma New Audio Technology handelt es sich um eine Toolbox für die Erstellung von räumlich wahrnehmbarem Audio. Neben den Postproduktionsschritten für die Erstellung klassischer Raumklangformate, wie z.B. 5.1-Surround, wird auch die binaurale Ausgabe von 3D-Audio ermöglicht. Neben den bisher vorgestellten Systemen, welche hardwarebasiert sind, handelt es sich hierbei um eine Softwarelösung. Das Plugin dient der Erstellung von Inhalten und ist nicht für Echtzeitsimulationen ausgelegt. Für die Wiedergabe werden aber Lautsprecher setups in verschiedenen Umgebungen in Echtzeit simuliert.

Diese Echtzeitfunktionalität kann auch separat als Software-Soundkarte *Spatial Sound Card*<sup>7</sup> verwendet werden. Diese stellt sich gegenüber der Anwendungssoftware, wie z.B. Medienabspielern oder Videospiele, als Mehrkanalgerät dar, gibt das Tonsignal aber als bearbeitetes Stereosignal an die physische Soundkarte weiter. In dem Bearbeitungsschritt wird per Software eine HRTF-Filterung durchgeführt und so der Eindruck einer Surround-Lautsprecheranordnung erzeugt. Die virtuelle Umgebung, in der die Wiedergabe stattfindet, kann parametrisiert werden.

Eine Distanzsimulation ist allerdings genauso wie eine dynamisch an die Audioinhalte angepasste Raumschallsimulation nicht möglich. Das Richtungshören basiert weiterhin auf Panning zwischen den virtuellen Lautsprechern. Dies ist hier, wie auch bei den vorgestellten Hardwarelösungen, fehlenden Standards für das objektbasierte 3D-Audio geschuldet: Zurzeit ist eine binaurale Wiedergabe, welche nicht auf Kanälen basiert, nur bei direkter Integration in die 3D-Engine, das Grundgerüst der Multimedia-Anwendung, möglich.

---

<sup>6</sup> Informationen abrufbar unter: 3D-Audio <https://newaudiotechology.com/products/spatial-audio-designer/>

<sup>7</sup> Informationen abrufbar unter: <https://newaudiotechology.com/products/spatial-sound-card-2/>

## 4.4 Das Vorgängerprojekt "Augmented-Reality-Audio-SOC" als Ausgangspunkt

Das Vorgängerprojekt ist die Abschlussarbeit von David Hemmer, zu welcher auch eine Veröffentlichung existiert [8]. Es handelt sich um ein verteiltes System zur Echtzeiterzeugung von 3D-Audio aus Monotonquellen durch HRTF-Filterung. Es ist in eine in Java geschriebene Serverkomponente und ein FPGA mit integrierten DSPs für die Berechnungen aufgeteilt. Es verfügt über angeschlossene Sensorik für Headtracking. Die grundlegende Methodik ist mit [20] und die Realisierung mit [28] vergleichbar. Die Aufteilung in verschiedene Komponenten und die Nutzung spezieller Recheneinheiten wird bei dem für diese Studie entwickelten Modell wiederaufgenommen.

Auf einem FPGA vom Typ XILINX Virtex5 ML 507 werden alle Echtzeitberechnungen durchgeführt. An dieses kann direkt über einen Klinkestecker ein Kopfhörer angeschlossen werden. An das FPGA ist als Sensor zur Ausrichtungsbestimmung ein Magnetkompass der Firma Oceanserver über die RS232 Schnittstelle angeschlossen. Zur Positionierung der Tonquellen existiert eine iPhone-App. Die Kommunikation mit dieser und dem zentralen Java-Server wird über den WLAN-Adapter Netgear WNCE3001 ermöglicht.

Dieser WLAN-Adapter ist an die Ethernetschnittstelle des FPGAs angeschlossen. Der physische Aufbau des Systems mit seinen verschiedenen Komponenten ist in Abbildung 4.1 dargestellt.

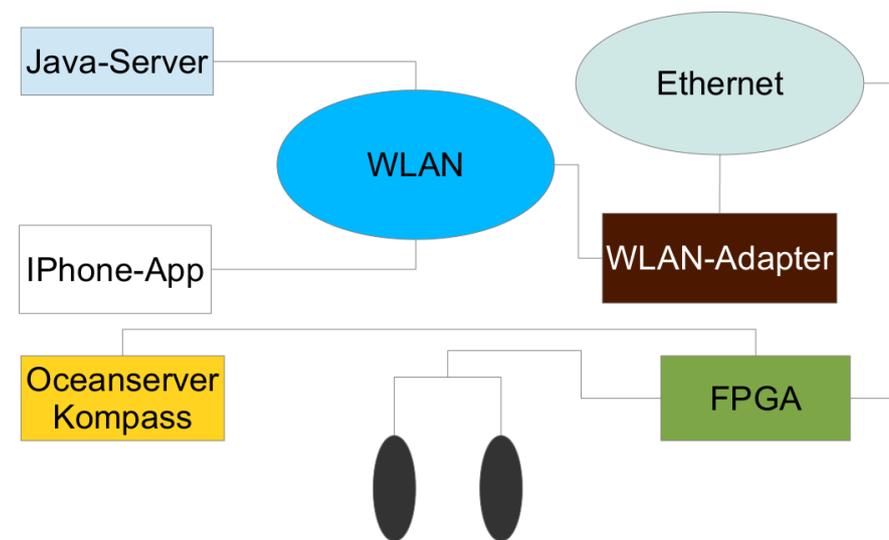


Abbildung 4.1: Physikalischer Aufbau des ARA-Systems

Auf dem FPGA befindet sich eine MicroBlaze-Central Processing Unit (CPU). Die CPU startet die Berechnungen der HRTF und füllt den Puffer des Audio Codec '97 (AC97)-Soundchips. Die Ethernetschnittstelle wird von dieser angesteuert. Die eigentlichen Filteroperationen führen die auf dem FPGA vorhandenen DSP durch. Eine Algorithmic State Machine (ASM) unterstützt sie bei der bilinearen Interpolation, da für jede der acht virtuellen Tonquellen für jede beliebige Richtung die Koeffizienten der HRTF unterschiedlich sind und nur als Stützstellen vorliegen. Die in der Arbeit verwendeten Filterkoeffizientensätze sind in einer vorhergehenden Arbeit<sup>8</sup> erstellt worden.

Die verschiedenen Soft- und Hardwarekomponenten arbeiten über verschiedene Schnittstellen zusammen. Dies ist in Abbildung 4.2 dargestellt.

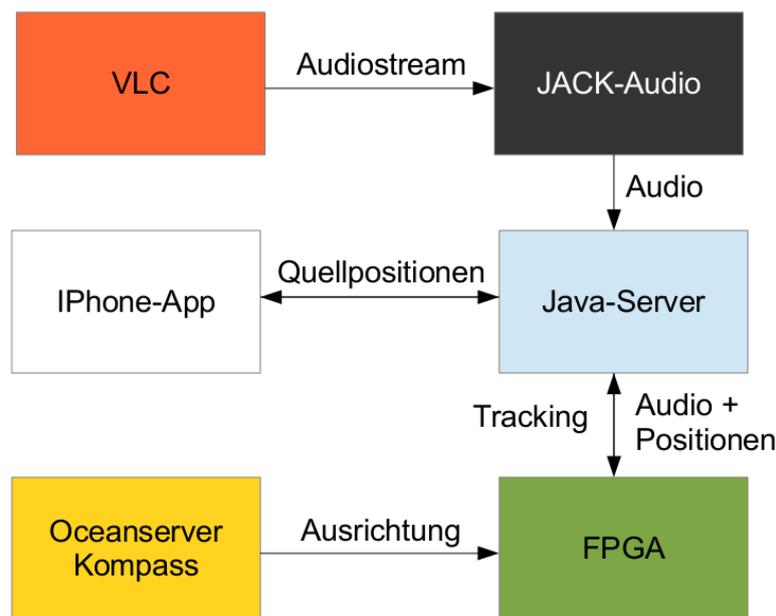


Abbildung 4.2: Komponentenkommunikation

Der Java-Server bekommt über JACK Audio Connection Kit (JACK) ein Audiosignal von der Abspielsoftware. JACK ist eine Software zur Verwaltung von Audioströmen. Der Server zeigt in der Graphical User Interface (GUI) die Ausrichtung des Kopfhörers und die Positionen der Tonquellen an, welche über das Open Sound Control Protocol (OSC-Protocol) in einer iPhone-App bewegt werden. Der Server sendet die Zielpositionen und die Audiosignale über ein proprietäres Protokoll zum FPGA und erhält wiederum die Ausrichtungsdaten des Oceanserver-Kompasses.

---

<sup>8</sup> Sima, Sylvia: HRTF Measurements and Filter Design for a Headphone-Based 3D-Audio System, HAW-Hamburg, Bachelorthesis, 2008

# 5 Aufbau und Auswertung von Studien im Audiobereich

In einer Benutzerstudie sollen qualitative und quantitative Aussagen zu dem gewählten Ansatz mit der Kombination aus HRTF und Spiegelschallquellenmodell gewonnen werden. Es soll untersucht werden, ob das Modell einen Distanzeindruck erzeugt, wie präzise sich dadurch virtuelle Tonquellen lokalisieren lassen und wie es sich auf die anderen Aspekte des räumlichen Hörens auswirkt.

Die allgemeinen Grundlagen für wissenschaftliche Studien werden zuerst beschrieben. Hier werden auch die für Studien im Audiobereich typischen Untersuchungsaspekte und Probleme diskutiert. Anschließend werden wichtige Begriffe aus dem Bereich Statistik eingeführt, welche für die Auswertung benötigt werden.

## 5.1 Theoretische Grundlagen zur Studienrealisierung

Ziel der Durchführung einer Studie ist der Erkenntnisgewinn. Mit ihrer Hilfe soll eine Fragestellung beantwortet, eine belastbare Aussage zu einem Thema ermöglicht oder das Ergebnis einer anderen Forschungsgruppe verifiziert oder falsifiziert werden. Im ersten Schritt müssen zu diesem Zweck verwertbare Daten durch Befragungen, Messreihen etc. erlangt werden. Damit diese Daten aussagekräftig sind, müssen einige Randbedingungen bei der Durchführung beachtet werden.

Eine wichtige Voraussetzung ist eine ausreichend hohe Anzahl an Versuchspersonen. Je größer die Anzahl der teilnehmenden Personen ist, desto geringer wirken sich einzelne Ausreißer aus. Die Zunahme der Genauigkeit steht jedoch nicht in einem linearen Zusammenhang mit der Anzahl der Personen. Schon bei einem relativ kleinen Anteil an befragten Personen sind ausreichend genaue Aussagen über die gesamte Gruppe möglich, aus der die Teilnehmenden stammen. Dies kann an dem Beispiel einer Wahlumfrage verdeutlicht werden.

Um Aussagen über das Abstimmungsverhalten in einem Abstimmungskörper, wie z.B. der Bevölkerung eines Bundeslandes oder einer Stadt, zu treffen, ist eine Befragung im Bereich von tausend Personen für eine erfolgreiche Prognose ausreichend.

Die durchzuführende Studie wird selbstverständlich nicht mit tausenden von Teilnehmern durchgeführt und hat daher eher den Charakter einer vorbereitenden Studie. Es werden grundlegende, qualitative Aussagen zur untersuchten Materie geliefert. Die geringe Probandenzahl führt aber zu einem relativ großen Fehler bei qualitativen Aussagen.

Neben der Anzahl ist auch das Auswahlverfahren für Testpersonen relevant. Diese Auswahl sollte möglichst zufällig geschehen, ansonsten wird schon durch die Auswahl der befragten Personen das Ergebnis unwillentlich in eine Richtung beeinflusst. Eine nachweisbar zufällige Auswahl der teilnehmenden Personen für kleinere Studien ist jedoch schwierig. Die Auswahl der Personen für die Studie wird deshalb so gestaltet, dass diese möglichst verschiedene Hintergründe haben. Hintergrund bedeutet in diesem Fall, dass es sich nicht nur um Studierende handelt, welche sich stark mit dem Thema Audio beschäftigen, sondern auch andere Studienrichtungen, Berufs- und Altersgruppen an der Befragung teilnehmen. Die für die Studie relevanten Unterschiede bei den Personen werden erfasst, sodass Zusammenhänge zwischen dem Abschneiden in bestimmten Teilaufgaben der Studie und den erfassten Eigenschaften hergestellt werden können.

Dabei muss darauf geachtet werden, dass das Zusammenfallen einer Beobachtung mit einer Auffälligkeit bei einer Gruppe von Testpersonen mit einer bestimmten gemeinsamen Eigenschaft nicht zwingend bedeutet, dass das beobachtete Ergebnis durch die Gemeinsamkeit ausgelöst wurde. Es ist eher ein Hinweis auf einen eventuellen Zusammenhang. Es muss untersucht werden, ob das auftretende Ereignis, z.B. eine Beobachtung, tatsächlich in einem Wirkzusammenhang mit der damit korrelierten Eingangsgröße liegt. Ansonsten liegt eine Scheinkausalität vor [22]. Die Auffälligkeit kann schließlich auch durch eine nicht beachtete Gemeinsamkeit oder nur zufällig zustande gekommen sein. Ein Beispiel hierfür ist eine Korrelation von hoher Lebenserwartung und der Ausübung von Reitsport. Das eigentliche Reiten ist nicht besonders gesundheitsfördernd, es ist aber nur denen möglich, die einer höheren sozialen Schicht angehören. Die höhere soziale Schicht ermöglicht eine bessere Gesundheitsversorgung und Arbeitsbedingungen, was zu der beobachteten höheren Lebenserwartung führt.

Ein Begriff aus Soziologie und Psychologie ist der des Bias. Das Bias beschreibt eine Voreingenommenheit gegenüber einem Sachverhalt, einem Vorgang oder einer Sache, die dazu führt, dass die gleiche Situation von verschiedenen Personen unterschiedlich wahrgenommen und interpretiert wird. Diese Voreingenommenheit kann z.B. durch Sozialisation, gesammelte Erfahrungen und vorhandenes Vorwissen entstehen.

In der Studie äußert sich dies an den teilnehmenden Personen, welche sich intensiv mit dem Thema 3D-Audio beschäftigt haben. Diese Gruppe achtet vermutlich auf den Einsatz üblicher Techniken oder antizipiert bestimmte Vorgehensweisen in der Studie und könnte so merklich andere Ergebnisse als die übrigen Teilnehmer liefern. Das Auftreten solch einer Verzerrung der Wahrnehmung lässt sich nicht verhindern. Durch eine Erfassung von Merkmalen der Versuchspersonen, wie z.B. dem Vorwissen, soll das Problem sichtbar gemacht und bei der Auswertung der Studie kompensiert werden.

Ähnliche Auswirkungen hat die Autosuggestion. Es handelt sich dabei um eine Erwartungshaltung, die sich selbst erfüllt. Für den Versuch bedeutet dies, dass eine Behauptung über eine Sachlage, die dem Versuch vorangestellt wird, diesen dahingehend beeinflusst. Die Behauptung, dass eine Technik zur Distanzmodellierung eingesetzt wird, kann dazu führen, dass diese von den teilnehmenden Personen auch bei unbearbeitetem Audio als vorhanden angegeben wird.

Die Auswirkung von Autosuggestion als auch das Auftreten von durch die Studie erzeugtem Bias lassen sich reduzieren, wenn den teilnehmenden Personen möglichst wenig, d.h. nur die nötigen Informationen für den Versuch vorliegen. Im medizinischen Bereich hat sich deshalb die Durchführung von Doppelblindtests als Lösung etabliert [3].

In einem einfachen Blindtest wissen die teilnehmenden Personen nicht, ob diese zur Kontroll- oder Experimentalgruppe gehören, ob sie also z.B. ein Placebo oder den experimentellen Wirkstoff erhalten haben. Dadurch wird der subjektiven Erwartungshaltung der Proband\_Innen als Einflussfaktor entgegengewirkt. Das Präparat kann als wirksam gelten, wenn es sich als wirksamer als das Placebo erwiesen hat.

Der Versuchsleitung ist in diesem Fall bekannt, wer welches Präparat erhalten hat. Dies kann dazu führen, dass die Patienten unterschiedlich behandelt werden oder bei der Diagnose unbewusst in eine Richtung gearbeitet wird. Aus diesem Grund kann der doppelte Blindtest eingesetzt werden. Bei dieser haben auch die Personen aus der Versuchsdurchführung bis zur Auswertung der Studie keine Kenntnis zu welcher Versuchsgruppe die Proband\_Innen gehören. Auf diese Weise werden alle Testpersonen sicher gleich behandelt und auch die Auswertung kann unvoreingenommen durchgeführt werden.

Auf die Audio-Studie übertragen heißt dies, dass die Teilnehmer nicht wissen ob z.B. in dem Testfall der Raumschall hinzugefügt wurde oder keine Distanzmodellierung eingesetzt wird. Sie erhalten lediglich die Hörprobe und die Möglichkeit den bei sich erzielten Höreindruck zu notieren. Die Bearbeitung der Testfälle findet ohne Interaktion statt und die Auswertung ist größtenteils automatisiert. Das entspricht nicht vollständig einem Doppelblindtest, greift aber die Grundideen auf, um die Ergebnisqualität zu verbessern.

## Herausgestellte Untersuchungsaspekte im Audiobereich

Für Studien im Audiobereich existieren übliche Untersuchungsaspekte und Vorgehensweisen. Für die subjektiven Aspekte eignet sich eine Likert-Skala zur Zustimmung.

**Auflösungsvermögen:** Das Auflösungsvermögen beschreibt die Fähigkeit, die Position von zwei Tonquellen zu unterscheiden. Dabei kann es sich um eine Abweichung der Richtung, der Entfernung oder eine Kombination aus beidem handeln. Durch einen Test, ob sich zwei Geräusche an derselben Stelle befinden, kann das Auflösungsvermögen objektiv bestimmt werden. Durch Variation der Abweichung nähert man sich dem Auflösungsvermögen der Testperson.

**Azimet-/ Elevationsfehler:** Auch den Fehler bei der Bestimmung der Richtungs- und des Höhenwinkels kann man objektiv bestimmen. Hierzu werden verschiedene Testgeräusche und eine Möglichkeit zur Richtungsangabe verwendet.

**Externalisierung:** Externalisierung beschreibt den Eindruck, dass ein Geräusch von außerhalb des eigenen Kopfes kommt. Ungefilterte Signale wirken bei Kopfhörernutzung so, als ob sich diese in der eigenen Kopfmitte befinden. Da es sich um eine subjektive Wahrnehmung handelt, kann mit direkten Fragen zu dem Thema gearbeitet werden. Implizit kann auch über eine korrekte Richtungs- oder Entfernungsangabe ohne explizite Thematisierung eine Aussage über die Externalisierung getroffen werden. Wenn der Tonquelle erfolgreich eine Richtung oder Entfernung zugeordnet wurde, dann kann diese nicht in der Kopfmitte gehört werden.

**Natürlichkeit:** Ob ein Höreindruck als natürlich wahrgenommen wird, ist ebenfalls subjektiv. Menschen sind verschiedene Geräusche und Hörsituationen, wie z.B. die Auswirkung der Umgebung auf diese Geräusche, gewöhnt. Auf diese Weise liegen Referenzwerte für die Nutzung in Vergleichstests vor. Die Verwendung verschiedener bekannter Geräusche und die Variation der Umgebung durch Raumschall können als Testfälle zur Überprüfung dieser Eigenschaft genutzt werden.

**Hinten-Vorne-Fehler:** Ohne Kopfbewegungen ist die Entscheidung, ob sich ein Geräusch vor oder hinter der Person befindet, mit einem Fehler behaftet. Mit Spiegelung von Geräuschquellen über eine zwischen den Ohren gedachten Achse und die Verwendung beider Positionen mit identischen Randbedingungen kann dieser Punkt ausgewertet werden.

### 5.1.1 Statistische Auswertung einer Studie im Audiobereich

Nach der Gewinnung der Daten mit der oben dargestellten Methodik in einer Studie, müssen die Ergebnisse der Teilversuche jeder einzelnen Personen ausgewertet werden. Dazu werden die Einzelergebnisse mit Hilfe von Mitteln der Statistik sinnvoll zusammengefasst und aufbereitet. Dadurch wird ihre Interpretation und das Treffen einer Aussage über den Untersuchungsgegenstand möglich. Eine grundlegende Methode ist die Bildung des arithmetischen Mittelwertes  $\bar{x}$  mit Formel 5.1. So kann das Versuchsergebnis von vielen Versuchsdurchläufen in einer einzelnen Zahl zusammengefasst werden.

$$\text{Mittelwert} = \bar{x} = \frac{1}{n} \sum_{i=0}^n x_i \quad (5.1)$$

Dazu ist ebenfalls der Median geeignet. Bei dem Median handelt es sich um einen Wert aus der Mitte der Messreihe. 50% der übrigen Werte sind größer und 50% der übrigen Werte sind kleiner als dieser. Im Gegensatz zum arithmetischen Mittel ist es ein real existierender Wert und nicht eine aus den Werten berechnete künstliche Größe.

Um die Aussagekraft des Mittelwertes zu beurteilen, wird zusätzlich die Angabe der durchschnittlichen Abweichung von diesem benötigt, die Standardabweichung  $\sigma$ . Diese kann nicht direkt berechnet werden. Zuerst muss die Varianz  $\sigma^2$ , die mittlere quadratische Abweichung, nach der Formel 5.2 bestimmt werden.

$$\text{Varianz} = \sigma^2 = \sum_{i \geq 1}^n (x_i - \bar{x})^2 * p_i \quad (5.2)$$

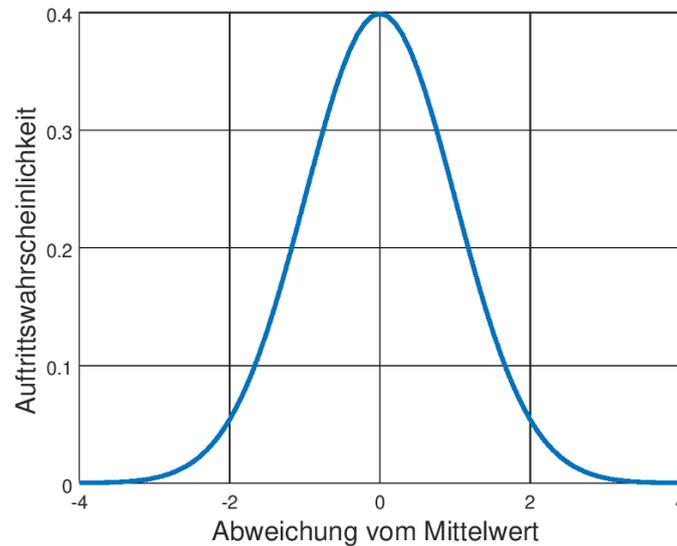
Wobei es sich bei  $x_i$  um einen einzelnen Messpunkt, bei  $p_i$  um dessen Auftrittswahrscheinlichkeit und bei  $\bar{x}$  um den Mittelwert handelt, der hier statt des Erwartungswertes eingesetzt wird. Durch das Ziehen der Wurzel in 5.3 wird aus der Varianz die Standardabweichung berechnet.

$$\text{Standardabweichung} = \sigma = \sqrt{\sigma^2} \quad (5.3)$$

Streuen die Ergebnisse stark, dann ist die Standardabweichung groß, ist diese klein, dann entsprechen die einzelnen Messungen dem Durchschnittswert größtenteils. Zusammen mit dem Maximum und Minimum als größten und kleinstem Wert lassen sich so Ergebnisse mit dem Mittelwert interpretieren.

Gerade bei einer großen Probenzahl tritt bei statistischer Auswertung eine charakteristische Verteilung der Abweichung der Messwerte vom Erwartungswert auf. Diese wird als Normal- oder Gauß-Verteilung bezeichnet (Diagr. 1).

Diagramm 1: GNU-Octave-Plot der Normalverteilung



Der Erwartungswert befindet sich bei  $x = 0$ . Auf der X-Achse ist die Abweichung vom Erwartungswert eingetragen und auf der Y-Achse ist die entsprechende Auftrittswahrscheinlichkeit abzulesen. Die Normalverteilung ist von der Varianz abhängig, das Diagramm zeigt den Verlauf für  $\sigma^2 = 1$ . Durch Veränderung der Varianz wird die Kurve gestreckt oder gestaucht, mit dem Erwartungswert wird diese auf der X-Achse verschoben. Ein Beispiel für eine normalverteilte Größe ist die Körperhöhe von Menschen, jeweils für das biologische Geschlecht separiert.

In Studien wird mit Stichproben versucht, eine Aussage über die Grundgesamtheit zu treffen, aus der die Stichprobe stammt. Mit Hilfe der Signifikanz kann bestimmt werden, wie wahrscheinlich es ist, ob eine Beobachtung auf die Grundgesamtheit übertragbar ist oder ob es sich nur um eine zufällige Schwankung in der Stichprobe handelt. Dazu wird ein Signifikanzniveau festgelegt, unter dem sich die Irrtumswahrscheinlichkeit, der  $p$ -Wert, befinden soll. Häufig wird ein Signifikanzniveau von  $\alpha = 0,05$  vorausgesetzt. In diesem Fall ist das Ergebnis mit 95 %iger Wahrscheinlichkeit keine zufällige Schwankung. Als Nullhypothese kann die Annahme verwendet werden, dass es sich um eine zufällige Schwankung handelt. Anschließend wird ein Hypothesentest genutzt, um diese abzulehnen und den  $p$ -Wert zu berechnen. Ist der  $p$ -Wert kleiner als das nötige Signifikanzniveau, dann wird von einem signifikanten und aussagekräftigen Ergebnis gesprochen.

## 6 Durchführung einer Studie zur Bewertung der Distanzmodellierung

Die Wirksamkeit der ausgewählten Modellierung auditiver Distanz über ein Spiegelschallquellenmodell in einem binauralen 3D-Audio-System wird in einer Studie untersucht. Die Studie wird auf Basis der Grundlagen aus Abschnitt 5 durchgeführt. Es wird überprüft, ob sich dieser Ansatz für den Einsatz in VR-Systemen eignet und welche weiteren Effekte zu beobachten sind. Für die Studie wird ein selbst erstelltes 3D-Audio-Modell verwendet. Das System ist so entworfen, dass die verwendeten Algorithmen und die Aufteilung der verschiedenen beteiligten Komponenten auch in einer Echtzeitimplementierung eingesetzt werden könnten. Dieses Modell wird verwendet, um aus trockenen Tonaufnahmen die einzelnen Audioanteile für die Testfälle zu generieren. Diese Audioabschnitte werden mit Bildinhalten zu einem Video kombiniert und den Proband\_Innen vorgespielt. Dabei füllen diese einen Fragebogen in Form eines PDF-Formulars aus. Die Auswertung erfolgt anschließend in der Programmiersprache R. Bei der Durchführung mussten Einschränkungen durch die Maßnahmen gegen die Corona-Pandemie beachtet werden.

### 6.1 Erstellung des 3D-Audio-Modells für die Studie

Die Softwareraumschallsimulation erzeugt aus trocken aufgenommenen Monosignalen ein Stereosignal, um auf einem Kopfhörersystem binaurales 3D-Audio zu ermöglichen. Dieses Stereosignal enthält durch HRTF-Filterung Richtungsinformationen und durch den Einsatz eines Spiegelschallquellenmodells und Anpassung der Lautstärke Distanzinformationen. Durch diese Kombination wird eine räumlich ortbare, virtuelle Tonquelle erzeugt. Für die Untersuchung werden Koeffizienten für die HRTF-Filterung eingesetzt, welche in einer vorhergehenden Messung erstellt wurden. Die Software ist zweigeteilt und in den Programmiersprachen C sowie GNU-Octave geschrieben.

### 6.1.1 Fähigkeiten der entwickelten Softwareraumschallsimulation

Die Software liest eine Einkanal-Wave-Datei ein und wandelt diese in ein Zweikanal-Signal für die binaurale Wiedergabe um. Im einfachsten Fall wird nur eine HRTF-Filterung mit einer Anpassung der Lautstärke für die vorgegebene Position durchgeführt. Das so erzeugte Signal entspricht dem isolierten Direktschall einer Tonquelle in der Realität. Hauptziel ist jedoch das Hinzufügen von Raumschall, um einen Distanzeindruck und Externalisierung zu erzeugen. Dazu wird ein Spiegelschallquellenmodell eingesetzt, dieses wird in Abschnitt 3.4.2 beschrieben. Die maximale Ordnung der Spiegelschallquellen und damit die Genauigkeit des Modells sind wählbar. Alle Schallquellen werden als ungerichtete Punktschallquellen angenommen. Die Signale der einzelnen Spiegelschallquellen werden abschließend durch jeweils eigene Filter verarbeitet, welche per MLS-Koeffizienten eine Korrelation und damit Artefakte durch Überlagerung verhindern.

Es werden zwei Methoden zur Wiedergabe der Spiegelschallquellen realisiert:

Die erste Variante ist das präzise, aber filterungsintensive direkte Modell. Bei diesem werden die Spiegelschallquellen einzeln mit einer richtungsgebenden HRTF-Filterung versehen (Abb. 6.1).

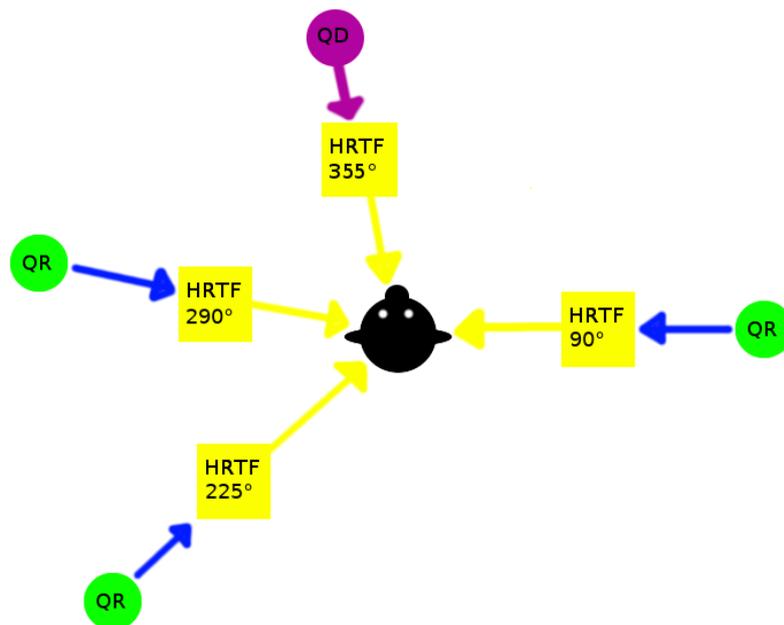


Abbildung 6.1: Direkte Filterung von Direktschall und Spiegelschallquellen

Der Direktschallanteil  $QD$  und die Spiegelschallquellen  $QR$  werden im direkten Modell jeweils separat für die vorgegebene Richtung gefiltert.

Die zweite Variante ist weniger präzise und basiert auf einer geringeren Anzahl von Filterungsvorgängen und dem Einsatz von Panning. Der eingesetzte Algorithmus erweitert hierzu das bekannte Stereopanning von zwei auf vier Lautsprecher und entspricht damit Vector Base Amplitude Panning (VBAP) [23]. Um die Hörerposition werden für jede der Raumrichtungen um  $90^\circ$  versetzt vier virtuelle Lautsprecher aufgestellt (Abb. 6.2). Liegt die Tonquelle exakt auf der Richtung des Lautsprechers, so wird auf diesem das Signal zu 100 % wiedergegeben. Steht die Quelle im Uhrzeigersinn gesehen zwischen dem vorherigen und dem nächsten Lautsprecher, so wird das Signal gleichmäßig aufgeteilt. Es sind auch Abstufungen, wie z.B. das Verhältnis von 30 % zu 70 %, möglich. Wenn  $+90^\circ$  erreicht werden, wird nur noch der neue Lautsprecher eingesetzt. Wird weiter gedreht, findet die Aufteilung des Signals zwischen den nächsten Lautsprechern auf der Kreisbahn um die Hörerposition statt.

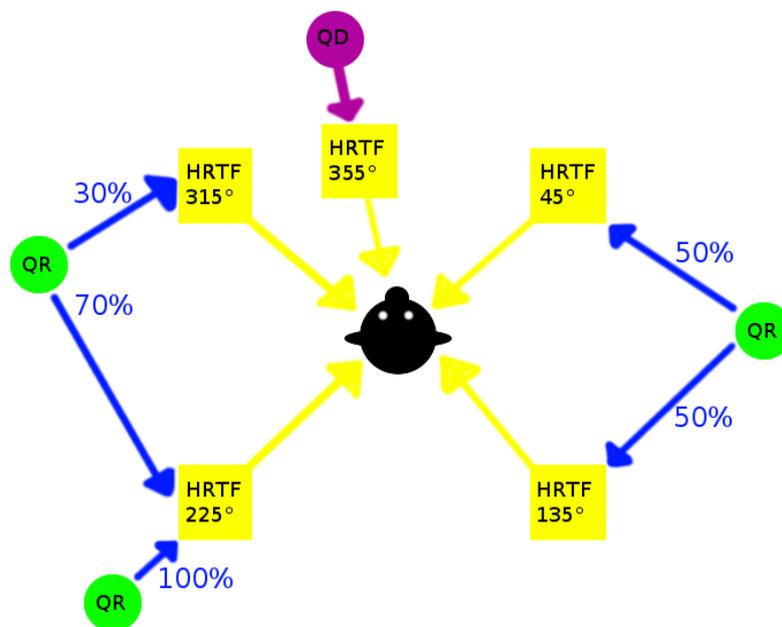


Abbildung 6.2: Panningbasierte Wiedergabe der Spiegelschallquellen

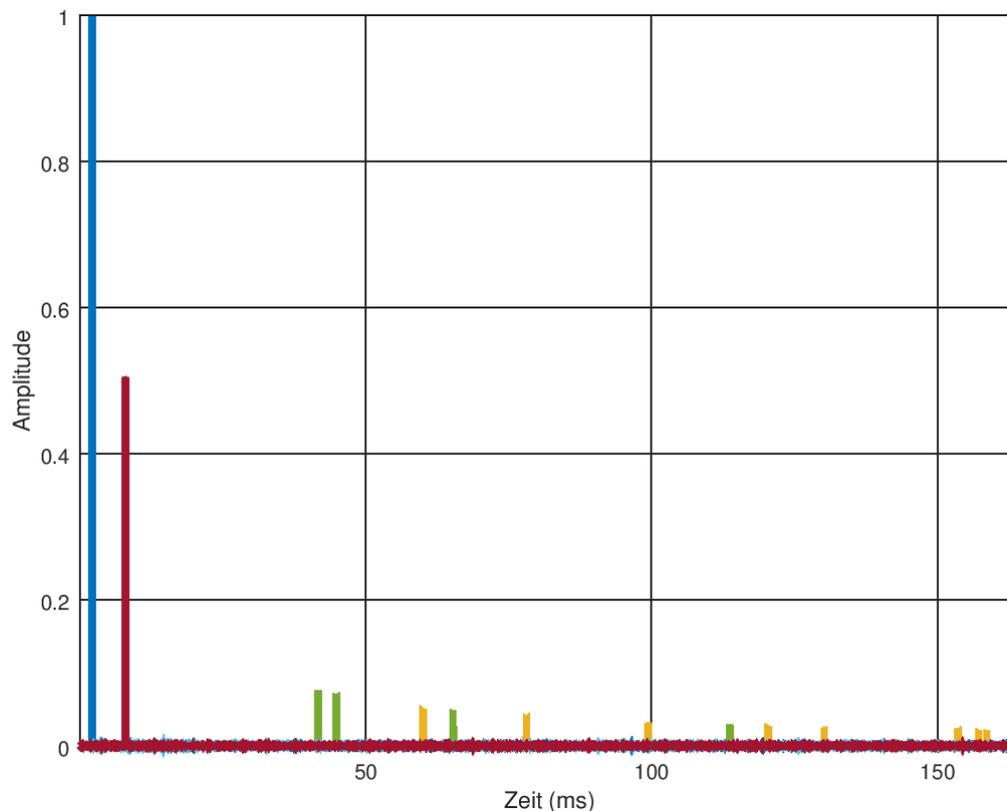
Im panningbasierten Modell werden die Spiegelschallquellen  $QR$  je nach Position auf die virtuellen Lautsprecher (HRTF  $45^\circ$ , HRTF  $135^\circ$ , HRTF  $225^\circ$  und HRTF  $315^\circ$ ) verteilt. Der Direktschallanteil  $QD$  wird wieder direkt gefiltert.

Es ist möglich, neben ruhenden Tonquellen auch bewegliche Tonquellen einzusetzen. Die Bewegungsmuster einer Rechteckfahrt um die aktuelle Position und eine Fahrt von einer Position zur nächsten sind möglich. Die Angabe der Position der Tonquelle, der Wegstellen und die Position der hörenden Person werden relativ zu dem virtuellen Raum aus der

Draufsicht angegeben. Auf diese Weise ist es möglich, verschiedene Beobachterpositionen für die im Raum stehenden oder sich bewegenden Quellen einzustellen.

Die Wirkungsweise der Software wird an einem Beispiel veranschaulicht, bei dem die HRTF-Filterung abgeschaltet ist. Es wird ein Saal simuliert, in dem sich die Tonquelle im Winkel von  $-20^\circ$  in einer Entfernung von 2 m vor der Hörerposition befindet. Als Eingangsdatei wird ein Impuls eingesetzt und die Software mit verschiedenen Ordnungen von Spiegelschallquellen durchlaufen (Diagr. 2).

Diagramm 2: Impuls ohne, mit niedriger und mit hoher Spiegelschallquellenordnung



Der Impuls als Eingangssignal wird in Blau, der Direktschallanteil in Rot, die Spiegelschallquellen erster Ordnung in Grün und die Spiegelschallquellen zweiter Ordnung in Gelb dargestellt. Die Lautstärke ist auf 1 m Entfernung normiert. Der rote Direktschallanteil hat einen Pegel von 50 %, da die Amplitudenabnahme linear ist. Zu erkennen ist die entfernungsbedingte Laufzeit der Spiegelschallquellen. Es ist ersichtlich, wie sich einem natürlichen Raumschall durch höhere Spiegelschallquellen Ordnungen genähert wird. Für den Versuch wird die mittlere, grüne Variante verwendet. Diese enthält die für die Distanzwahrnehmung wichtigen frühen Reflexionen.

### 6.1.2 Aufbau der Software für die Raumschallsimulation

Die Software ist in Anlehnung an eine Echtzeitimplementierung in zwei Stufen aufgeteilt. In der ersten Stufe findet die Berechnung der Position der Spiegelschallquellen in einem C-Programm statt und in der zweiten Stufe werden die Filteroperationen und Überlagerungen mit einem GNU-Octave-Script durchgeführt. Die Berechnung der Positionen der Spiegelschallquellen kann als Software effizient auf einer CPU stattfinden. Sie ist nur bei Änderungen der Position oder der Ausrichtung nötig. Für die Filteroperationen und Überlagerung bieten sich DSPs an. Diese Operationen müssen dauerhaft wiederholt werden und der Durchsatz ist entscheidend.

Eine Möglichkeit zur Realisierung des Systems wäre die Verwendung eines FPGAs. Hier kann ein Softcore-Prozessor die Berechnungen durchführen und z. T. vorhandene DSPs mit dem nötigen IP-Block zur Überlagerung und Verzögerung kombiniert werden.

#### Positionsberechnung der Spiegelschallquellen in einem C-Programm

Schon bei der Kompilierung wird durch ein `#define` die Ordnung der Spiegelschallquellen angegeben. Diese ändert sich nicht zur Laufzeit und so kann die Performanz erhöht werden. Im ersten Schritt der Ausführung werden die Eingabeparameter ausgewertet sowie die Initialisierung ausgeführt. Dabei werden die Geometrie des Raumes, die an den Wänden stattfindende Absorption, die Hörerposition samt Ausrichtung und die Koordinaten der Schallquelle im zweidimensionalen Raum, sowie gegebenenfalls eine zweite Koordinate für Bewegungen der Tonquelle, aus dem Aufruf übernommen. Der darauf folgende zweite Schritt läuft rekursiv ab. Hier wird der Raum samt Schallquelle gespiegelt, bis die gewünschte Anzahl an Räumen und damit Ordnung erreicht ist. Dadurch liegen die Koordinaten für alle Tonquellen in einer Liste vor. Im dritten Berechnungsschritt wird die Liste der Tonquellen durchiteriert, eine Unterscheidung von Direktschallquelle und Spiegelschallquellen ist nicht notwendig. Es wird jeweils die Entfernung zu der Hörerposition, die dadurch entstehende Abnahme der Lautstärke und die durch die Laufzeit entstehende Verzögerung berechnet. Für die Berechnung der Lautstärke wird der Abstand  $R$  zur Hörerposition verwendet. Die Amplitude nimmt mit  $\frac{1}{R}$  ab. Die Lautstärke ist so normiert, dass diese in einem Meter Entfernung den Ursprungspegel des Tonsignals hat. Die Verzögerung entspricht der Zeit, welche die Entfernung geteilt durch die Schallgeschwindigkeit ist. Zusätzlich wird anhand der Ordnung die Dämpfung durch die Reflexion an den Wänden mit einbezogen.

### Filterung und Überlagerung in einem GNU-Octave-Script

Durch den Fokus auf mathematische Operationen ist die Scriptsprache für die Faltungsoperationen der HRTF geeignet. Grundsätzlich wird in diesem Rechenschritt jede Schallquelle mit dem passenden HRTF-Filter versehen, in der Lautstärke angepasst und nach den Vorgaben verzögert. Für alle Schallquellen wird das identische Verfahren verwendet, sodass über die Liste der Quellen iteriert wird. Abschließend wird die Überlagerung der Signale mit Dekorrelation durchgeführt. Für jede Schallquelle wird dabei das gleiche trockene Signal als Ausgangssignal verwendet. Im Falle der panningbasierten Lösung wird statt der direkten HRTF-Filterung ein Panning auf die vier filternden virtuellen Lautsprecher durchgeführt. Es werden FFT-Funktionen einer GNU-Octave-Bibliothek eingesetzt. Als Filterkoeffizienten werden die selbst gemessenen HRTFs aus Abschnitt 6.1.3 verwendet.

Bei Bewegungen ist die Verzögerung und die Lautstärke über die Zeit nicht konstant, sodass bei Übergängen zwischen den Teilstücken abhängig vom Ursprungssignal Tonstörungen auftreten könnten. Um die Übergänge zu verdecken, wird Crossfading genutzt und die für die Berechnung verwendeten Teilstücke überlappen sich (Abb. 6.3).

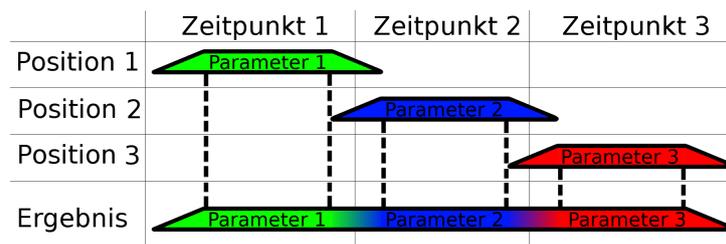


Abbildung 6.3: Überlagerung der Teilstücke bei Bewegung der virtuellen Schallquelle

Das Verfahren basiert auf dem gleichzeitigen Einblenden des neuen Abschnittes, während das vorherige Teilstück ausgeblendet wird. Bei der Überlagerung der Ton-Signale wird ein Korrekturfaktor verwendet. Dieser Korrekturfaktor für die Aussteuerung wird auf beide Kanäle angewendet, sodass die für die Ortung hilfreichen Rechts-Links-Lautstärkeunterschiede erhalten bleiben.

Die Korrelation der einzelnen gleichartigen Signale wird verhindert. Die Technik zur Dekorrelation wird in Abschnitt 3.2.1 beschrieben. So treten trotz der Überlagerung der Kopie eines Signals keine unnatürlich wirkenden Korrelationsartefakte auf.

Die verwendeten Filtersätze liegen jeweils als eigene Dateien vor. So können der inverse Filter zur Kopfhörerkompensation sowie der Satz an Koeffizienten bei Bedarf ausgetauscht werden.

### 6.1.3 Anfertigung benötigter Filter mit Hilfe von Kunstkopfmessungen

Für die Realisierung der Raumschallsimulation auf Basis der Filterung mit HRTFs wird ein entsprechender Filterkoeffizientensatz benötigt. Solche Koeffizientensätze sind teilweise öffentlich verfügbar, wie z.B. der des MIT [11]. Für die Durchführung der Studie wurde ein eigener Satz erstellt, weil dies Referenz- und Vergleichsmessungen mit dem verwendeten Kunstkopf ermöglicht.

Die Messung wurde in einem Schallmessraum der Hochschule für Angewandte Wissenschaften (HAW) durchgeführt, welcher über eine schallabsorbierende Innenverkleidung an den Wänden, der Decke und dem Boden verfügt. Über dem schallabsorbierenden Boden ist ein Netz gespannt, auf dem die Messinstrumente aufgestellt werden konnten. Es hängt eine drehbare Plattform von der Decke, auf der in diesem Fall der Kunstkopf positioniert wurde. Vor diesem wurde ein Lautsprecher auf Ständern mit verschiedener Höhe in einem Abstand von einem Meter aufgestellt (Abb. 6.4).



Abbildung 6.4: Aufbauvarianten für die Messung der HRTFs im Schallmessraum

Auf diesem Lautsprecher wurde die MLS-Datei wiedergegeben und die Reaktion des Systems, des Kunstkopfes, aufgezeichnet. Analog zu der im Abschnitt 3.2.2 vorgestellten Filtererzeugung wurde anschließend die gewünschte Impulsantwort erzeugt. Für die Messung wurde der Kunstkopf mit einer Rasterung von  $5^\circ$  um insgesamt  $180^\circ$  gedreht. Um alle Daten für die  $360^\circ$  Positionierung zu erhalten, wurden ab der Marke von  $180^\circ$  die Werte für den linken und rechten Kanal gespiegelt.

Bei den Messungen wurde eine Entfernung von einem Meter zwischen Kunstkopf und Lautsprecher nicht unterschritten. Dies dient zum einen dem Vermeiden von unterschiedlichen Distanzwahrnehmungen durch Nahfeld-HRTFs [24] und es wurde die Bauart des Lautsprechers beachtet. Bei diesem sind Tief-, Mittel- und Hochtonlautsprecher nebeneinander angebracht. Durch die Einhaltung der Minimaldistanz wurden die verschiedenen Frequenzen annähernd aus der gleichen Richtung aufgenommen.

Es existieren Messungen für drei Höhenebenen: Eine Messung auf Kopfhöhe, sowie eine leicht darunter und darüber. Da der eingesetzte Lautsprecherständer ein festes Raster hat, sind die Höhenstufen nicht vollständig symmetrisch. Der maximale Winkel in positiver Elevation liegt bei  $13,47^\circ$  und im negativen Bereich bei  $-16,14^\circ$  (Abb. 6.5).

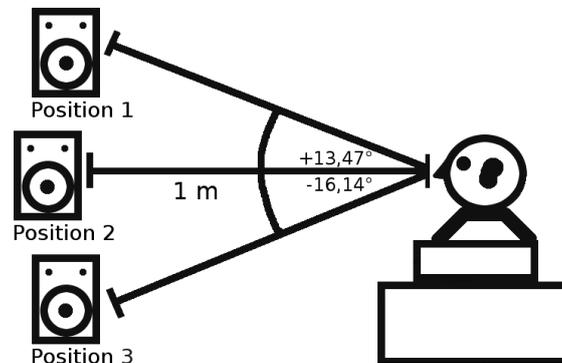


Abbildung 6.5: Winkel der drei Höhenstufen bei der Kunstkopfmessung

Für jede Höhenstufe ist eine Datei vorhanden, welche die Abfolge der jeweils gleich langen Impulsantworten enthält. Durch die Drehung des Kunstkopfes gegen den Uhrzeigersinn ist die Veränderung der ILD zu erkennen. Zuerst weist der linke Kanal (oben) und dann der rechte Kanal (unten) die höhere Lautstärke auf (Abb. 6.6).

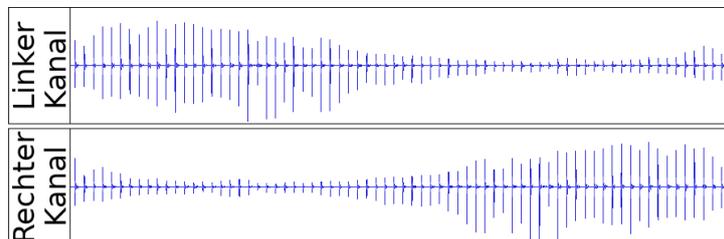


Abbildung 6.6: Stereospur der aneinandergereihten Impulsantworten der HRTF-Messung

Ursprünglich sollte die Studie mit Laborhardware durchgeführt werden. Aus diesem Grund wurde ein Kompensationsfilter für den Frequenzgang des für die Durchführung ausgesuchten Kopfhörers erstellt. Hierzu wurde der auf den Ohren des Kunstkopfes aufgesetzte Kopfhörer für die Wiedergabe der MLS verwendet. Diese Messung wurde mehrmals wiederholt und ein Mittelwert gebildet, da die optimale Positionierung des Kopfhörers schwierig ist. Mit den so bestimmten Impulsantworten wurde ein Filter erstellt und mit der GNU-Octave-Funktion *freqz(...)* wurde der Frequenzgang analysiert. Anschließend wurde dieser Frequenzgang invertiert und mit der Funktion *fir2(...)* ein FIR-Filter erzeugt, welches den Frequenzgang des Kopfhörers korrigiert.

Für die Messungen wurde der Kunstkopf HMS II.4, die Soundkarte Focusrite Scarlett 18i8, der Lautsprecher ADAM A8X und der Kopfhörer Sennheiser HD 600 verwendet (Abb. 6.7). Des Weiteren wurden ein Laptop für die Ausführung der Aufnahmesoftware, zwei unterschiedlich hohe Lautsprecherständer sowie zwei Holzplatten genutzt.



Abbildung 6.7: Darstellung der für die beiden Messungen verwendeten Geräte

Die zwei in den Ohren des Kunstkopfes befindlichen Mikrophone wurden mit der Soundkarte verbunden. Für die Ausgabe des Anregungssignals wurde für die HRTF-Messung der Monitorlautsprecher und für die Erzeugung des Kompensationsfilters der Kopfhörer eingesetzt. Über die per USB angeschlossene Soundkarte wurde gleichzeitig das Anregungssignal wiedergegeben und die Systemantwort gemessen (Abb. 6.8).



Abbildung 6.8: Logische Verbindungen der eingesetzten Messgeräte

Als Aufnahmesoftware wurde die Cubase Pro 8.0.0 eingesetzt. Um die Aufnahmelatenz zu reduzieren, wurde die Soundkarte per Audio Stream Input/Output (ASIO) und nicht über das Treibermodell Windows Driver Model (WDM) angesteuert.

## 6.2 Untersuchungsgegenstände und Aufbau der Studie

Es wird untersucht, wie sich die Einbettung von Spiegelschallquellen in ein HRTF-basiertes 3D-Audio-Modell auswirkt. Dazu wird das im Abschnitt 6.1 erstellte Modell in einer kopfhörergestützten Nutzerstudie eingesetzt. Hauptsächlich ist die durch diese erreichte Distanzwahrnehmung und die Externalisierung der virtuellen Schallquellen von Interesse. Hierzu wird die Erzeugung und die erzielte Genauigkeit bei der Distanzwahrnehmung beobachtet. Gleichzeitig wird die subjektiv wahrgenommene Externalisierung und Natürlichkeit erfragt. In diesen drei Bereichen werden durch den hinzugefügten Raumschall die größten Auswirkungen erwartet.

Neben diesen primären Aspekten wird untersucht, wie sich die Raumschallsimulation auf andere Bereiche der auditiven Wahrnehmung auswirkt, wie z.B. das Richtungshören. Untersuchungen zur Auswirkung auf die Richtungsbestimmung und das Auftreten von Hinten-Vorne-Fehlern sind deshalb in der Studie enthalten. In der Studie werden verschiedene Räume simuliert und die beiden Raumschallsimulationen mit einer Richtungs-simulation verglichen, welche nur den Direktschall beinhaltet. Die äußeren Umstände beeinflussen den Ablauf der Studie überdurchschnittlich.

### 6.2.1 Auftretende Einschränkungen durch die Pandemiesituation

Die Durchführung dieser Studie fällt in die Zeit der Corona-Pandemie und der gegen diese gerichteten Maßnahmen. Die erstmals am Ende des Jahres 2019 bei Menschen aufgetretene Zoonose Coronavirus Disease 2019 (COVID-19) führte zu einer weltweiten Pandemie. Die dagegen ergriffenen Maßnahmen beinhalten Regelungen<sup>1</sup> zur Beschränkung der Kontakte. Deshalb ist der Zugang zu vielen Institutionen beschränkt. Auch die Hochschulen schränkten den Zutritt stark ein. Dies ist der Grund, weshalb die Untersuchung als Studie im Home-Office durchgeführt wird. Die Proband\_Innen können nicht wie sonst üblich in einem vorbereiteten Raum in der HAW die Studie durchlaufen.

Hierdurch stellen sich deutlich schlechter kontrollierbare Versuchsbedingungen ein, was sich in mehreren Punkten auswirkt. Durch die Durchführung der Studie in den Räumlichkeiten und mit den Mitteln der Versuchspersonen liegen individuelle Umgebungen vor. Dies betrifft auch den verwendeten Kopfhörer, dessen Frequenzgang nicht mit einem Kompensationsfilter korrigiert werden kann. Auch war der Aufbau einiger Studiolautsprecher geplant, der bei den Proband\_Innen nicht durchführbar ist.

---

<sup>1</sup>Informationen Online abrufbar: <https://www.hamburg.de/allgemeinverfuegungen/>

Es sollten reale Lautsprecher um die Testpersonen herum aufgebaut werden, um so einen direkten Vergleich von realen und virtuellen Tonquellen zu ermöglichen (Abb. 6.9).



Abbildung 6.9: Ursprünglich geplanter Versuchsaufbau an der HAW.

Die Lautsprecher sollten auch der optischen Repräsentationen von realen und virtuellen Tonquellen dienen. Diese Funktion wird nun von dem Monitor übernommen. Um vergleichbare Resultate zu erhalten, muss der zu beurteilende Inhalt den gleichen Winkel im Field of View (FOV) der teilnehmenden Personen einnehmen (Abb. 6.10).

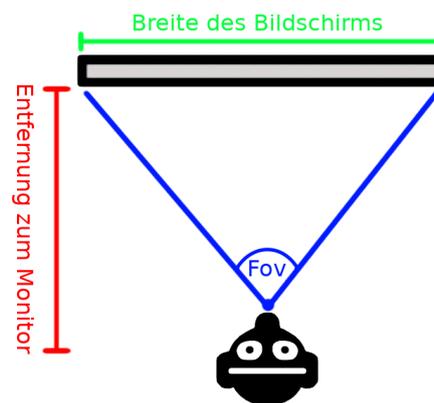


Abbildung 6.10: FOV für die Versuchsdurchführung

Das im englischen FOV genannte Sichtfeld beschreibt den Blickwinkel, in dem eine Wahrnehmung möglich ist. Der erreichbare FOV für Bildschirmhalte ist von der Monitorbreite und dem Sitzabstand abhängig und kann nach der Gleichung 6.1 berechnet werden.

$$FOV = 2 \cdot \arctan \left( \frac{\frac{1}{2} \cdot \text{Breite des Monitors}}{\text{Entfernung zum Monitor}} \right) \cdot \frac{180}{\pi} \quad (6.1)$$

Für das Verhältnis von eins zu eins für Monitorentfernung zu -breite beträgt der FOV 53,13°. Ein kleiner Teil des Sichtbereichs wird von dem Monitorrahmen belegt, so bleiben für die Verwendung in der Studie ca. 52° übrig.

Neben der eingeschränkten Kontrolle über die Versuchsumgebung erhöht die Durchführung der Studie im Home-Office auch die Wahrscheinlichkeit, dass ein Fehler auftritt. Die Proband\_Innen sind unbeobachtet, können Anweisungen missinterpretieren und es können Bedienfehler auftreten. Aus diesem Grund ist die Studie so angelegt, dass diese nicht alleine von der Nutzung des Monitors abhängig ist. Zusätzlich wird rein auditiv mit Referenzgeräuschen gearbeitet. Diese zweite, alternative Methode ist nicht von der Einhaltung der Sitzposition abhängig und es kann auch der Bereich außerhalb des Monitors genutzt werden.

Mit einem Vergleich der beiden Methoden kann überprüft werden, ob es systematische Ungenauigkeiten gibt. So können Rückschlüsse über die Auslegung der Studiendurchführung unter den beschriebenen Bedingungen gewonnen werden. Es ist also gleichzeitig eine Chance, die Möglichkeiten zur Untersuchung von Studien im Home-Office zu betrachten. Neben den fachlichen Untersuchungsaspekten wird deshalb auch die für die aktuelle Situation gewählte Vorgehensweise ausgewertet.

### 6.2.2 Ablaufplan der videobasierten Studie

Das Hauptaugenmerk beim Entwurf der Studie lag auf niedrigen Einstiegshürden und einer intuitiven Vorgehensweise. Die Studie besteht deshalb aus einem Video, welches die 3D-Tonspur der einzelnen Testfälle und die zugehörigen Grafiken enthält. Zu diesem Video wird ein Fragebogen ausgefüllt. So können die Proband\_Innen die Versuche selbst ohne Interaktion mit der Studienleitung durchführen. Die Proband\_Innen benötigen als Versuchsvoraussetzungen nur einen Rechner mit Bildschirm zum Abspielen des Videos und einen Kopfhörer für die Audiowiedergabe. Für den Versuch sind ca. 30 Minuten eingeplant, dazu kommt noch die Zeit zum Ausfüllen des Fragebogens, der Datenschutzerklärung und einer kleinen Einführung in den Themenbereich. Über diese Zeit sollte ein konzentriertes Zuhören möglich sein und bei 45-60 Minuten an zeitlichem Aufwand ist die Studie für potenziell teilnehmende Menschen nicht abschreckend.

Es gab die Alternative, statt eines Videos einzelne unkomprimierte Audio- und Bild-Dateien auszuliefern. So würde für die Audiospur kein verlustbehafteter Codec eingesetzt werden. In einer kontrollierten Versuchsumgebung bietet dies einen Vorteil, unter den gegebenen Umständen ist die Einsteigerfreundlichkeit wichtiger. Die geringere Fehleranfälligkeit wird einer potenziell genaueren Audiowiedergabe untergeordnet.

Die in Abschnitt 6.1.2 beschriebene Raumschallsimulation wird in einem Shell-Script genutzt, um die Tonspur der einzelnen Testfälle zu erzeugen und die Software GNU Image

Manipulation Program (GIMP) wird verwendet, um die dazugehörigen Grafiken zu erstellen. Für den Videoschnitt wird die ebenfalls unter der General Public License (GPL) stehende Software OpenShot Video Editor genutzt, um die erstellten Bilder mit den Audiodateien und der Einführung in das Themengebiet in einem Video zusammenzufügen. Der Fragebogen ist ein mit LibreOffice erzeugtes PDF-Formular, welches digital zurück gesandt werden kann.

Für die Durchführung der Studie erhält die Testperson eine E-Mail mit Downloadlinks zu dem PDF-Fragebogen, dem Video und der Datenschutzerklärung und lädt diese alles herunter. Die Einleitung wird in der PDF gelesen. Hiernach werden die Informationen der Person, zum Equipment und dem Ort der Versuchsdurchführung aufgenommen und die Datenschutzerklärung wird ausgefüllt. Der PDF-Fragebogen bleibt dann im Hintergrund geöffnet. Anschließend wird das Video gestartet. Es gibt eine kurze Einführung in das Themengebiet 3D-Audio. Es wird erläutert, dass es sich um Versuche zum dreidimensionalen Hören handelt.

Außerdem werden die für die Auswertung notwendigen Begriffe Natürlichkeit und Externalisierung erläutert. Schlussendlich wird auch erklärt, wie man das Video pausiert und zwischen dem Video-Player und dem PDF-Formular wechselt, um zwischen den Testfällen den Fragebogen auszufüllen. Der gesprochene Text befindet sich im Anhang.

Dann werden die in Abschnitt 6.2.3 dargestellten Versuche durchgeführt. Zu Beginn wird die jeweilige Versuchsnummer eingeblendet, um das richtige Auffinden auf dem Fragebogen zu ermöglichen. Anschließend wird die zu erfüllende Aufgabe erklärt. Die entsprechenden Bilder oder ein Platzhalter wird mit den Versuchstönen abgespielt und eventuell wiederholt. Hiernach wird eingeblendet *”Bitte das Video pausieren, den Fragebogen ausfüllen und anschließend fortsetzen”*. Durch die einfache Vorgehensweise - Video starten, Versuch anhören, pausieren und ausfüllen sowie anschließendem Fortfahren - wird erwartet, dass eine Fehlbedienung unwahrscheinlich ist.

### **Schema für die Positionierung der Tonquellen um die Testperson**

Für die Versuche mit unbewegten Tonquellen wird ein einheitliches Schema eingesetzt. Dies gilt sowohl für die Verwendung in den Versuchen mit optischen und akustischen Referenzpunkten als auch für die beiden simulierten Räume. Dies stellt zum einen eine Vergleichbarkeit der in der Studie verwendeten Methoden her und ermöglicht es zum anderen unterschiedliche Entfernungsbereiche abzudecken. Die Entfernung der Tonquelle zur Hörerposition wird entsprechend den Dimensionen des Raumes skaliert.

Im Radius von 1,5m (rot) um die Testpersonen werden keine Tonquellen positioniert (Abb. 6.11). In diesem Bereich liegen psycho-akustische Besonderheiten vor, die nicht untersucht werden sollen. Für die optisch ablaufenden Versuche gibt es eine Beschränkung auf den Sichtbereich (grün). Andernfalls kann auch der übrige Bereich (gelb) genutzt werden. Im Sichtbereich gibt es zwei Ringe äquidistanter Quellen vor der Hörerposition in 1,5 m (N1-N5) und 2 m (D1-D3) Entfernung.

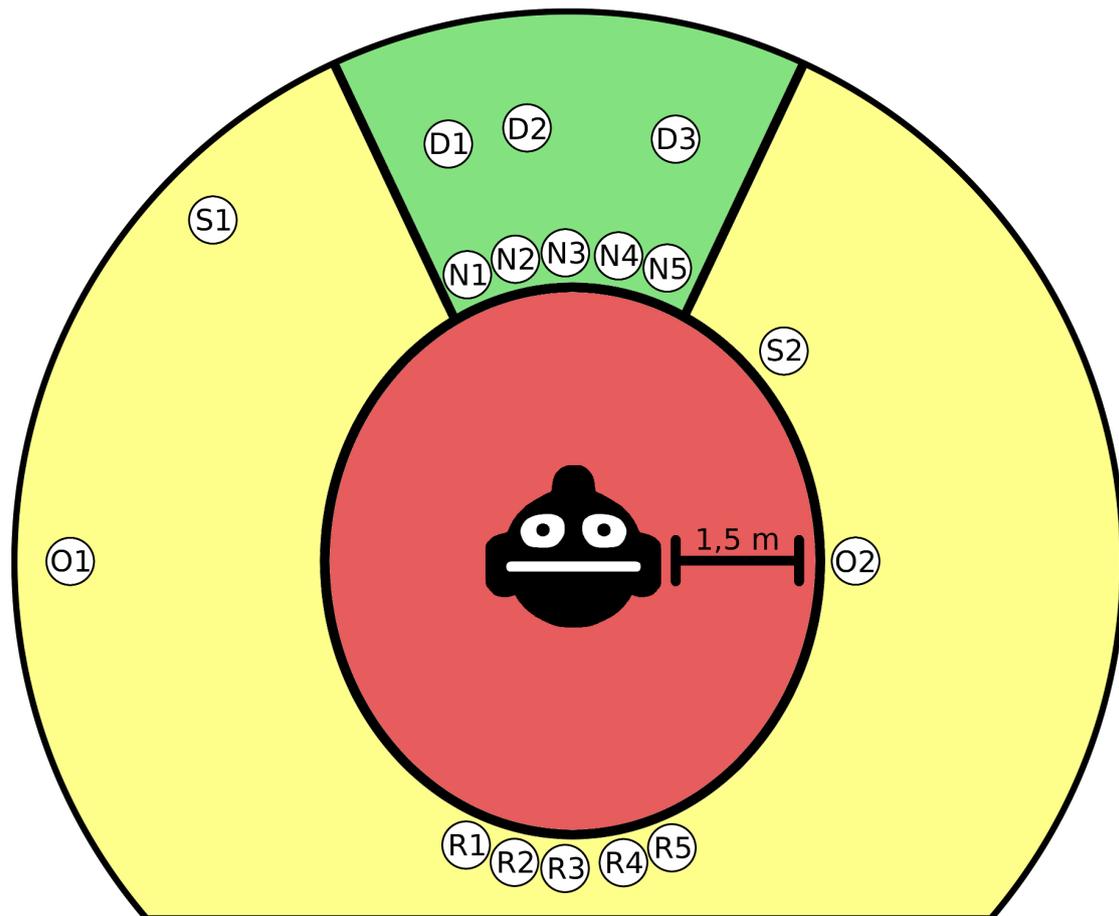


Abbildung 6.11: Positionierungsschema für die Tonquellen

Ein Teil dieser Quellen wird hinter die Hörerposition gespiegelt (R1-R5). Außerhalb des sichtbaren Bereichs sind zwei Tonquellen auf der zwischen den Ohren verlaufenden Achse in 2,5m und 1,5m Entfernung positioniert (O1 & O2). Auch befinden sich im 45° Winkel zwei Quellen schräg vor der Hörerposition in 2,5m und 1,5m Entfernung (S1 & S2).

### Eigenschaften der für die Studie simulierten Räume

Bei der Verwendung von Raumschall werden zwei verschiedene Räume simuliert, um die unterschiedlichen Hörsituationen erforschen zu können. Das Zimmer misst 5 m x 6 m, der Saal ist mit 27 m x 18 m deutlich größer. Um dasselbe Muster für die Tonquellen verwenden zu können, wird die Entfernung zu den Tonquellen im größeren Raum verdoppelt. Anhand der Grundrisse sind die Unterschiede bei Größe, Geometrie sowie bei der asymmetrischen Positionierung der Hörerposition zu erkennen (Abb. 6.12).

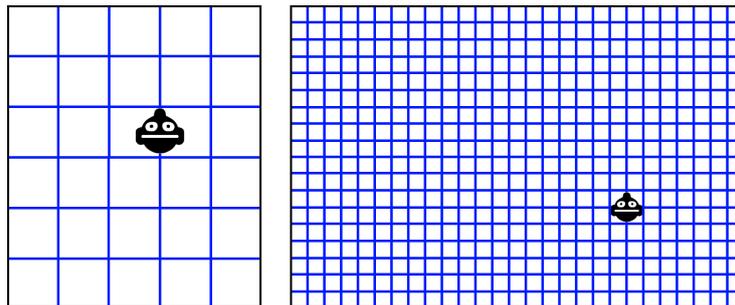


Abbildung 6.12: Grundrisse der mit dem Raumschallmodell simulierten Räume

Von verschiedenen Oberflächen wird der Schall frequenzabhängig unterschiedlich stark absorbiert. Dies wird mit dem Schallabsorptionsgrad  $\alpha$  bezeichnet, wobei für  $\alpha = 0,5$  gilt, dass 50 % des auftretenden Schalls absorbiert werden. Für verschiedene Materialien liegt der Absorptionsgrad aus DIN EN 12354-6:2004-04 [7] vor (Tab. 6.1).

Tabelle 6.1: Absorptionsgrad verschiedener Materialien im Frequenzbereich 1-4 kHz

Ausgewähltes Material	$\alpha$
Marmor, Fliesen, Klinker	0,02 - 0,03
PVC und Parkett	0,05 - 0,06
Fenster	0,06 - 0,02
Kalkzement	0,04 - 0,05
Kurzer Teppichboden	0,2 - 0,35
Vorhang	0,12 - 0,22

Für den Saal werden die folgenden Materialien angenommen: Marmorwände, PVC-Boden und einige Fenster. In dem Zimmer werden die Materialien verwendet: Kalkzement an den Wänden, Teppichboden und Vorhänge vor den Fenstern. Gemittelt ergibt sich für den Saal ein Schallabsorptionsgrad von  $\alpha = 0,04$  und für das Zimmer von  $\alpha = 0,16$ . Der Saal erhält durch die relativ stark reflektierenden Materialien deutlich mehr Raumschall als das Zimmer.

### Aufnahme und Eigenschaften der verwendeten Geräusche

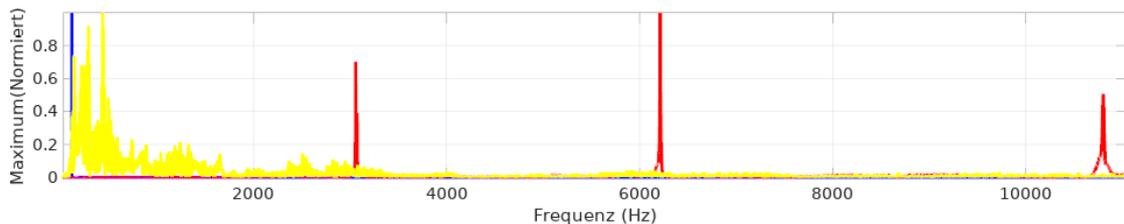
Es werden drei Geräusche verwendet, um unterschiedliche Frequenzbereiche abzudecken. Ein künstlich erzeugtes Brummen, das Schellen einer Fahrradklingel und eine männliche Stimme. Die Aufnahmen wurden in einem reflexionsarmen Raum mit dem Mikrofon AKG C3000B<sup>2</sup> durchgeführt (Abb.6.13).



Abbildung 6.13: Aufbau für die Aufnahme der Geräusche in der Studie

Der Brummtton ist das einzige künstlich erzeugte Geräusch, bei den anderen Geräuschen handelt es sich um Alltagsgeräusche. Der Vorteil bei der Verwendung bekannter Alltagsgeräusche liegt darin, dass die Lokalisierung von diesen bereits trainiert wurde. Bei der Fahrradklingel tritt zuerst eine Transiente auf, welche dann von einem Sustain gefolgt wird. Dies hat einen positiven Effekt auf die Distanzwahrnehmung.

Diagramm 3: FFT-Analyse der Frequenzmaxima der verwendeten Geräusche



Die Frequenzmaxima liegen für das Brummen (blau) bei 120 Hz, die Klingel (rot) bei 3, 6 und 11 kHz und die Stimme (gelb) bei für Sprache üblichen 0,3 - 4 kHz (Diagr. 3). Die Geräusche werden außerhalb des Bereichs des Nahfeldeffektes aufgenommen.

<sup>2</sup> Spezifikationen abrufbar unter: [https://de.ake.com/on/demandware.static/-/Sites-masterCatalog\\_Harman/default/dw32cbf567/pdfs/AKG\\_C3000\\_Cutsheet.pdf](https://de.ake.com/on/demandware.static/-/Sites-masterCatalog_Harman/default/dw32cbf567/pdfs/AKG_C3000_Cutsheet.pdf)

### 6.2.3 Ausführliche Beschreibung der durchlaufenen Testfälle

Die Studie ist in fünf Testblöcke unterteilt, auf diese werden die für alle Versuche eingeplanten 30 Minuten aufgeteilt. Häufig ist das Ziel, die gesuchte virtuelle Schallquelle zu lokalisieren, deshalb werden Markierungen eingesetzt, um den Geräuschen ein beschreibbares Objekt zuzuordnen. Dabei handelt es sich entweder um optische oder akustische Referenzwerte, an denen sich orientiert wird und welche benutzt werden können, um eine Tonquelle für den Fragebogen zu identifizieren. Bei den Versuchen mit optischen Referenzwerten gibt es jeweils eine Markierung in der Mitte des Monitors, die dem Zentrieren der Testperson dient. Für jede Aufgabe gibt es eine gesprochene Anweisung.

Für die erwartbare Genauigkeit sowohl der Richtungs- als auch die Entfernungsbestimmung wurden Annahmen gemacht. Diese wurden in einer Pilotstudie überprüft und nach deren Ergebnissen angepasst. Ziel ist es, den Bereich so zu wählen, dass weder die Testfälle die erreichbare Genauigkeit limitieren, noch die Versuchspersonen überfordert sind und deshalb keine verwertbaren Informationen entstehen. Für die Richtungsbestimmung gilt als Optimum des Auflösungsvermögens, Tonquellen im Winkel von  $3^\circ$  zu unterscheiden. Dies wird als untere Grenze angenommen. Es ist unwahrscheinlich, dass diese erreicht wird, da diese nur mit der Möglichkeit der Kopfdrehung bei dann fixen Tonquellen erreicht wird und Kopfbewegungen nicht erfasst werden. Da ein experimentelles Verfahren zur Erzeugung der Distanzwahrnehmung eingesetzt wird, werden verschiedene Entfernungsstufen genutzt. Diese entsprechen jeweils einer Entfernungszunahme um den Faktor drei. So werden sowohl geringe als auch große Distanzunterschiede abgedeckt. Ein Testblock über das maximale Auflösungsvermögen überprüft diese Annahmen.

#### **Richtungs- und Distanzhören mit optischen Markierungen als Einstieg**

Dieser Test dient der Eingewöhnung der Teilnehmenden an die Studie und dem Vergleich der Auswirkung verschiedener Geräusche auf die Richtungs- und Entfernungswahrnehmung. Es wird das im Abschnitt 6.1.1 beschriebene sparsame Modell der Raumschallsimulation eingesetzt, für welches es jeweils einen Testfall für Brummtton, Klingel und Stimme gibt. Als simulierter Raum wird der Saal verwendet. Zusätzlich gibt es noch Testfälle, bei denen nur der Direktschallanteil verwendet wird. Bei dem Richtungshören wird der Brummtton hierfür genutzt. Für das Distanzhören werden alle Tests mit Direktschall wiederholt, bei diesen ist die Lautstärke an die Entfernung angepasst. So kann die Auswirkung der Distanzmodellierung direkt beurteilt werden.

Für das Richtungshören werden hier zwei nebeneinander platzierte, optische Markierungen eingesetzt, deren Zentren  $15^\circ$  voneinander entfernt sind. An der Position einer der beiden Markierungen befindet sich die Tonquelle. Auf dem Fragebogen wird diese durch Ankreuzen ausgewählt.

Für das Distanzhören werden zwei Markierungen spiegelsymmetrisch angezeigt und an jeder der Positionen ein Geräusch abgespielt. Die Geräusche befinden sich in unterschiedlichen Entfernungen. Die Aufgabe ist nun das Geräusch auszuwählen, dessen Entfernung geringer ist. Die Entfernung unterscheidet sich dabei um 33 %.

Bei den optischen Markierungen handelt es sich um gefüllte Kreise (Abb. 6.14).

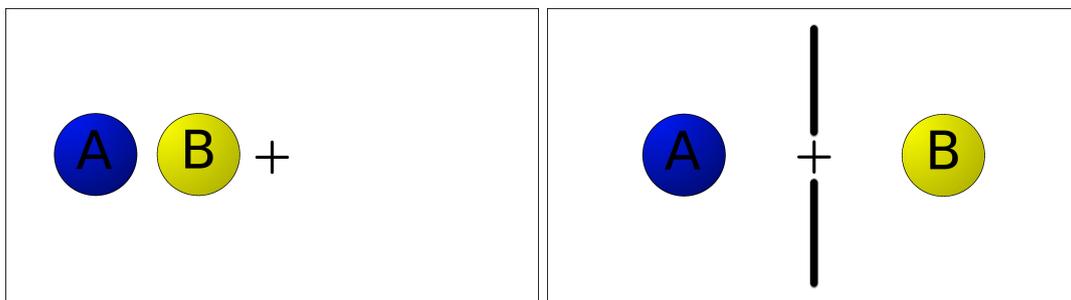


Abbildung 6.14: Monitorbild der Richtungs- (links) und Entfernungsfragen (rechts)

Der Testton wird jeweils mehrmals wiederholt, bevor um das Ausfüllen des Fragebogens gebeten wird. Ein + in der Mitte des Monitors bietet einen zentralen Punkt zur Ausrichtung auf den Monitor. Der Einführungstext lautet:

*”In diesem Abschnitt werden verschiedene Geräusche abgespielt, die dann Objekten auf dem Bildschirm zugeordnet werden sollen. Es gibt erst Aufgaben zum Richtungshören und dann zum Entfernungshören. Bei dem Richtungshören wird eine Geräuschquelle verwendet und einem der beiden dargestellten Objekte zugeordnet. Bei dem Entfernungshören werden zwei Geräusche abgespielt und anschließend soll angegeben werden, welches Geräusch weniger weit entfernt ist. Um die Ausrichtung auf den Monitor zu erleichtern, ist die Bildschirmmitte markiert.”*

Genutzte Positionen: N1-N5, D1-D3

Testanzahl: 3 Geräusche x  $3, \bar{3}$  (1 Richtung + 1 Entfernung +  $1, \bar{3}$  Direktschall) = 10

Testzeit:  $3, \bar{3}$  Minuten (20 Sekunden/Testfall)

### Fortgeschrittenes Richtungshören mit optischem Raster

Hier wird die Auswirkung auf die Richtungsgenauigkeit und das Auftreten von Hinten-Vorne-Fehlern erforscht. Es werden die beiden Raumschallmodellierungen und der isolierte Direktschall verwendet. Dazu wird das Geräusch Klingel im Sichtbereich oder exakt im gespiegelten Winkel hinter der der Testperson abgespielt.

Die Testpersonen richten sich zu Beginn des Testfalls auf den Bildschirm aus, schließen die Augen und hören das Geräusch. Sie deuten dann mit dem Finger in die Richtung der Tonquelle, öffnen die Augen und können die Gradzahl von der Skala ablesen (Abb. 6.15).

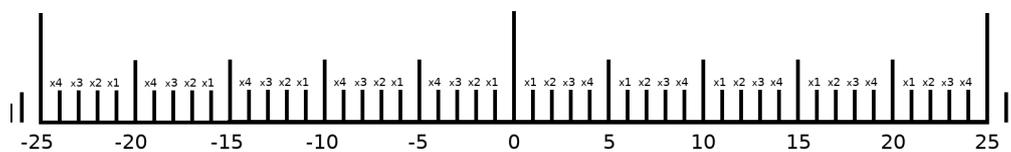


Abbildung 6.15: Raster für die Angabe des Winkels der Schallquelle

In den Fragebogen wird entweder die abgelesene Gradzahl oder ein *X* für eine Nichterkennung der Richtung eingetragen. Mit dem *X* werden die Hinten-Vorne-Fehler erfasst. Die Tonquellen befinden sich an denselben Positionen wie bei der Einstiegsrunde. Durch diesen Test können so durch Vergleiche zwischen den Aufgaben systematische Fehler durch die Monitorgröße/Sitzposition festgestellt werden.

Der Einführungstext lautet:

*”In diesem Testfall wird eine Klingel abgespielt und anschließend soll deren Richtung angegeben werden. Dazu wird auf dem Monitor ein Raster angezeigt. Zu Beginn der Aufgabe richtet man sich zentral auf den Monitor aus und schließt dann die Augen. Anschließend hört man das Geräusch und zeigt mit dem Finger in die Richtung. Mit der anderen Hand kann man das Video pausieren. Abschließend wird die Gradzahl abgelesen und in den Fragebogen eingetragen. Falls ein Geräusch nicht zugeordnet werden kann, dann bitte einfach ein 'X' eingeben.”*

Genutzte Positionen: N1-N5, R1-R5

Testanzahl: 1 Geräusch x (4 (2 Distanz Modelle x 2 Räume) + 1 Direktschall) x 2 Invertierung = 10

Testzeit: 3 Minuten / (20 Sekunden/Testfall)

### Bewegliche Tonquellen in verschiedenen virtuellen Umgebungen

Der Fokus dieses Tests liegt auf der Natürlichkeit und der wahrgenommenen Externalisierung. Nebenaspekte sind die Auswirkung optischer Reize und die Wahrnehmung der Bewegungen. Es wird eine kontinuierlich bewegte Stimme als Quelle eingesetzt. Diese werden durch einen Raum von links nach rechts (blau) und von hinten nach vorne bewegt (grün) (Abb. 6.16). Das Gesprochene lautet: *"Dies ist eine bewegliche Tonquelle."*

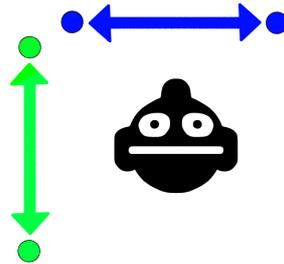


Abbildung 6.16: Bewegungsmuster für den Versuch mit bewegten Tonquellen

Bei der Hälfte der Testfälle ist die Bewegungsrichtung invertiert. Zur Testfallreduktion wird die Quelle pro Kombination aus Modell und Raum entweder von vorne nach hinten oder von links nach rechts bewegt. Hierbei werden beide Räume simuliert und der isolierte Direktschall mit beiden Raumschallmodellen verglichen. Zur Kontrolle wird für jedes dieser Modelle eine ruhende Tonquelle abgespielt. Auf dem Monitor werden zu den simulierten Räumen Bilder gezeigt (Abb. 6.17).



Abbildung 6.17: Angezeigte Bilder: Saal (links) und kleines Zimmer (rechts)

Um die Auswirkung von Hör- und Sehsinn zu beurteilen, werden zwei Testfälle wiederholt und die Abbildung des Raumes so gewählt, dass diese nicht mit der akustischen Simulation übereinstimmt. In den Fragebogen wird eingetragen, ob und in welche Richtung sich das Geräusch bewegte. Zusätzlich wird für jeden Testfall nach Externalisierung und der Natürlichkeit in fünf Abstufungen gefragt.

Der Einführungstext lautet:

*”In diesem Testfall stehen bewegliche Tonquellen und deren Wahrnehmung im Vordergrund. Diese bewegen sich entweder von links nach rechts oder von vorne nach hinten bzw. in die entgegengesetzte Richtung. Diese Bewegungsrichtung soll für die Testfälle angegeben werden. Zusätzlich wird nach der Externalisierung und der Natürlichkeit gefragt. Bei der Natürlichkeit bitte eintragen, ob das Geräusch einen natürlichen, d.h. nicht künstlichen oder ungewohnten Eindruck erzeugt hat. Unter Externalisierung wird im Audiobereich verstanden, dass ein Geräusch außerhalb des Kopfes wahrgenommen wird. Diesen Eindruck auch bitte für jeden Testfall festhalten.”*

Tonquellen: Bewegte Tonquellen aus dem Muster

Testanzahl: 1 Geräusch x 3 Modelle x 2 Räume x (2 x 0.5) Bewegungsrichtungen + 2 Optisch + 3 Unbewegt=11

Testzeit: 5,5 Minuten (30 Sekunden / Testfall)

### **Bestimmung des Auflösungsvermögens der Distanzwahrnehmung**

Dieser Test bestimmt bei, welchen Entfernungen sich die Distanz von zwei Tonquellen noch unterscheiden lässt. Dazu werden jeweils zwei Tonquellen mit demselben Geräusch verwendet. Diese befinden sich in unterschiedlicher Entfernung, aber in dem gleichen Winkel zur Testperson. Es wird an beiden Positionen das Geräusch nacheinander abgespielt. Anschließend wird gefragt, ob sich die Geräusche an derselben Position befanden. Es wird nur binär geantwortet, damit werden die Externalisierung und der Realismus nicht untersucht. Da die Distanzauflösung im Fokus liegt, wird eine Fahrradklingel verwendet und ausschließlich mit beiden Raumschallmodellen gearbeitet. Es werden wieder die zwei Räume mit den unterschiedlichen Nachhallzeiten verwendet. Es wird darauf hingewiesen, dass in dem Teil der Studie das Auflösungsvermögen untersucht wird, obwohl es eventuell ein Bias erzeugt. So können die Teilnehmenden sich maximal auf die Abweichungen der Quellpositionen konzentrieren. Der Einleitungstext lautet hierfür:

*”Dieser Test dient der Erforschung der Fähigkeit zur Unterscheidung der Entfernung von zwei Tonquellen. Dazu wird zweimal dasselbe Geräusch abgespielt. Anschließend soll angegeben werden, ob diese sich an der gleichen oder an unterschiedlichen Positionen befinden.”*

Tonquellen: 33% 12,5%, 5% und 0% Abstand zu den Tonquellen S1 & S2

Testanzahl: 1 Geräusch x 2 Räume x 2 Modelle x 4 Distanzvarianten=16

Testzeit: 5,3 Minuten / (20 Sekunden/Testfall)

### Richtungs- und Distanzhören mit akustischen Referenzpunkten

Dieser Test läuft rein auditiv ab. Es finden Versuchsaufgaben zur Bestimmung der Richtungsgenauigkeit als auch für die Entfernungswahrnehmung statt. Das grundsätzliche Vorgehen hierfür ist dasselbe, aber es wird in einem Fall der Winkel und im anderen Fall die Entfernung variiert. Es werden jeweils zwei Geräuschquellen als Referenz eingesetzt. Zwischen diesen wird eine dritte Tonquelle positioniert. Für alle drei Quellen wird das selbe Geräusch, die Klingel, verwendet. Auf dem Fragebogen wird dann auf einer Skala mit fünf Abstufungen die Quelle zwischen den beiden Referenzpositionen eingetragen. Zuerst wird die erste Referenzquelle, dann die Zielquelle und abschließend die zweite Referenzquelle abgespielt.

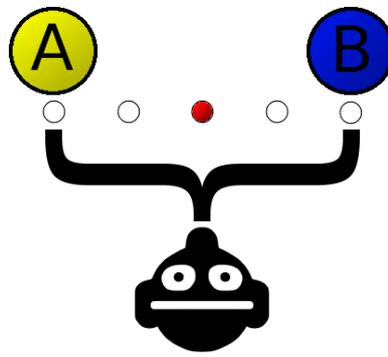


Abbildung 6.18: Richtungshören mit akustischen Referenzpunkten

Das gelbe Objekt *A* und das blaue Objekt *B* stellen die Referenzquellen dar (Abb. 6.18). Die kleinen Kreise unter diesen sind Repräsentationen der Auswahlmöglichkeiten auf dem Fragebogen. Die Position der gesuchten Tonquelle wird mit einem rot gefüllten, kleinem Kreis verbildlicht. Das Raster hat  $10^\circ$  Schritte für die Richtungs- und 5 % Schritte für die Entfernungsbestimmung.

Für die Messungen werden wieder die Positionen der Tonquellen aus den Versuchen mit optischen Referenzpunkten verwendet, um diese vergleichen zu können. Zusätzlich werden Tonquellen verwendet, welche auf der Symmetrieachse zwischen den Ohren liegen. So wird genutzt, dass der mögliche Bereich für Tonquellen hier nicht auf das Sichtfeld begrenzt ist.

Als Tonquelle wird ausschließlich das Klingeln verwendet. Es werden im Raumschallmodell beide möglichen Räume simuliert. Für die Richtungsbestimmung und Distanzwahrnehmung werden Direktschall und die beiden Raumschallmodelle verwendet.

Der Einleitungstext lautet hierfür:

*”Bei diesem Test wird dreimal das gleiche Geräusch abgespielt. Das erste und das letzte Geräusch spannen einen Bereich auf, in dem das mittlere Geräusch eingeordnet werden soll. Man gibt auf der Skala an, ob sich das Geräusch eher an der Position des ersten Geräusches, hier als 'A' bezeichnet, oder an der des letzten Geräusches, hier als 'B' bezeichnet, befindet. Die fünf Kugeln symbolisieren die Auswahlmöglichkeiten, wobei die äußersten Auswahlmöglichkeiten exakt den Positionen von 'A' oder 'B' entsprechen.”*

Tonquellen: N1-N5, R1-R5 und O1 & O2

Testanzahl: 1 Geräusch x (10 (3 Modelle Richtung x 2 Räume x Invers)) + (6(3 Modelle Distanz x 2 Räume x 1 Bereich))=16

Testzeit: 5,3 Minuten / (20 Sekunden/Testfall)

### **Vor der Erprobung verworfene Testfälle**

Einige Ideen für Testfälle werden nicht verwendet. Entweder, weil diese den Zeitrahmen der Studie gesprengt hätten, es einen effizienteren Weg gibt, an die gewünschten Informationen zu gelangen, oder die Vorgehensweise nicht praktikabel ist.

**Einzeichnen in ein Koordinatensystem:** Die Testperson zeichnet in eine Draufsicht des virtuellen Raumes mit dem eigenen Aufenthaltsort und der Ausrichtung die Position ein, an der die Tonquelle gehört wurde. Dies kann über Abstufungen geschehen, z.B. zuerst nur die Wahl des Quadranten und anschließend das Einzeichnen der konkreten Position im Raster. Hierbei sind die Gründe für die Nichtverwendung sowohl auftretende Schwierigkeiten bei der Auswertung vieler Grafiken als auch die Fähigkeit der Proband\_Innen die subjektive Wahrnehmung in ein Koordinatensystem aus der Vogelperspektive einzutragen.

**Kreisende Tonquelle für die Distanzeinschätzung:** Eine virtuelle Tonquelle kreist um die Testperson in einer festen Distanz. So wird eine richtungsunabhängige Distanzeinschätzung ermöglicht. Dieses Vorgehen wurde nicht gewählt, da das Hinzufügen der Referenztonquellen problematisch ist. Entweder müssten diese gleichzeitig abgespielt werden und würden den Versuch stören oder sie würden mit einem zu großen zeitlichen Abstand zur gesuchten Quelle abgespielt werden. Außerdem würden so nur Informationen über sich bewegende Tonquellen erhalten werden.

### 6.2.4 Aufbau des für die Studie verwendeten PDF-Fragebogens

Der Fragebogen ist zweigeteilt. Der erste Abschnitt enthält die Angaben zur Person und der Versuchsumgebung. Dies ist notwendig, da die Versuche nicht in einer kontrollierten Umgebung stattfinden. Auf diese Weise können statistische Zusammenhänge zwischen den beteiligten Personen und den Gegebenheiten vor Ort festgestellt werden.

Bei dem zweiten Abschnitt handelt es sich um die Fragen zur Beurteilung der Testfälle. Das Verfahren für einzelne Fragetechniken wird mit einer kurzen Begründung für die Vorgehensweise dargestellt.

#### Grundinformationen zur teilnehmenden Person und der Versuchsumgebung

Der Fragebogen beginnt mit einer kurzen Erläuterung der Studie. Anschließend werden die Angaben zur Person erfasst. Dies beinhaltet die Punkte Alter, Körpergröße, Geschlecht, Besonderheiten, wie z.B. Hörverluste, genutzte Audio-Unterhaltung pro Tag und Vorkenntnisse im Bereich Audio. Aus den Bereichen Musik, Filme/Serien, Hörbücher und Podcasts wird die präferierte Audiounterhaltung ausgewählt. Zusätzlich wird nach der Nutzung von Videospielen und nach der Berufsausbildung gefragt. Die Erfassung dieser Merkmale ermöglicht es durch Korrelation Zusammenhänge zwischen den ProbandInnen und Versuchsergebnissen herzustellen.

So wirkt sich das Alter negativ auf die Fähigkeit zur Wahrnehmung hoher Frequenzen aus. Die Ohrengröße ist proportional zur Körpergröße. HRTF-Filterung liefert die beste Genauigkeit, wenn der Ausgangskunstkopf des Filters den eigenen Ohren am ähnlichsten ist und könnte so unterschiedliche Richtungseinschätzungen der Testpersonen erklären. Die Frage nach Besonderheiten dient dazu, Probleme bei der Lokalisation zu erklären, welche z.B. durch Schwerhörigkeit oder unilateralen Hörverlust verursacht werden. Die Zeit, welche pro Tag mit Audio-Unterhaltung verbracht wird und die Art der konsumierten Inhalte ist aus folgender Überlegung heraus von Interesse: In interaktiven Inhalten wie Computerspielen oder VR-Anwendungen wird mit den genutzten Soundeffekten räumliches Hören trainiert. Dies sollte in abgeschwächter Form auch bei Serien und Filmen der Fall sein. Ausgiebiger Musikkonsum könnte dazu führen, dass der Realismus von Geräuschen besser beurteilt werden kann. Es werden auch teilnehmende Personen aus dem Audibereich erwartet, welche die verwendeten Techniken kennen und sich so u.U. anders verhalten. Der gelernte oder ausgeübte Beruf ist von Interesse, da dies Vorkenntnisse oder unterschiedliche Erfahrungswelten beinhalten kann.

Zusätzlich zu den personenbezogenen Daten werden die eingesetzte Technik und Informationen zur Umgebung festgehalten:

Abspielgeräte, also Soundkarte und Kopfhörer, verwendete Software, sowie Raumgröße und Uhrzeit. Zwischen den verfügbaren Kopfhörern und Soundkarten bestehen Qualitätsunterschiede, welche Auswirkungen auf die Versuche haben. Visuelle Reize und die Wirkung des Raumes, in dem der Versuch durchgeführt wird, beeinflussen die Wahrnehmung. Deshalb werden die Raumgröße (B, T, H) und der Bodenbelag erfasst. Durch die Gewöhnung an die örtlichen Verhältnisse liegt zwar ein Bias vor. Die schriftlichen Angaben dienen dazu, eine qualifizierte Einschätzung über deren Auswirkung treffen zu können.

### **Datenschutzerklärung zur DSGVO-konformen Studiendurchführung**

Damit die Studie unter Beachtung der Datenschutz-Grundverordnung (DSGVO) durchgeführt wird, erhalten die Studienteilnehmer\_Innen eine Datenschutzerklärung. In dieser wird die Verwendung der erhobenen Daten auf diese Studie und die damit zusammenhängende Auswertung beschränkt. Die von den Teilnehmer\_Innen erhaltenen E-Mails mit den ausgefüllten Fragebögen werden nach dem Abspeichern des Anhangs gelöscht. So können die Versuchsergebnisse und Angaben zur Person keiner Einzelperson zugeordnet werden und werden auf diese Weise anonymisiert.

## 7 Auswertung der Studie zur auditiven Distanz

Von den 24 verschickten Fragebögen wurden 16 ausgefüllt zurück gesendet und mit der Sprache R ausgewertet [25]. Diese Sprache ist für statistische Auswertungen konzipiert und mit R-Studio existiert eine frei verfügbare Entwicklungsumgebung. Die Informationen aus den PDF-Fragebögen wurden in eine csv-Datei übertragen. Zusätzlich wurde eine csv-Datei mit den Sollwerten erstellt und ermöglicht so eine Berechnung des Azimut-Fehlers und anderer Abweichungen. In einem Fragebogen fehlten Werte für zwei Teilaufgaben, weshalb diese Werte nicht in die betroffenen Aufgaben eingingen.

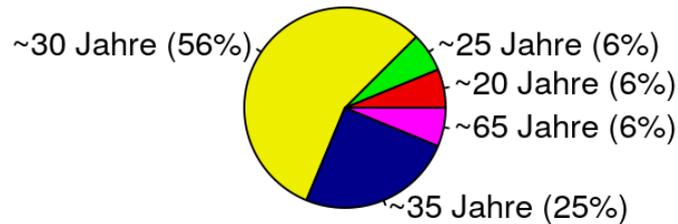
Die Auswertung der Studie beginnt mit der Übersicht über die Proband\_Innen und die jeweiligen Versuchsumgebungen. Anschließend folgt zuerst eine direkte Auswertung der einzelnen Teilversuche und anschließend die Detailauswertung einzelner Versuchsaspekte und Nebenbetrachtungen. Abschließend gibt es eine einordnende Zusammenfassung der Ergebnisse mit einer Methodenkritik.

### 7.1 Übersicht über die Proband\_Innen und deren jeweilige Versuchsumgebungen

Die Studie fand nicht unter kontrollierten Bedingungen statt, weshalb erweiterte Angaben genutzt werden, um auffällige Unterschiede aufzudecken. Die Räumlichkeiten entsprachen eher dem modellierten Zimmer, als dem modellierten Saal. Lediglich die Fußböden waren im Durchschnitt schallhärter, da sehr selten Teppich vorhanden war. Für den Versuch wurde überwiegend die integrierte Soundlösung und selten eine externe oder hochwertige Soundkarte eingesetzt. Der Großteil der verwendeten Kopfhörer war geschlossen, sodass die Geräusche von außen abgeschwächt wurden. Der VLC- und der Windows-Mediaplayer waren die hauptsächlich verwendeten Medienabspieler.

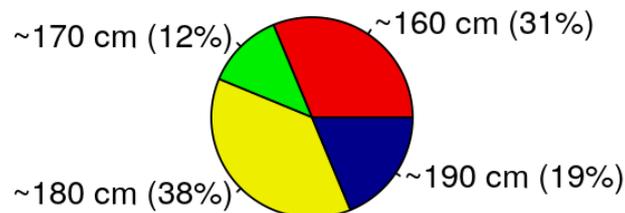
Neben diesen Ähnlichkeiten gab es große Unterschiede bei dem Preisniveau der Kopfhörer und der genutzten Bildschirmdiagonale. Die Unterschiede bei der Bildschirmdiagonale sind auf die Verwendung von Laptop und Desktop-Rechnern zurückzuführen und könnten Auswirkungen auf die Versuche mit optischen Markierungen haben.

Diagramm 4: Auf 5 Jahre gerundetes Alter der Proband\_Innen



Das Alter der Proband\_Innen liegt im Durchschnitt bei 32 Jahren. Die Werte reichen von 22 bis 66 Jahre, ein Großteil hat jedoch ein Alter nahe des Durchschnitts (Diagr. 4).

Diagramm 5: Auf 10 cm gerundete Körpergröße der Proband\_Innen



Die Körpergröße beträgt im Durchschnitt 175 cm und reicht von 158 cm bis 196 cm. Hier sind drei Bereiche jeweils ähnlich stark vertreten (Diagr. 5). Auf Grund der Unterschiede in diesem Bereich wird die Genauigkeit bei der Richtungseinschätzung untersucht.

Von den 16 Proband\_Innen besuchte die Mehrheit (56 %) Hochschulen, die übrigen verfügen über eine Berufsausbildung oder es wurden keine Angaben gemacht. Die Angaben zu den Geschlechtern teilen sich in 50 % männlich und 50 % weiblich auf. 25 % der Teilnehmenden gaben an über Expertise im Audibereich zu verfügen. Die meisten konsumieren regelmäßig Filme und Musik, seltener sind Podcasts. Die pro Tag konsumierte Menge an Audioinhalten variiert mit zwei bis 360 Minuten stark. Die Hälfte der Befragten spielt Videospiele wie Egoshooter und VR-Inhalte, welche häufig 3D-Audio enthalten.

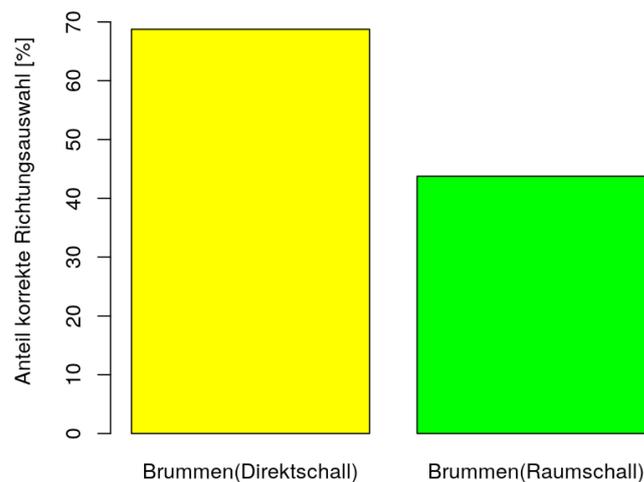
## 7.2 Detailauswertung der einzelnen Teilversuche

In diesem Abschnitt werden die Teilversuche entsprechend der Reihenfolge in der Studie ausgewertet. Ziel ist es, die damit verbundenen Fragestellungen zu beantworten. In manchen Fällen ist dies nicht auf die geplante Weise möglich, dann werden die dadurch erzielten Ergebnisse oder dabei aufgetretenen Probleme thematisiert.

### 7.2.1 Versuch 1: Überblick und Vergleich der Geräuschquellen

Im ersten Teilversuch wurde bestimmt, wie häufig die Richtung eines Brummtones einem von zwei Objekten auf dem Monitor richtig zugeordnet wurde. Dazu wurde einmal nur mit Richtungsfilterung und einmal zusätzlich mit dem durch das sparsame Modell hinzugefügten Raumschall gearbeitet. Das Brummen wird mit hinzugefügtem Raumschall (grün) schlechter lokalisiert als in der Variante mit Direktschall (gelb) (Diagr. 6).

Diagramm 6: Brummen mit und ohne Raumschall bei der Richtungsauswahl

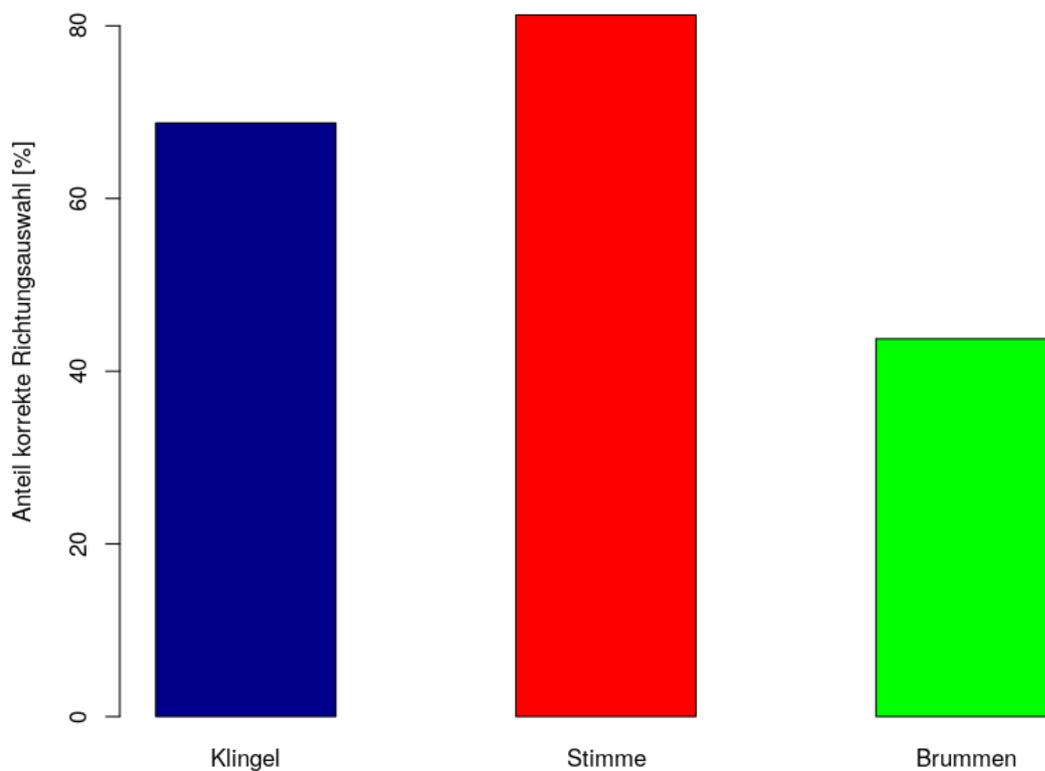


Die schlechtere Lokalisierbarkeit trotz erweiterter Simulation kann wie folgt erklärt werden: Im Gegensatz zur Realität mit diffusem Raumschall trifft durch das Spiegelschallquellenmodell aus verschiedenen Richtungen präzise lokalisierbarer Schall ein und stört so die Lokalisation. Zusätzlich ist das künstlich erzeugte Brummen durch sein enges Spektrum störanfällig, denn es ist auf wenige Richtungs-Hinweise beschränkt. So kann durch die niedrige Frequenz von 120 Hz nur die ITD und nicht die ILD verwendet werden.

Im nächsten Teilversuch wurde die Richtungsgenauigkeit verschiedener Geräusche verglichen. Es wurde wieder aus zwei Objekten auf dem Monitor ausgewählt und die Prozentwerte für die korrekte Auswahl für die drei Geräusche Klingel, Stimme und Brummen verglichen. Für alle Geräusche wurde das sparsame Raumschallmodell in einem simulierten Saal verwendet.

Für die Stimme (rot) konnte die Richtung am häufigsten richtig angegeben werden, gefolgt von der Klingel (blau). Das eingangs untersuchte Brummen (grün) ist das in der Richtung am ungenauesten eingeschätzte Geräusch der Auswahl (Diagr. 7).

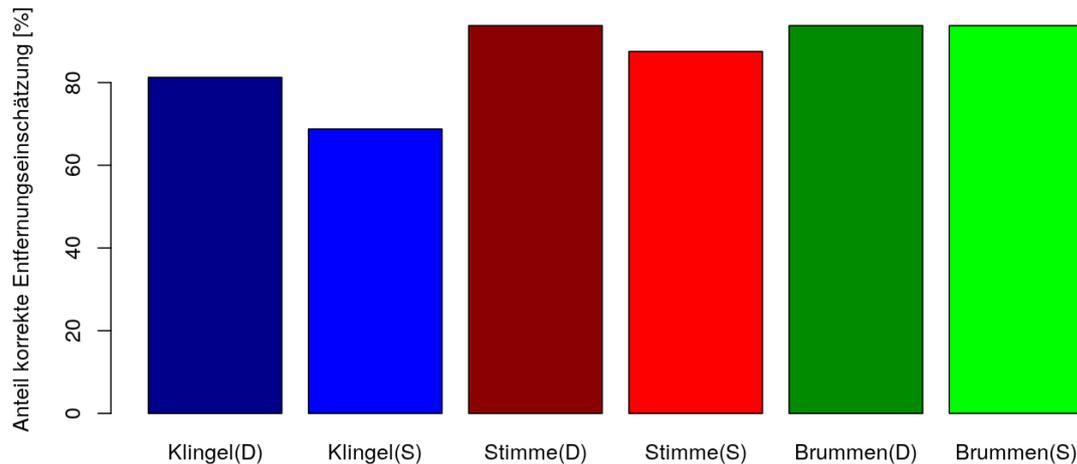
Diagramm 7: Vergleich verschiedener Geräusche bei der Richtungsauswahl



Im vorherigen Abschnitt wurde erläutert, warum Brummen schlecht zu lokalisieren ist. Sowohl Stimme als auch Klingel verfügen über einen größeren Frequenzbereich als das Brummen. Bekannte Geräusche lassen sich besser einschätzen als ungewohnte. Da die meisten Proband\_Innen an Stimmen und deren Lokalisation gewöhnt sind, lässt sich die Reihenfolge der Stimuli für die Richtungswahrnehmung hiermit erklären.

Anschließend wurden zwei Geräuschquellen verwendet und diesmal die Entfernung um jeweils 33 % variiert. Es wurde erfragt, welche die Nähere ist. Es wurden zwei Entfernungsmodellierungen eingesetzt: In einem Fall wurde nur die Lautstärke variiert, im anderen Fall wurde zusätzlich Raumschall mit dem sparsamen Modell hinzugefügt. Daraus wurde dann ein Prozentsatz für die korrekte Erkennung des Distanzunterschiedes berechnet.

Diagramm 8: Vergleich verschiedener Geräusche bei der Distanzauswahl



Nur bei der Klingel (blau) treten häufig Fehler auf, bei Stimme (rot) und Brummen (grün) findet meist eine korrekte Zuordnung der Entfernung statt (Diagr. 8). Es sieht so aus, als ob in den meisten Fällen der Versuchsaufbau der limitierende Faktor ist. Die hohe Fehlerrate bei der Klingel ist unerwartet, da diese für die Distanzmodellierung gut geeignet sein sollte. Es handelt sich um eine Vergleichssituation, bei der die Entfernung des dreimaligen Anschlages der Klingel verglichen wird. Eventuell klingen die drei aufeinanderfolgenden Anschläge untereinander zu unterschiedlich, um einen einheitlichen Entfernungseindruck zu erzeugen.

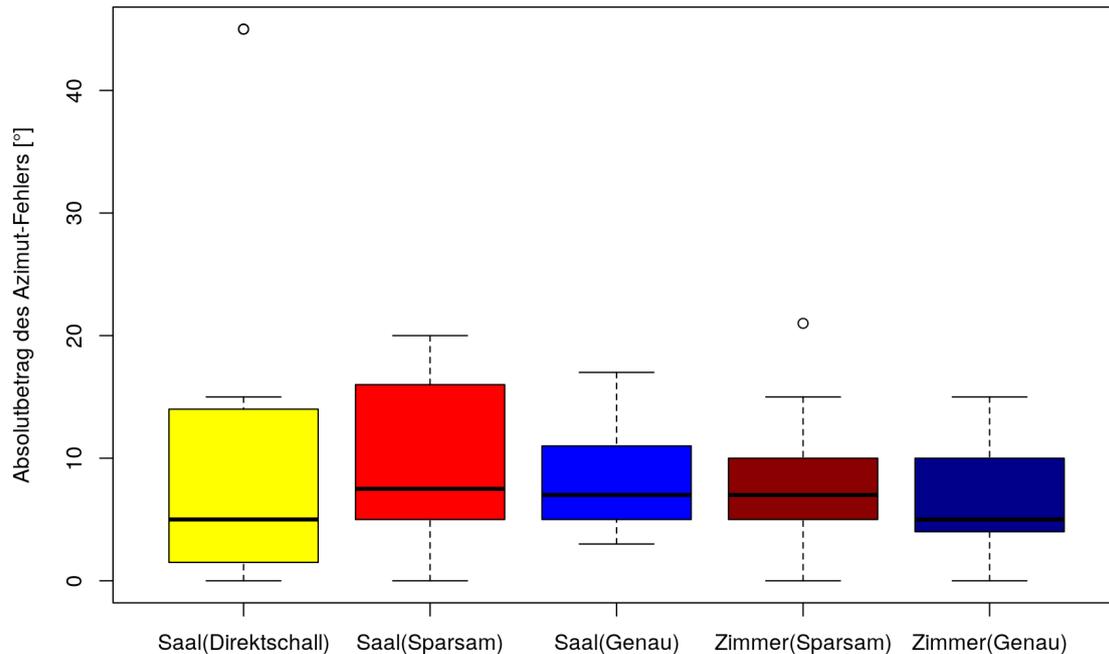
Die Modellierung mit Direktschall wird in dunkleren Farbtönen und die mit Raumschallmodell in den helleren Tönen dargestellt. Auffällig ist, dass die Variante mit Spiegelschallmodell bei der Klingel und der Stimme weniger häufig korrekt erkannt wird. Bei der Stimme kann es sich noch um einen einzelnen Ausreißer handeln, während die Abweichung bei der Klingel etwas größer ist.

Dieser Versuch stand jedoch relativ am Anfang der Studie und zielte auf einen Vergleich und die generelle Funktion der Distanzmodellierung ab, welche mit einem Anteil von meist über 80 % korrekter Antworten gegeben ist.

### 7.2.2 Versuch 2: Richtungshören mit optischen Markierungen

In diesem Versuch wurde der Absolutbetrag des Azimut-Fehlers in ° bei dem Richtungshören mit einem optischen Raster als Hilfsmittel bestimmt. Die Ergebnisse werden in einem Box-Whisker-Plot dargestellt (Diagr. 9).

Diagramm 9: Azimut-Fehler beim Richtungshören mit optischen Markierungen



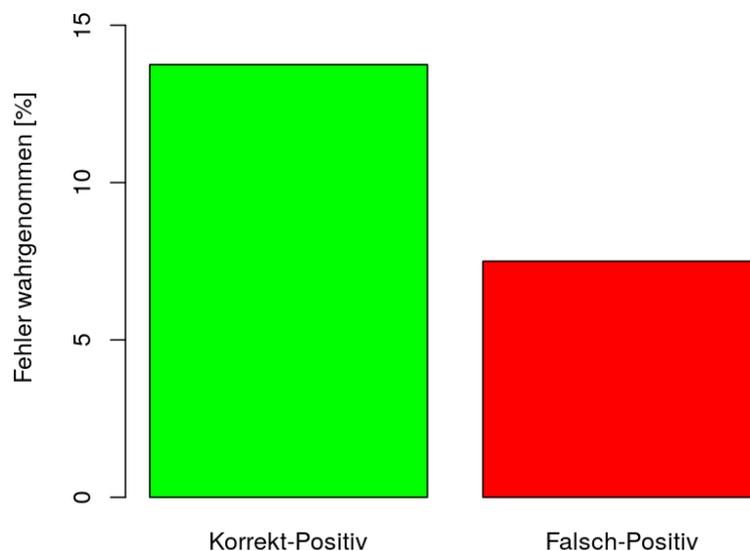
Betrachtet wird der für kleinere Stichproben aussagekräftigere Median (schwarze Linie). Die Variante mit Direktschall (gelb) ist am genauesten und der Fehler bei dem sparsamen Modell für beide Räume (Rot-Töne) und mit dem genauen Modell im Saal (hell-blau) größer. Nur bei dem genauen Modell ist der Fehler im Zimmer (dunkel-blau) vergleichbar. Die Wahrnehmung im simulierten Zimmer (dunklere Farbtöne) ist genauer als in dem Saal (hellere Farbtöne). Für beide Räume gilt, dass der Azimut-Fehler vom sparsamen (Rot-Töne) zum genaueren Modell (Blau-Töne) abnimmt.

An der Größe der Boxen, welche jeweils das Quartil über und unter dem Median enthalten, ist zu erkennen, dass bei dem Direktschall die Streuung hoch ist. Hier ist auch ein großer Ausreißer vorhanden. Eventuell gab es Missverständnisse bei dem Ausfüllen, wie z.B. ein vergessenes Minuszeichen. Der Unterschied bei der Streuung des Azimut-Fehlers zwischen sparsamer und genauer Modellierung im modellierten Saal kann mit

dem Vorhandensein mehrerer Phantomschallquellen erklärt werden, die sich störend auf die Richtungswahrnehmung auswirken.

Ziel dieses Versuchs war es auch, dass der Hinten-Vorne-Fehler untersucht wird. Die Hälfte der Geräusche wurde hinter den Proband\_Innen positioniert und konnte nicht auf dem Monitor gezeigt werden. Deshalb wurde erwartet, dass an dieser Stelle ein *X* in den Fragebogen eingetragen wird. Von dieser Möglichkeit wurde nur sehr selten Gebrauch gemacht, meist wurde eine Richtungsangabe eingetragen. Die abgegebenen Antworten wurden in Korrekt-Positiv (grün) und Falsch-Positiv (rot) aufgeteilt (Diagr. 10).

Diagramm 10: Hinten-Vorne-Fehler beim Richtungshören mit optischen Markierungen



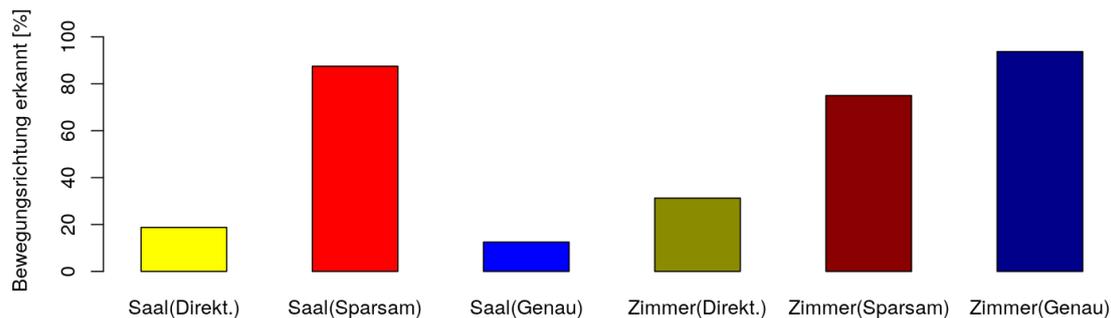
Es wurde im niedrigen zweistelligen Bereich erkannt, dass sich die Geräusche nicht im Bereich des Monitors befinden. Die Rate der fälschlicherweise als nicht im Monitorbereich angezeigten Tonquellen war halb so groß.

Es hätte klarer kommuniziert werden sollen, dass sich nicht alle Tonquellen im Bereich des Monitors befinden. So waren die Proband\_Innen evtl. gewillt, die Aufgabe eher zu beantworten, als die Nicht-Lokalisierbarkeit zuzugeben. Andererseits sind Hinten-Vorne-Fehler schwer ohne eine Kopfbewegung auflösbar, welche z.B. durch Head-Tracking ermöglicht wird. Die Proband\_Innen waren sich nicht sicher genug, dass die Geräuschquelle hinter ihnen ist. Deshalb ist davon auszugehen, dass die gewählte Raumschallmodellierung keine positive Auswirkung auf das Auftreten von Hinten-Vorne-Fehlern hat.

### 7.2.3 Versuch 3: Bewegung, Natürlichkeit und Externalisierung

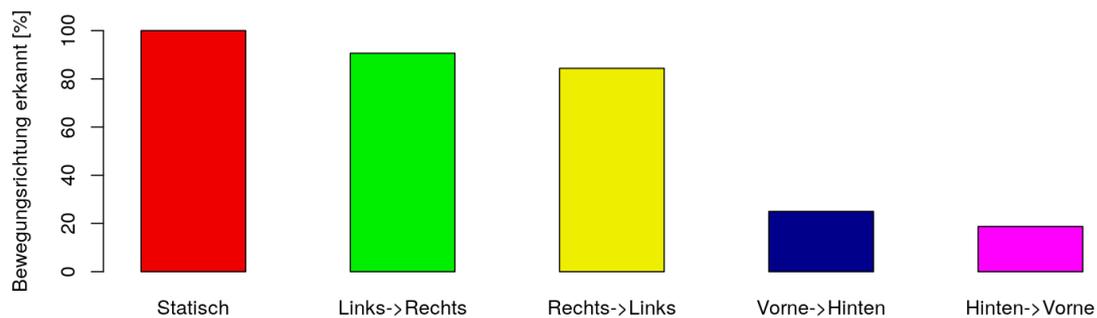
In diesem Abschnitt wurde zu einer bewegten oder unbewegten Tonquelle das Bewegungsmuster, sowie die empfundene Natürlichkeit und Externalisierung erfragt und dabei Bilder von einem Saal und einem Zimmer eingeblendet. Wenn die Bewegungserkennung nach Modell aufgelistet wird, ist kein Zusammenhang zu erkennen (Diagr. 11).

Diagramm 11: Bewegungserkennung nach verwendetem Raum und Modell



Die Modellierung mit Direktschall (gelb), sparsamer Raumschallmodellierung (rot) und genauer Raumschallmodellierung (blau) wird für den simulierten Saal (helle Farbtöne) und das Zimmer (dunkle Farbtöne) dargestellt. Mit dem gewählten Bewegungsmuster als Basis erkennt man, dass die Richtungserkennung nur vom Bewegungsmuster abhängig ist (Diagr. 12).

Diagramm 12: Bewegungserkennung nach gefragter Bewegungsrichtung



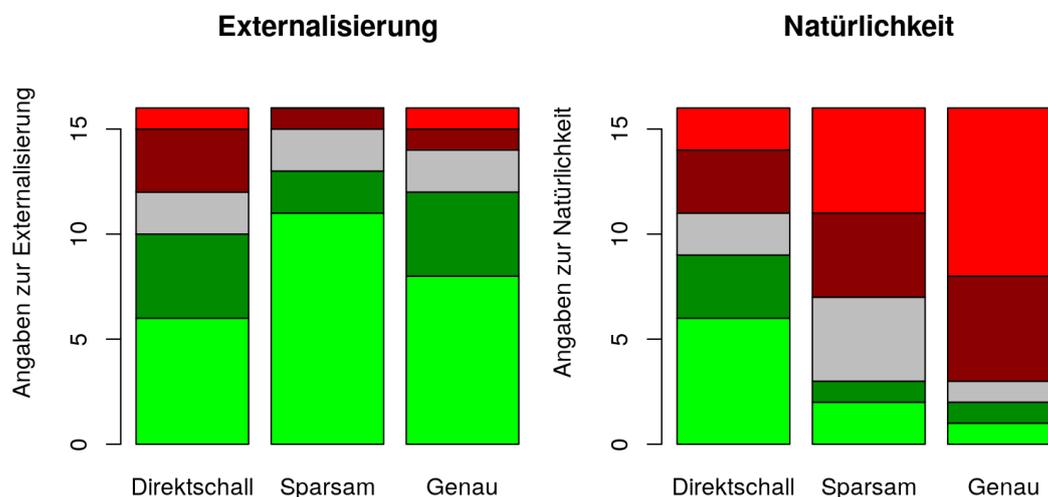
Bewegung zwischen links und rechts (grün und gelb) werden genauso wie die statischen Tonquellen (rot) zuverlässig und Bewegungen zwischen vorne und hinten deutlich seltener korrekt erkannt.

Dies kann mit dem nicht simulierten Doppler-Effekt erklärt werden, welcher bei der sich nähernden oder entfernenden Tonquelle als dynamischer Distanzhinweis fehlt. Eventuell handelt es sich auch um eine nicht eindeutige Definition von vorne und hinten, welche man mit einem Hörbeispiel hätte einführen können.

Die statischen Tonquellen wurden immer als solche erkannt, weshalb sich die Auswertung der subjektiv eingeschätzten Natürlichkeit und Externalisierung auf diese beschränkt.

Die Externalisierung konnte in dem Bereich von *Außerhalb* bis *Innerhalb* angegeben werden. Der Grad der Natürlichkeit reichte von *Natürlich* bis *Künstlich*. Dazu kommt jeweils eine neutrale Position und die beiden Zwischenstufen. Das Farbschema geht von hellgrün für *Außerhalb* bzw. *Natürlich* über dunkelgrün zu grau, um dann über dunkelrot zu hellrot zu gelangen, welches für *Innerhalb* bzw. *Künstlich* steht.

Diagramm 13: Übersicht Externalisierung (links) und Natürlichkeit (rechts)



In dem Diagramm ist ersichtlich, dass die Externalisierung durch die Verwendung eines Raumschallmodells zunimmt, während die Natürlichkeit abnimmt (Diagr. 13). Für die Natürlichkeit ergibt sich eine Reihenfolge von keiner Raumschallmodellierung über die sparsame zur genauen Raumschallmodellierung.

Der fehlende diffuse Nachhall sorgt in beiden Raumschallmodellen für eine künstliche Wahrnehmung. Bei der genauen Modellierung können auch noch die Spiegelschallquellen präzise lokalisiert werden, während das sparsame Modell durch das Panning einen etwas diffuseren Eindruck erzeugt. In beiden Fällen tritt durch das Kopieren des Ausgangssignals und das ständige zeitliche Verschieben durch das MLS-Filter Phasing auf. Der Kontrast zwischen dem diffusen Raumschall und dem präzise ortbaren Raumschall erklärt, warum das sparsame Modell die höchste Externalisierung aufweist.

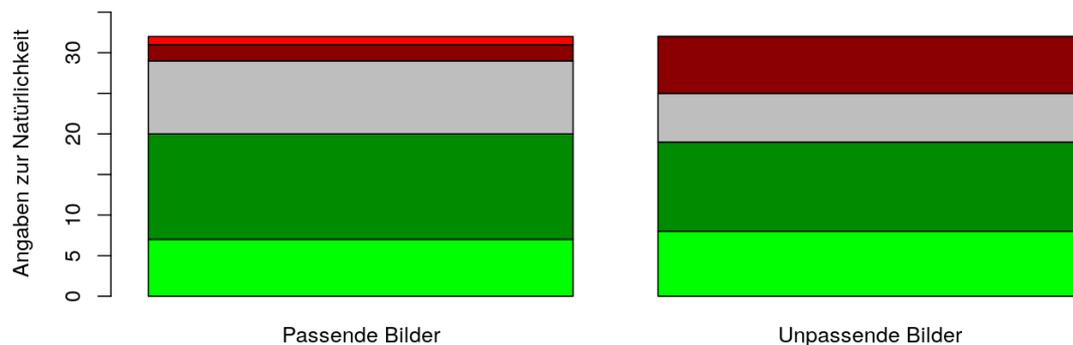
Ziel der Verwendung der Raumschallmodelle ist es, die Immersion durch Externalisierung zu erhöhen. Die Verwendung von alleinigem Direktschall und der sparsamen Raumschallmodellierung unterscheiden sich bei der Externalisierung stark.

Um die Signifikanz dieser Aussage zu bestimmen, wird die als *Außerhalb* wahrgenommene Externalisierung gegenübergestellt. Für den Direktschall gaben sechs Personen (38 %) und für das sparsame Raumschallmodell gaben elf Personen (69 %) diese Wertung. Die Ausgangswerte befinden sich im Anhang. Die Nullhypothese  $H_0$  besagt, dass auch für das Direktschallmodell zu 69 % oder mehr die Wertung *Außerhalb* auftritt. Die Alternativhypothese  $H_1$  lautet, dass der Direktschall eine schlechtere Wertung erzielt.

$$p\text{-Wert} = \sum_{k=m}^N \binom{n}{k} * p^k * (1-p)^{(n-k)} \quad (7.1)$$

Die Gleichung 7.1 wird mit  $m=11$  und  $N=16$  verwendet, um die Wahrscheinlichkeit für das Eintreten der Fälle ( $11 \leq k \leq 16$ ) unter der Auftrittswahrscheinlichkeit  $p = \frac{6}{16} = 0,38$  zu berechnen. Dies ergibt einen  $p$ -Wert von 0.0112. Das gewünschte Signifikanzniveau von  $\alpha = 0,05$  wird erreicht, da ( $\alpha > p\text{-Wert}$ ) gilt. Das Direktschallmodell erzeugt eine signifikant schlechtere Externalisierung als das sparsame Raumschallmodell.

Diagramm 14: Auswirkung der gezeigten Bilder auf die Natürlichkeit



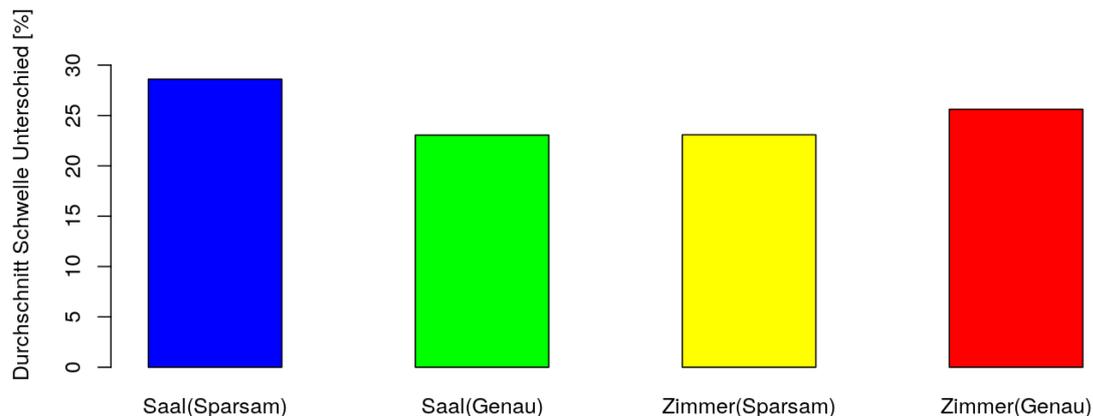
Als Letztes wurde die Auswirkung der Kombination von passendem Raum mit dem Raumschall und unpassendem Raum und Raumschall getestet. (Diagr. 14). Die Farbgebung wurde aus den vorherigen Diagrammen übernommen. Die zwei Teilversuche, in denen Abbildung und Modellierung identisch waren und die zwei Teilversuche bei denen dies nicht so war, wurden zusammengefasst. In dem Diagramm ist zu erkennen, dass das Vertauschen des Bildes keine nachweisbare Auswirkung auf die Natürlichkeit hat.

### 7.2.4 Versuch 4: Bestimmung des Auflösungsvermögens

Mit diesem Teilversuch wurde das Auflösungsvermögen der einzelnen Modellierungen in den verschiedenen simulierten Räumen untersucht. Dazu wurde zweimal das Geräusch Stimme abgespielt. Zwischen der Wiedergabe wurde die Entfernung um 0 %, 5 %, 12,5 % und 33 % variiert. Anschließend wurde erfragt, ob ein Unterschied wahrgenommen wurde. Die maximale Entfernung von 33 % Abweichung entspricht derjenigen, welche im ersten Versuch in Abschnitt 7.2.1 für die Distanztests eingesetzt wurde.

Betrachtet man nur die Entfernungsvariation, dann nahmen die ProbandInnen bei der Stufe mit 33 % Entfernung zuverlässig Unterschiede wahr, bei 12,5 % gab es eine deutlich schlechtere Trefferquote und bei 5 % wurde der Distanzunterschied selten erkannt. Die Werte für die Entfernungsstufen wurden nach Modell und Räumlichkeit sortiert, um die Auswirkung der Modellierungsarten auf das Auflösungsvermögen zu beurteilen. Diese Angaben wurden anschließend zu einem einzigen Wert kumuliert, der dem für die Erkennung nötigen, prozentualen Entfernungsunterschied entspricht (Diagr. 15).

Diagramm 15: Distanzauflösungsvermögen verschiedener Modellierungsarten und Räume

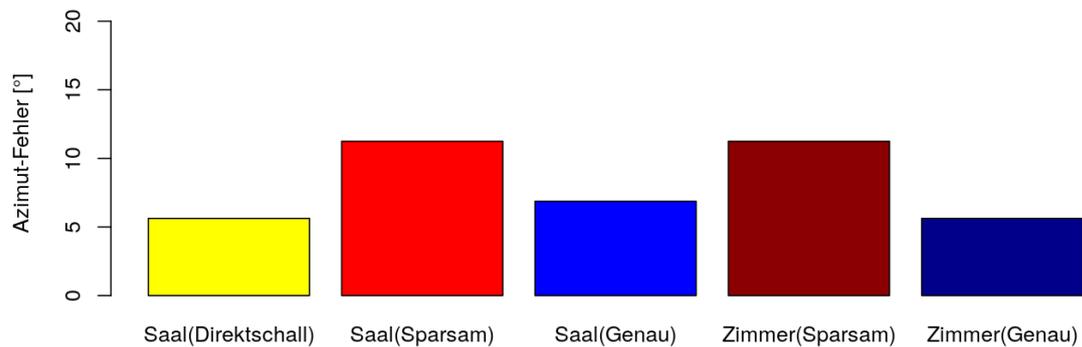


Für den Saal mit sparsamer Raumschallmodellierung ist die nötige Entfernung am größten, für die Modellierung mit dem genauen Modell mit am kleinsten. Für das simulierte Zimmer ist dies für die Raumschallmodellierungen invertiert. Es lässt sich kein Trend für eine bestimmte Räumlichkeit oder Modellierung erkennen. Dies kann daran liegen, dass nicht explizit nach einem Unterschied in der Entfernung gefragt wurde und stattdessen z.B. die Natürlichkeit verglichen wurde. Auch können die Versuchsschallquellen im Hallradius liegen und durch eine Kombination aus Positionierung und Raumgeometrie trotz eines Entfernungsunterschiedes ähnlicher klingen als erwartet.

### 7.2.5 Versuch 5: Tests mit akustischen Markierungen

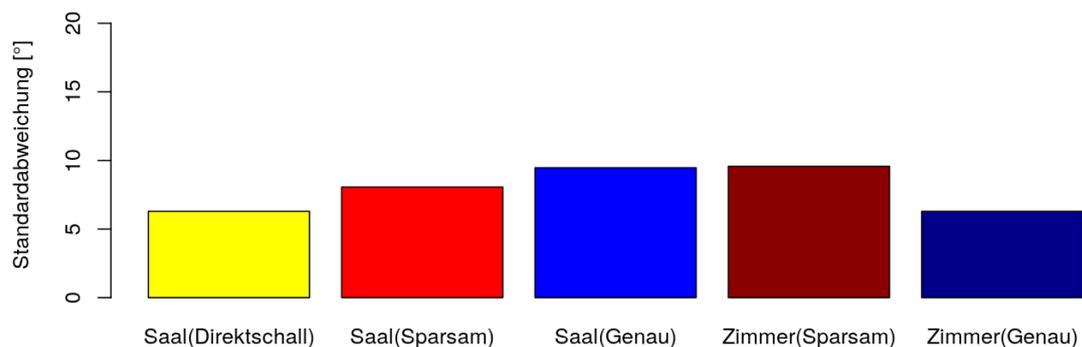
Zu Vergleichszwecken wurde die Messung der Richtungsgenauigkeit mit rein akustischen Mitteln wiederholt. Die Position einer Tonquelle wurde zwischen zwei anderen variiert. Die fünf Stufen der Antwortmöglichkeiten bei einer Schrittweite von  $10^\circ$  ergaben Probleme bei der Deutung des Medians (siehe A.4), weshalb der Mittelwert verwendet wird.

Diagramm 16: Mittelwert des Azimut-Fehlers beim akustischen Richtungshören



In dem Diagramm (Diagr. 16) ist zu erkennen, dass die Annahmen zum Azimut-Fehler aus dem Versuch mit optischen Hilfsmitteln grundsätzlich bestätigt werden. Der Direktschall (gelb) ist genauer als die meisten Varianten mit Raumschall (rot Töne und hellblau). Die Simulation des Zimmers mit dem genauen Raumschallmodell (dunkelrot) ist wieder die einzige Ausnahme. Die Reihenfolge der Raumschallmodelle findet sich ebenfalls wieder.

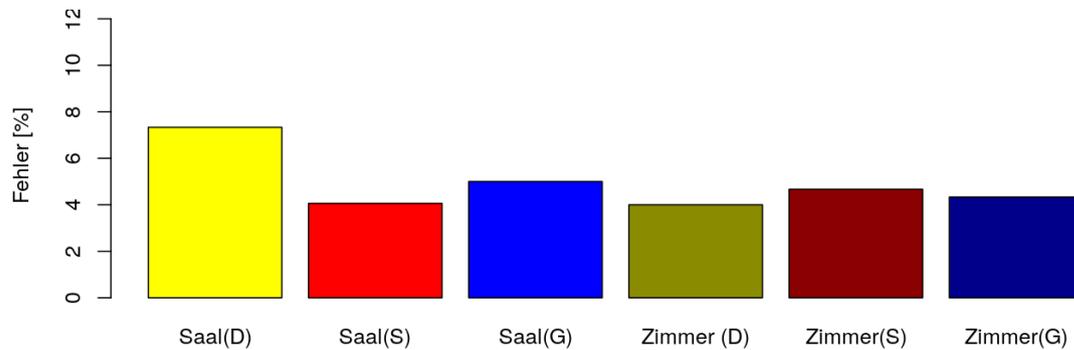
Diagramm 17: Standardabw. des Azimut-Fehlers beim akustischen Richtungshören



Die Standardabweichung ist zwar hoch, aber relativ gleichmäßig verteilt (Diagr. 17). Dies rührt daher, dass ein  $10^\circ$  Raster genutzt wurde und der Azimut-Fehler gering ist.

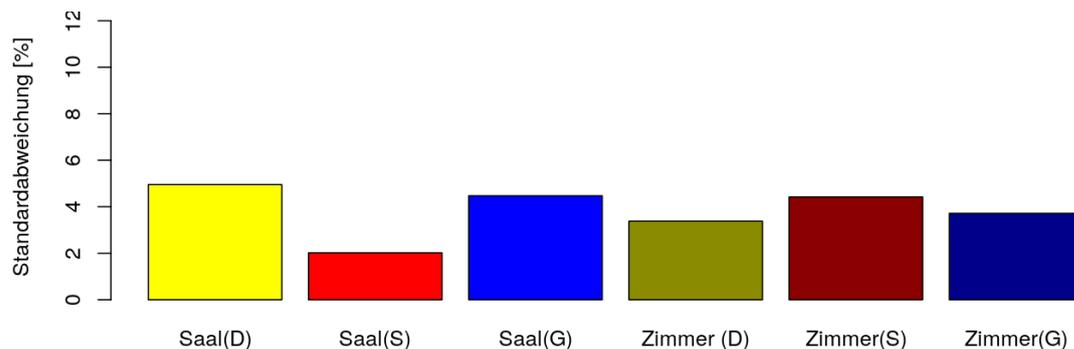
Die Entfernungseinschätzung wurde mit einer ähnlichen Methode untersucht. Die Einteilung der fünf Antwortmöglichkeiten erfolgte in 5% Schritten. Vorher lag der Schwerpunkt auf anderen Aspekten, wie z.B. dem Vergleich der Geräusche untereinander oder der Modelle miteinander. Dieser Teilversuch dient der Messung der Distanzgenauigkeit.

Diagramm 18: Mittelwert des Fehlers bei der akustischen Distanzeinschätzung



Während es für das simulierte Zimmer (dunkle Farbtöne) keine großen Unterschiede zwischen den Modellen gibt, ist der Unterschied für den simulierten Saal (helle Farbtöne) ausgeprägt (Diagr 18). Der Fehler bei der Entfernungsbestimmung mit dem Direktschall (gelb) ist deutlich höher als bei den mit Raumschall versehenen Modellen (rot und blau). Für den Saal, mit den absolut gesehen größeren Entfernungen, wird das erwartbare Ergebnis gezeigt: Das Hinzufügen von Raumschall fördert die Distanzwahrnehmung.

Diagramm 19: Standardabw. des Fehlers bei der akustischen Distanzeinschätzung



In der Standardabweichung sind zwar Schwankungen vorhanden, jedoch gefährden diese die Aussagen des Mittelwertes nicht (Diagr. 19), da diese bis auf einen größeren Ausreißer nach unten relativ gleich verteilt sind.

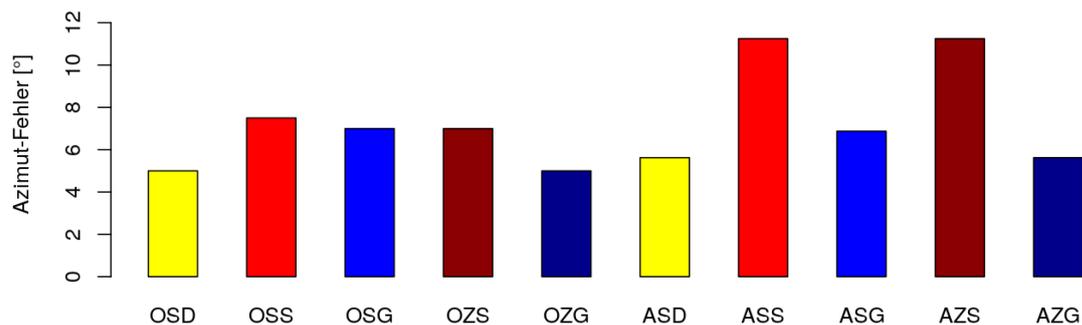
## 7.3 Methodenvergleich und Nebenbetrachtungen

Vorhergehend wurden die Versuche direkt ausgewertet. Nun folgt ein Vergleich der beiden Herangehensweisen. Durch die Analyse der Zusammensetzung der Proband\_Innen ergaben sich außer der Auswirkung der Körpergröße zusätzliche Nebenbetrachtungen.

### 7.3.1 Azimut-Fehler bei akustischen und optischen Markierungen

Es wurde angekündigt, die zwei Herangehensweisen für die Richtungsauswahl zu vergleichen. Die ermittelten Werte sind dabei nicht direkt übertragbar, denn bei der Optik wurde in  $1^\circ$  Schritten abgelesen, während bei Akustik  $10^\circ$  Schritte verwendet wurden. Auch wurde bei dem Versuch mit optischen Hilfsmitteln der Median gewählt und nicht der bei der akustischen Richtungsmessung verwendete Mittelwert.

Diagramm 20: Vergleich der optischen und akustischen Richtungsgenauigkeit

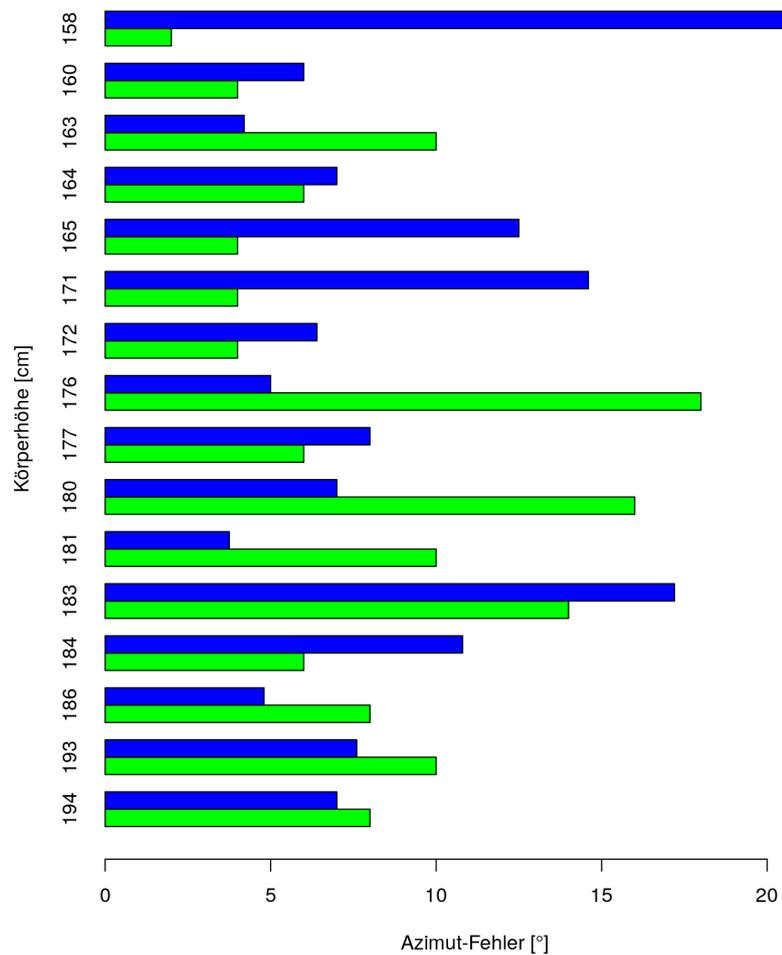


Beide Messungen bestätigen jedoch die Grundaussage, dass die beiden Raumschallmodelle die Richtungsgenauigkeit meist reduzieren (Diagr. 20). Der Versuch mit optischen Hilfsmitteln zeigt einen geringeren Azimut-Fehler und der Unterschied zwischen sparsamer und genauer Raumschallmodellierung ist im Gegensatz zu den akustischen Markierungen weniger stark ausgeprägt. Das folgende Abkürzungsschema wird verwendet: Die Versuche Optisch (O) und Akustisch (A) werden mit dem simulierten Saal (S) oder Zimmer (Z) kombiniert. Daran wird das verwendete Modell, Direktschall (D), sparsame (S) oder genaue (G) Raumschallmodellierung angehängt.

### 7.3.2 Auswirkung der Körpergröße auf die Richtungswahrnehmung

HRTF-Filterung funktioniert am besten, wenn das verwendete Filter dem eigenen Ohr entspricht. Der für die Erzeugung der HRTF-Filter verwendete Kunstkopf entspricht in seiner Gestaltung dem internationalen Durchschnittsmenschen nach IEC: 60318-7. Da die Ohrgröße mit der Körpergröße skaliert, kann eine Korrelation zwischen Azimut-Fehler und Körpergröße auftreten und sich eine Normalverteilung zeigen.

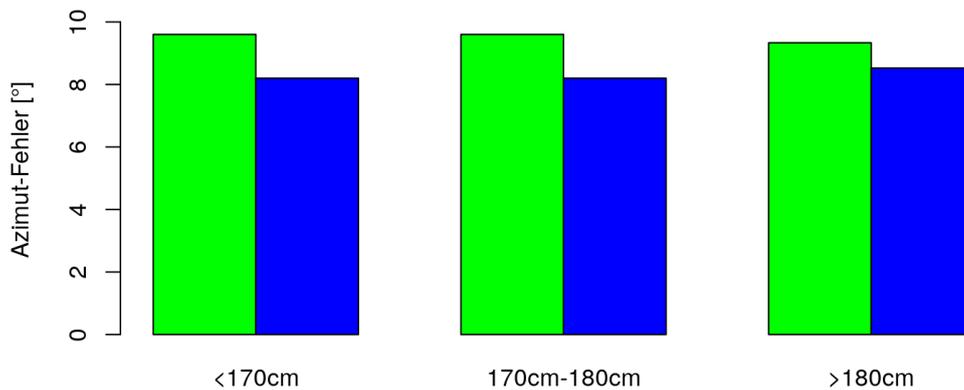
Diagramm 21: Azimut-Fehler bezogen auf die Körpergröße



In dem Diagramm sind auf der Y-Achse die Körpergrößen absteigend eingetragen, die X-Achse entspricht dem Mittelwert des Azimut-Fehlers (Diagr. 21). Die Werte werden für die Teilversuche akustisches Richtungshören (grün) und optisches Richtungshören (blau) getrennt angegeben. Es zeichnet sich keine Normalverteilung ab.

Zur Verdeutlichung wurden um den Durchschnitt der Körpergröße von 175 cm drei 10 cm große Bereiche aufgespannt, in denen jeweils ca. ein Drittel der Proband\_Innen liegt. Im ersten Bereich liegen Körpergrößen unter 170 cm, im zweiten Bereich Proband\_Innen zwischen 170 cm und 180 cm und im letzten Bereich Personen über 180 cm (Diagr. 22).

Diagramm 22: Mittelwert des Azimut-Fehlers bezogen auf Gruppen von Körpergrößen



Die Farben (grün und blau) stehen wieder für die beiden Ermittlungsmethoden des Azimut-Fehlers. Auch bei den gruppierten Körpergrößen ist keine Korrelation mit dem Azimut-Fehler erkennbar.

Dass kein Zusammenhang zu erkennen ist, kann damit begründet werden, dass die HRTF nicht alleine von der Größe der Ohren abhängig ist, welche mit der Körpergröße korreliert. Die HRTF wird auch von der Form der Ohren, dem Ohrabstand und dem Oberkörperbereich bestimmt. Selbst wenn eine Gruppe eine mit dem Kunstkopf vergleichbare Größe der Ohren hat, unterscheidet sich trotz dessen die individuelle HRTF durch die anderen Merkmale, sodass kein Vorteil bei der Lokalisierung vorliegt. Dies gilt auch für die Distanzwahrnehmung, für welche sich ein Diagramm nach Körpergröße geordnet mit erreichter Genauigkeit im Anhang befindet.

### 7.3.3 Effekt von Expertise, Monitor, Audiokonsum und Videospiele

Neben der Körpergröße gab es noch vier weitere Punkte, in denen sich die Proband\_Innen unterschieden: Expertise, Monitorgröße, Audiokonsum und Videospiele-Erfahrung. Für diese Nebenbetrachtungen wurde die Gesamtheit der Teilnehmenden jeweils in zwei Gruppen aufgeteilt, um die Auswirkung eines Merkmals zu untersuchen. Da sich hierdurch die Anzahl der Messpunkte reduziert, wird statt des Mittelwertes der Median verwendet. Die Ergebnisse werden in einer zusammenfassenden Tabelle dargestellt (Tab. 7.1).

Die größte Auffälligkeit zeigt die Aufschlüsselung nach Expertise. Die Gruppe, welche Expertenkenntnisse angab, hat besonders im Versuch mit optischen Hilfsmitteln einen hohen Azimut-Fehler. In dieser nur 25 % der Teilnehmer ausmachenden Gruppe liegen die Ausreißer bei den Versuchen zum Azimut-Fehler.

Tabelle 7.1: Median des Azimut-Fehlers nach Gruppen

Ausgewähltes Kriterium	Betrachtete Teilgruppe	Akustischer Azimut-Fehler	Optischer Azimut-Fehler
Expertise	Expertise	9°	12,4°
	Keine Expertise	6°	7°
Monitorgröße	Großer Monitor	7°	7,8°
	Kleiner Monitor	7°	6,7°
Audiokonsum	Wenig Audio	7°	7,3°
	Viel Audio	7°	7°
Videospiele	3D-Audio	8°	7,8°
	Kein 3D-Audio	7°	6,5°

Bei der Monitorgröße wird zwischen Laptop-Monitoren (<46 cm) und Desktop-Monitoren (>=46 cm) unterschieden. Ein großer Monitor könnte die genauere Einhaltung des Abstandes davor ermöglichen. Auch könnte dieser es erleichtern, die Werte der optischen Markierung präzise abzulesen. Dem ist nicht so, denn Proband\_Innen mit kleinen Monitoren haben einen leicht geringeren Azimut-Fehler als diejenigen mit großen Monitoren. Die Grenze zwischen wenig und viel Audiokonsum wird bei dem Median dieser Kategorie gezogen und liegt bei 165 min. Die Zeit, welche mit dem Konsum von Audioinhalten verbracht wird, führt zu keinem nachweisbaren Unterschied.

In dem Fragebogen konnten Angaben zum Thema Videospiele in ein Freitextfeld eingetragen werden. Dies führte zu einem breiten Spektrum an Antworten. Diese Antworten wurden aufgeteilt in solche, in denen 3D-Audio eine Rolle spielt und in solche, bei denen dies nicht der Fall ist. Die Angaben *VR* und *First-Person* bilden die Gruppe *3D-Audio*, die anderen Eintragungen die Gruppe *Kein 3D-Audio*. Auch hierfür lassen die geringen festgestellten Unterschiede keine Auswirkungen vermuten.

Die Analyse der Teilgruppen zeigt, dass die Unterschiede zwischen den Proband\_Innen bis auf den Grad der Expertise keine großen Auswirkungen auf die Ergebnisse haben. Bei diesen handelt es sich wahrscheinlich um eine zufällige Häufung von Ausreißern. Für die Ergebnisse der Gesamtgruppe bedeutet dies, dass es sich nicht um eine Überlagerung in sich verschiedener Teilergebnisse handelt, sondern dass diese tatsächlich die betrachtete Gruppe repräsentieren.

## 7.4 Zusammenfassung der Ergebnisse und Methodenkritik

Die Studie zeigt, dass die Richtungsgebung mit den gemessenen HRTF-Filtern funktioniert. Auch kann nur durch Lautstärkeanpassung ein leichter Distanzeindruck erzeugt werden. Wird zusätzlich ein Raumschallmodell verwendet, sinkt die Genauigkeit bei der Richtungseinschätzung ab, dafür steigt im Gegenzug die Genauigkeit bei der Distanzeinschätzung in großen Räumen an.

Bei den subjektiven Parametern ist der Unterschied noch deutlicher. Hier sorgen die Raumschallmodelle für eine deutlich höher empfundene Externalisierung. Allerdings wird das so bearbeitete Material auch als deutlich künstlicher empfunden. Die positive Auswirkung der Raumschallmodellierung wird sowohl bei dem subjektiven Eindruck als auch durch Messung der Fähigkeiten bei der Distanzeinschätzung bestätigt.

Wenn man die Modelle vergleicht, zeigt das ressourcensparende Modell mit den geringeren Anforderungen an die Hardware keine Nachteile. Es erzielt bei der Externalisierung und der Natürlichkeit sogar höhere Werte als das genaue Modell. Auch könnte es für höhere Ordnungen des Spiegelschallquellenmodells verwendet werden, ohne den Filteraufwand zu erhöhen.

Tabelle 7.2: Abschließender Vergleich der drei Modellierungsvarianten

Betrachtete Messgröße	Isolierter Direktschall	Sparsame Modellierung	Genaue Modellierung
Azimet-Fehler opt. Markierungen	5°	7,5°	7°
Azimet-Fehler ak. Markierungen	5,6°	11,3°	6,9°
Distanz-Fehler	7,3%	4,1%	5%
Natürlichkeit	38%	12%	6%
Externalisierung	38%	69%	50%

Die Resultate der drei Modellierungsvarianten, also isoliertem Direktschall, sparsamer sowie genauer Raumschallmodellierung, werden der Übersicht halber in einer Tabelle gegenübergestellt (Tab. 7.2). Es wird jeweils nur der simulierte Saal betrachtet, da hier für alle dargestellten Messgrößen Resultate vorliegen. Für den Azimet-Fehler, der mit optischen Markierungen ermittelt wurde, wird der Median dargestellt. Der Mittelwert wird für den mit akustischen Markierungen ermittelten Azimet-Fehler und den Distanz-Fehler verwendet. Die Aussage zur Natürlichkeit bezieht sich auf die prozentuale Angabe der Wertung *Natürlich* und die Aussage zur Externalisierung auf die Wertung *Außerhalb*.

Während der Analyse der Datensätze fiel auf, dass es in vielen Situationen hilfreicher wäre, über mehr Datenpunkte als wenige große Datensätze zu verfügen. Allgemein war es schwierig Proband\_Innen für die Studie zu finden. Besonders schreckten der Zeitaufwand und die Notwendigkeit ab, einen bestimmten Sitzabstand einzuhalten. Die Studie war auf Robustheit ausgelegt, damit trotz der widrigen Umstände Ergebnisse erzeugt werden. Diese Redundanz schlug sich in verschiedenen Methoden für die Beantwortung der selben Fragestellung und einen höheren Zeitaufwand nieder.

Bei manchen Proband\_Innen machte der Fragebogen Probleme und wurde in einem Fall ausgedruckt und eingescannt. Es wurden Schwierigkeiten bei den Aufgaben mit den akustischen Markierungen angemerkt. Diese Probleme fanden sich in der Auswertung allerdings nicht wieder. Zwei Teile der Studie funktionierten nicht so wie erwünscht: So scheiterte die Untersuchung des Hinten-Vorne-Fehlers und auch die Ergebnisse aus dem direkten Vergleich der beiden Raumschallmodellierungen waren nicht sinnvoll einzuordnen. Aussagen zum Vergleich der beiden Modelle konnten aber aus den anderen Teilaufgaben gewonnen werden.

## 8 Verwendung der Erkenntnisse über das erzeugte VR-3D-Audio-Modell

Nachdem die Funktionsweise und der Aufbau des VR-Audio-Modells dargestellt und dessen Fähigkeiten durch eine Nutzerstudie belegt wurden, werden nun Möglichkeiten zur weiteren Arbeit erläutert. Dieser Abschnitt ist dreigeteilt, zuerst werden Überlegungen zur Erweiterung des Raumschallmodells vorgestellt, dann Anpassungen und Erweiterungen der Studie diskutiert und abschließend ein Anwendungsszenario skizziert.

### 8.1 Mögliche Funktionserweiterungen des Modells

Die Studie hat gezeigt, dass die Externalisierung funktioniert, aber gleichzeitig ein künstlicher Eindruck entsteht. Dies könnte durch das Hinzufügen eines diffusen Nachhalls behoben werden. Dieser wäre in der Berechnung nicht aufwändig, könnte die Wirklichkeit aber genauer abbilden. Eine Erhöhung der Ordnung der Spiegelschallquellen könnte hierfür ebenfalls nützlich sein. In einer Studie wurde gezeigt, dass sechs ausgewählte frühe Reflektionen mit diffusen Nachhall einen natürlichen Eindruck erzeugen [4]. Dabei mussten häufig weder Decken noch Böden geometrisch betrachtet werden. Da sich das sparsame Modell gleichwertig zeigte, können ohne gesteigerten Aufwand bei der Filterung acht Spiegelschallquellen verwendet werden.

An der Parametrisierung des Modells könnten Optimierungen vorgenommen werden, indem Materialeigenschaften für Wände und Boden in einer auswählbaren Form zur Verfügung stehen. Die für die Studie verwendeten Absorptionskoeffizienten wurden aus Tabellen nachgeschlagen. Durch die Spezifikation von Materialien könnte auch deren frequenzabhängige Absorption berücksichtigt werden.

Ein durch das Erfordernis vieler Messungen aufwändiger Punkt ist die Integration von HRTFs mit Distanzinformationen für den Nahbereich.

## 8.2 Optimierung und Ausbau der Nutzerstudie

Hauptschwachstelle der Studie war die geringe Teilnehmerzahl, weshalb es sich um eine Vorstudie mit teilweise beschränkter Aussagekraft handelt. Ziel ist es, eher viele als komplexe Datensätze zu erhalten. In dem die Schwelle zur Teilnahme abgesenkt wird, könnte eine höhere Anzahl erreicht werden. So könnte z.B. eine Website mit Audio/Video-Integration und interaktivem Fragebogen verwendet werden. Damit könnten bei Unklarheit Hörbeispiele leichter wiederholt und Zeit gespart werden, wenn keine Wiederholung benötigt wird. Zusätzlich kann in einem interaktiven Format eine Motivation durch Gamifizierung erreicht werden. Gamifizierung beschreibt die Anwendung von Spielregeln und Spielbelohnungen. Nach der Durchführung könnte eine Punktwertung ausgegeben werden. Einige Proband\_Innen äußerten den Wunsch, ihre Ergebnisse zu erhalten.

Auch wäre eine Beschränkung des Zeitaufwands sinnvoll, die Studie dauerte zu lange. In der Studie zeigte sich, dass die Stimme als Geräusch gute Ergebnisse lieferte, sodass man diese ausschließlich einsetzen kann. Auch ist nur noch die Verwendung des sparsamen Raumschallmodells notwendig. Man könnte auch nur den Saal simulieren, da es der relevantere Fall ist. Ebenfalls funktionierte die Arbeit mit akustischen Marken gut, sodass man Monitorinhalte, wenn überhaupt, in der Eingewöhnung einsetzt. Mit der weniger zeitaufwendigen Studie könnte die Auswirkung von höheren Spiegelschallordnungen, die Kombination mit diffusem Nachhall und der HRTF-Filterung mit Distanzinformationen untersucht werden.

## 8.3 Anwendungsszenarien für das erstellte Modell

Neben der Realisierung ähnlich dem ARA-SOC mit FPGA und Softcore-Prozessor, ist auch die Integration in eine für VR geeignete 3D-Engine, wie z.B. Unity denkbar. Im Gegensatz zu den vorhandenen kanalbasierten Lösungen, kann in der Spielengine auf Informationen zur Entfernung der Tonquellen zugegriffen werden. Dazu wäre eine automatische oder manuelle Angabe der Raumdimension und eine Übertragung der verwendeten Materialien für die Absorptionskoeffizienten nötig.

Das Modell ermöglicht eine Entfernungseinschätzung für große Räume, was der Realität in vielen VR-Anwendungen entspricht. In diesen könnte man mit schlechten Lichtverhältnissen oder anderweitig verdeckten Geräuschquellen arbeiten, um eine Anwendung mit einem Fokus auf akustische Lokalisation zu erstellen.

## 9 Fazit

In der vorhergehenden Arbeit wurde das Anwendungsszenario dargelegt, ein Überblick über die Grundlagen und Methodik gegeben und die Implementierung bzw. deren Realisierung dargestellt. Auf Basis dieser Vorarbeiten konnte eine Studie durchgeführt werden, welche bewies, dass durch die Kombination von HRTF-Filterung mit einem Spiegelschallquellenmodell eine Verbesserung der Distanzwahrnehmung erreicht wird. Dabei hat sich die ressourcenschonende Implementierung, welche eine Panning-Komponente besitzt, als geeignete Wahl erwiesen.

Die Studiendurchführung fand während der Anti-Corona-Maßnahmen und den damit einhergehenden Kontaktbeschränkungen statt. Es wurde nicht nur gezeigt, dass eine Studiendurchführung auch aus dem Home-Office möglich ist, sondern auch Erkenntnisse über die Studiendurchführung unter diesen Bedingungen gewonnen. Auch geschah dies nicht im üblichen Rahmen mit den gut zu kontrollierendem Bedingungen in den Räumlichkeiten der HAW. Dafür ist die Studie deutlich näher an der Realität der Anwender\_Innen, denn es wurde nicht hochwertige Labortechnik, sondern stattdessen tatsächlich im Alltag genutzte Audiokomponenten und Kopfhörer verwendet.

Die durch die Anti-Corona-Maßnahmen erhöhte Verbreitung von Videokonferenzen und Telepräsenz zeigt nebenher ein weiteres Anwendungsfeld von räumlichem Echtzeit-3D-Audio auf. In vielen Bereichen ist binaurales 3D-Audio bereits präsent, auch in Windows 10 ist eine Raumklangsimulation integriert <sup>1</sup> und aktuelle Forschung im Rahmen der Microsoft HoloLens [27] zeigt, dass binaurales 3D-Audio für VR- und AR-Systeme weiterhin relevant ist.

---

<sup>1</sup> Informationen Online abrufbar: <https://blogs.windows.com/windowsexperience/2017/11/13/windows-10-tip-make-movies-games-sound-better-windows-sonic/>

# Literaturverzeichnis

- [1] BLAUERT, Jens: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1996. – ISBN 978-0262268684
- [2] BLAUERT, Jens: *Räumliches Hören, 2. Nachschrift - Neue Ergebnisse und Trends seit 1982 bis 1997*. S. Hirzel Verlag, 1997. – ISBN 3-7776-0738-X
- [3] BOUTRON, I u. a.: *Methods of Blinding in Reports of Randomized Controlled Trials Assessing Pharmacologic Treatments: A Systematic Review*. 2006. – URL <https://doi.org/10.1371/journal.pmed.0030425>. – [Online; Abgerufen am 1. Oktober 2021]
- [4] BRINKMANN, Fabian ; GAMPER, Hannes ; RAGHUVANSHI, Nikunj ; TASHEV, Ivan: Towards encoding perceptually salient early reflections for parametric spatial audio rendering. In: *AES convention*, URL <https://www.microsoft.com/en-us/research/publication/towards-encoding-perceptually-salient-early-reflections-for-parametric-spatial-audio-rendering/>, June 2020
- [5] COLEMAN, P.D.: Failure to localize the source distance of an unfamiliar sound. In: *J. Audio Engr. Soc.* 34, 1962, S. 345–346
- [6] D. BEGAULT, M. A.: Direct comparison of the impact of the head tracking, reverberation, and individualized head-related transfer functions on the spartial perception of a virtual speech source. In: *J. Audio Engr. Soc.*, Audio Engr. Soc., 2001, S. 49:904–916
- [7] *Bauakustik - Berechnung der akustischen Eigenschaften von Gebäuden aus den Bauteileigenschaften - Teil 6: Schallabsorption in Räumen; Deutsche Fassung EN*. April 2004. – DIN EN 12354-6:2004-04
- [8] FOHL, Wolfgang ; HEMMER, David: A FPGA-Based Virtual Reality Audio System. In: *Proc. 138th AES Convention*. Warsaw, Poland, 2015

- [9] FU, Zhong-Hua ; XIE, Lei ; JIANG, Dong-mei ; ZHANG, Yan-ning: Fast 3D audio image rendering using equalized and relative HRTFs. In: *Orange Technologies (ICOT), 2013 International Conference on*, March 2013, S. 47–50
- [10] G. MÜLLER, M. M.: *Taschenbuch der Technischen Akustik*. Springer, 2003. – ISBN 978-3540412427
- [11] GARDNER, Bill ; KEITH MARTIN, MIT Media L.: *HRTF Measurements of a KEMAR Dummy-Head Microphone*. May 1994. – URL <http://sound.media.mit.edu/resources/KEMAR.html>. – [Online; Abgerufen am 1. Oktober 2021]
- [12] GERZON, Michael A.: Periphony: With-Height Sound Reproduction. In: *J. Audio Eng. Soc.* 21 (1973), Jan./Feb., Nr. 1/2, S. 2–10
- [13] GERZON, Michael A.: General Metatheory of Auditory Localisation. In: *Audio Engineering Society Convention 92*, URL <http://www.aes.org/e-lib/browse.cfm?elib=6827>, Mar 1992
- [14] GOLOMB, Solomon W.: *Shift Register Sequences*. Laguna Hills, CA, USA : Aegean Park Press, 1981. – ISBN 0894120484
- [15] HAAS, Helmut: *Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache*, Technische Hochschule Braunschweig, Dissertation, 1949
- [16] KILIAN, Dr. U.: *Lexikon der Physik: Hörfläche*. 1998. – URL <https://www.spektrum.de/lexikon/physik/hoerflaeche/6904>. – [Online; Abgerufen am 1. Oktober 2021]
- [17] L. MERZ, H. J.: *Grundkurs der Regelungstechnik - Einführung in die praktischen und theoretischen Methoden*. Oldenbourg Wissenschaftsverlag, 1990. – ISBN 3-486-21603-1
- [18] LARSEN, Camilla H. ; LAURITSEN, David S. ; LARSEN, Jacob J. ; PILGAARD, Marc ; MADSEN, Jacob B.: Differences in Human Audio Localization Performance Between a HRTF- and a non-HRTF Audio System. In: *Proceedings of the 8th Audio Mostly Conference*. New York, NY, USA : ACM, 2013 (AM '13), S. 5:1–5:8. – URL <http://doi.acm.org/10.1145/2544114.2544118>. – ISBN 978-1-4503-2659-9
- [19] MATSMURA, T. ; IWANAGA, N. ; KOBAYASHI, W. ; ONOYE, T. ; SHIRAKAWA, I.: Embedded 3D sound movement system based on feature extraction of head-related transfer function. In: *Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers. International Conference on*, Jan 2005, S. 281–282

- [20] MAURO, Davide A. ; MEKURIA, Rufael ; SANNA, Michele: Binaural Spatialization for 3D Immersive Audio Communication in a Virtual World. In: *Proceedings of the 8th Audio Mostly Conference*. New York, NY, USA : ACM, 2013 (AM '13), S. 8:1–8:8. – URL <http://doi.acm.org/10.1145/2544114.2544115>. – ISBN 978-1-4503-2659-9
- [21] MION, Luca ; AVANZINI, Federico ; MANTEL, Bruno ; BARDY, Benoit ; STOFFREGEN, Thomas A.: Real-time Auditory-visual Distance Rendering for a Virtual Reaching Task. In: *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*. New York, NY, USA : ACM, 2007 (VRST '07), S. 179–182. – URL <http://doi.acm.org/10.1145/1315184.1315217>. – ISBN 978-1-59593-863-3
- [22] MYERS, David G.: *Psychologie*. Heidelberg, Deutschland : Springer, 2008. – 30 S. – ISBN 978-3-540-79032-7
- [23] PULKKI, Ville: Virtual Sound Source Positioning Using Vector Base Amplitude Panning. In: *J. Audio Eng. Soc.* 45 (1997), Nr. 6, S. 456–466. – URL <http://www.aes.org/e-lib/browse.cfm?elib=7853>
- [24] QU, T. ; XIAO, Z. ; GONG, M. ; HUANG, Y. ; LI, X. ; WU, X.: Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. In: *IEEE Trans. Audio Speech, and Lang. Process.*, IEEE, 2009, S. 17:1124–1132
- [25] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (Veranst.), 2018. – URL <https://www.R-project.org/>
- [26] STEIN, Ulrich: *Programmieren mit MATLAB*. 6., neu bearbeitete Auflage. Carl Hanser Verlag, 2017. – ISBN 3446448640, 978-3446448643
- [27] TASHEV, Ivan: Capture, representation, and rendering of 3D audio for virtual and augmented reality. In: *International Journal on Information Technologies and Security* 11 (2019), September, Nr. SP2, S. 49–62. – URL <https://www.microsoft.com/en-us/research/publication/capture-representation-and-rendering-of-3d-audio-for-virtual-and-augmented-reality/>
- [28] THEODOROPOULOS, Dimitris ; CIOBANU, Catalin B. ; KUZMANOV, Georgi: Wave Field Synthesis for 3D Audio: Architectural Perspectives. In: *Proceedings of the 6th ACM Conference on Computing Frontiers*. New York, NY, USA : ACM, 2009 (CF

- '09), S. 127–136. – URL <http://doi.acm.org/10.1145/1531743.1531764>. – ISBN 978-1-60558-413-3
- [29] THEODOROPOULOS, Dimitris ; KUZMANOV, Georgi ; GAYDADJIEV, Georgi: A 3D-audio Reconfigurable Processor. In: *Proceedings of the 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays*. New York, NY, USA : ACM, 2010 (FPGA '10), S. 107–110. – URL <http://doi.acm.org/10.1145/1723112.1723131>. – ISBN 978-1-60558-911-4
- [30] WÖHR, Martin ; DICKREITER, Michael ; DITTEL, Volker ; HOEG, Wolfgang: *Handbuch der Tonstudioteknik*. Walter de Gruyter, 2014. – ISBN 978-3-11-028978-7
- [31] XIE, Bei S. Bo-sun ; XIANG, Ning: Audio Signal Decorrelation Based on Reciprocal-Maximal Length Sequence Filters and Its Applications to Spatial Sound. In: *AES 133rd Convention*, 2012
- [32] ZAHORIK P, Bronkhorst A.: Auditory Distance Perception in Humans: A Summary of Past and Present Research. In: *Acta Acustica united with Acustica 91*, 2005, S. 409–420
- [33] ZÖLZER, Udo: *Digitale Audiosignalverarbeitung*. 3. Auflage. Vieweg+Teubner Verlag, 2005. – ISBN 3519261804, 978-3519261803
- [34] ZÖLZER, Udo: *DAFX: Digital Audio Effects*. 2nd. Wiley Publishing, 2011. – ISBN 0470665998, 9780470665992

# A Anhang

## A.1 Gesprochener Begrüßungstext

Der Text lautet:

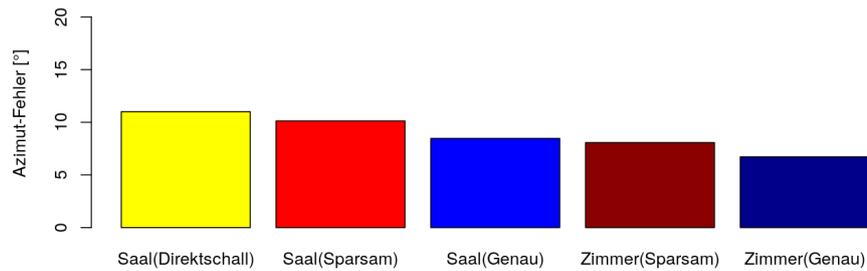
*”Herzlichen Willkommen bei der Studie zum Thema kopfhörerbasiertes 3D-Audio. Ziel ist die Erforschung der Richtungs- und Distanzwahrnehmung von virtuellen Tonquellen. Dazu werden in der kommenden halben Stunde verschiedene Testfälle durchlaufen. Wichtig ist, dass ein Kopfhörer verwendet wird und dieser richtig herum aufgesetzt wurde. Der linke Kanal und der rechte Kanal müssen auf den entsprechenden Ohren wiedergegeben werden. Die Lautstärke sollte so gewählt sein, dass meine Stimme jetzt klar verständlich ist. Das Video ist darauf ausgerichtet, im Vollbild abgespielt zu werden. Außerdem muss für die verschiedenen Testfälle der Abstand zu dem Monitor so gewählt sein, dass dieser der Breite des Monitors entspricht. Ansonsten gibt es Probleme mit der Zuordnung der Bildinhalte mit der Richtung aus der ein Geräusch wahrgenommen wird. Es gibt verschiedene Testblöcke. Am Anfang von jedem dieser Testblöcke gibt es eine kurze Erklärung, um was für eine Aufgabe es sich handelt. Allgemein ein kurzer Hinweis: Das Video lässt sich bei den meisten Abspielprogrammen mit der [Leertaste] pausieren. Wenn nur der Fragebogen und das Abspielprogramm geöffnet sind, dann kann nach dem Anhören einer Teilaufgabe mit der Tastenkombination [Alt] und [Tabulator] zu dem PDF-Fragebogen gewechselt und dieser ausgefüllt werden. Anschließend wechselt man wieder mit [Alt]-[Tab] zurück und setzt das Video mit der [Leertaste] fort. Bitte nach dem Ende des Versuchs nicht vergessen den Fragebogen abzuspeichern. Viel Erfolg und vielen Dank für die Teilnahme!”*

*”Die nachfolgenden Aufgaben sind so angelegt, dass zu Beginn eines Aufgabenblocks eine Erklärung erfolgt. Die Informationen der Anleitung befinden sich jeweils auf den dazugehörigen Fragebogenseiten. In den einzelnen Aufgaben wird entweder nur ein Geräusch abgespielt oder dieses mit Darstellungen kombiniert. Nach der Aufgabe wird jeweils einblendend, dass das Video pausiert werden soll, damit die Antwort in den Fragebogen eingetragen werden kann. Anschließend wird das Video fortgesetzt.”*

## A.2 Mittelwert, Standardabweichung und Median optisches Richtungshören

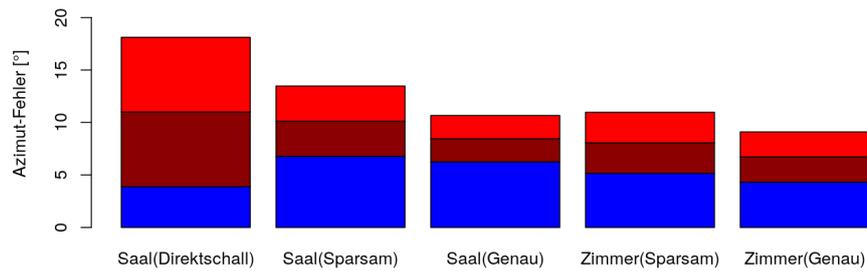
Mittelwert Azimut-Fehler Richtungshören mit optischen Markierungen (Abb. A.1).

Abbildung A.1: Mittelwert des Azimut-Fehlers beim optischen Richtungshören



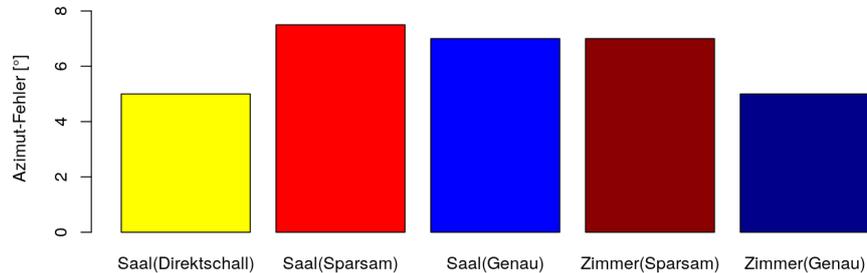
Standardabweich. Azimut-Fehler Richtungshören mit optischen Markierungen (Abb. A.2).

Abbildung A.2: Standardabweichung des Azimut-Fehlers beim optischen Richtungshören



Median Azimut-Fehler Richtungshören mit optischen Markierungen (Abb. A.3).

Abbildung A.3: Median des Azimut-Fehlers beim optischen Richtungshören



### A.3 Externalisierung und Natürlichkeit

Bei den nachfolgenden Diagrammen entspricht 6.25 % eine\_r Proband\_In.

Abbildung A.4: Externalisierung und Natürlichkeit Direktschall

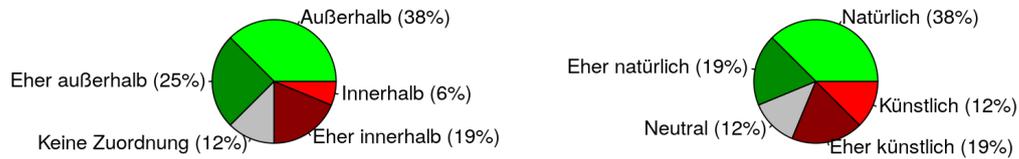


Abbildung A.5: Externalisierung und Natürlichkeit sparsame Modellierung

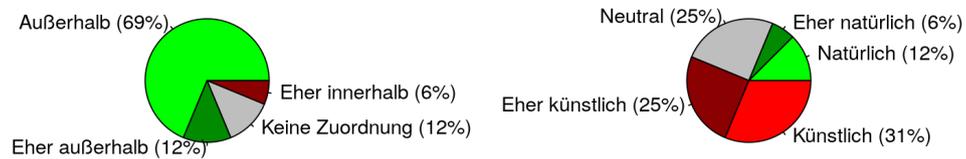
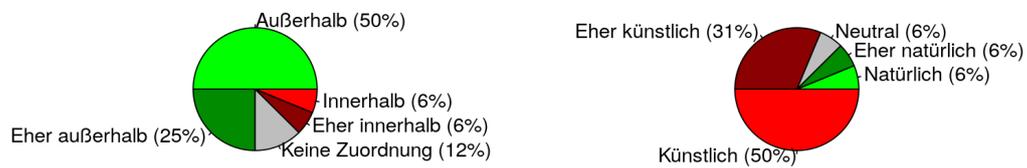


Abbildung A.6: Externalisierung und Natürlichkeit genaue Modellierung



## A.4 Box-Whisker-Plots rein akustische Versuche

Die Box-Whisker-Plots für die rein akustischen Versuche werden dargestellt.

Abbildung A.7: Box-Whisker-Plot akustisches Richtungshören

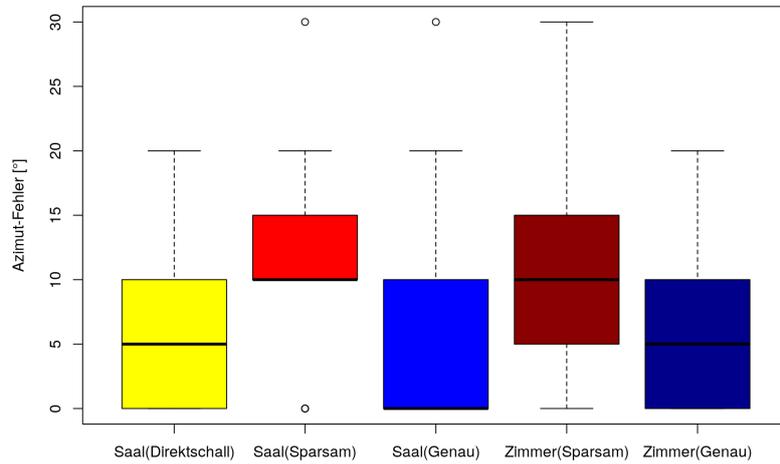
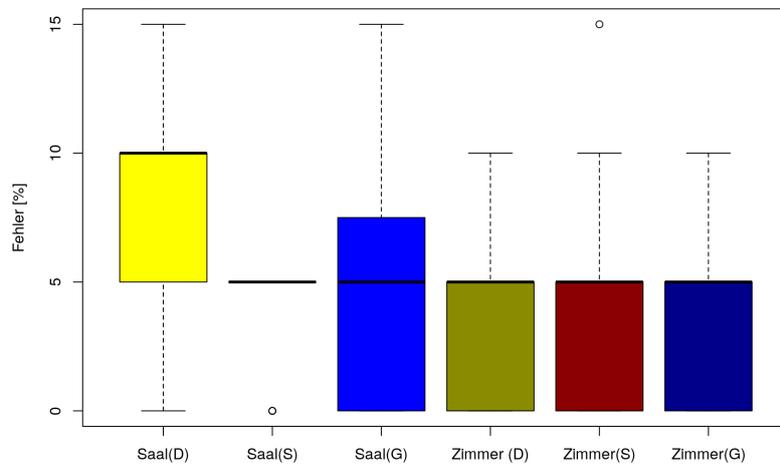
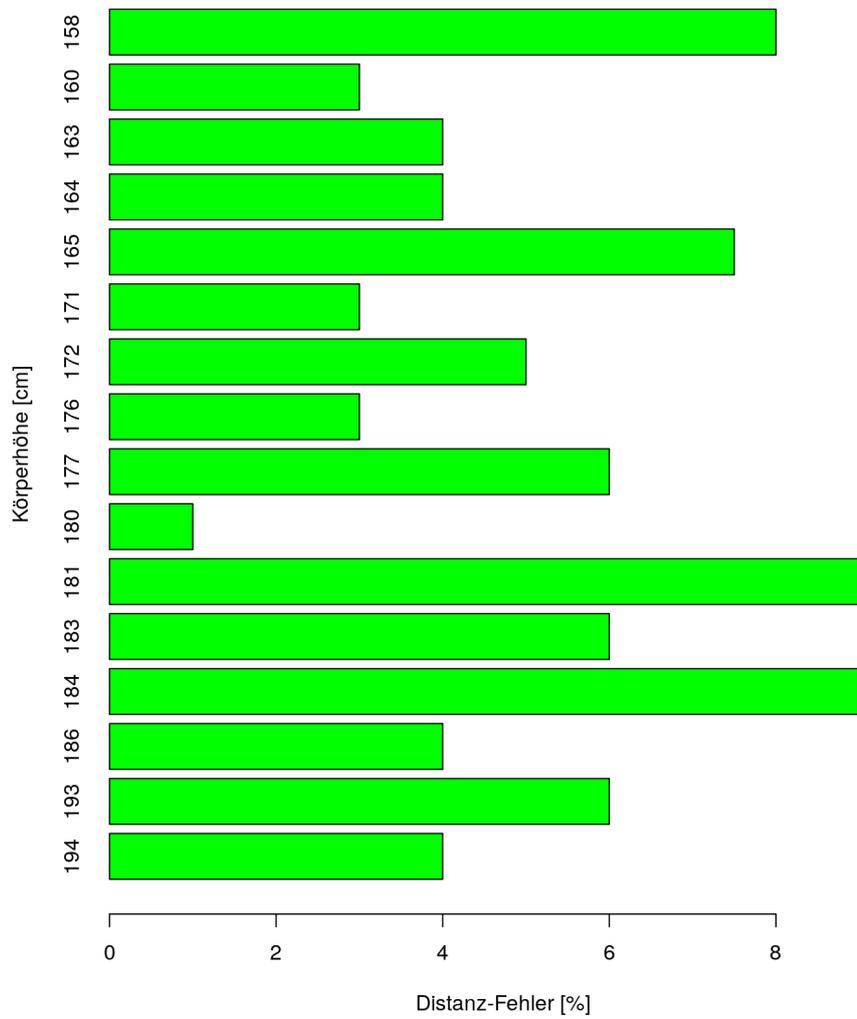


Abbildung A.8: Box-Whisker-Plot akustisches Distanzhören



## A.5 Distanzeinschätzung bezogen auf die Körpergröße

Abbildung A.9: Fehler bei der Distanzeinschätzung bezogen auf die Körpergröße



# Fragebogen zur Studie: Kopfhörerbasiertes 3D-Audio

An der Studie wird teilgenommen, indem das Video angeschaut wird und die Anweisungen im Video beachtet werden. Parallel wird dieser Fragebogen ausgefüllt. Die Felder zu der Person und die Angaben zu der Versuchsumgebung bitte **vor** dem Start des Versuchs ausfüllen.

Dieses Formular kann digital als PDF ausgefüllt werden (z.B. mit Firefox). Falls es analog ausgefüllt wird, bitte den Fragebogen einscannen. Ansonsten: Abspeichern nicht vergessen!

Wenn ein Punkt nicht angegeben werden möchte, dann kann das entsprechende Feld leer gelassen werden. Für die Versuchsdurchführung bitte **Stereo-Kopfhörer verwenden** und mittig vor dem Monitor sitzen. Dabei soll der **Abstand zum Monitor dessen Breite betragen**.

**Die Angaben werden vertraulich behandelt und sind selbstverständlich freiwillig.**

(Es gelten die Bedingungen der Datenschutzerklärung.)

## Persönliche Angaben

Hier bitte die persönlichen Angaben und Informationen zum Musik-/Audiokonsum eintragen.

Alter:  Jahre

Körpergröße:  cm

Geschlecht:  (Männlich, Weiblich, Divers, etc.)

Besonderheiten:  (z.B. Hörgeräte, Erkrankungen, etc.)

Audio pro Tag:  Minuten (Täglicher Konsum z.B. Musik, Hörbücher, etc.)

Expertenkenntnisse:  Vorhanden (Expertenkenntnisse im Bereich 3D-Audio)

Präferierte Audio-Unterhaltung (Mehrfachauswahl möglich):

Musik  Filme/Serien  Hörbücher  Podcasts

Videospielerfahrung:   
(z.B. Keine, Casual, Strategie, Action, Rennspiele, First-Person, VR)

Berufsausbildung:   
(z.B. Studium, Berufsausbildung, Ehrenamt, professionelle Hobbys)

## Angaben zur Versuchsumgebung

Die Informationen zu der Versuchsumgebung sind nötig, da der Versuch nicht in einer kontrollierten Umgebung stattfindet. Bitte füllen Sie diese deshalb vollständig aus.

### Abspielgeräte (Versuchsmittel)

Soundkarte: , Art  (z.B. integriert, extern, PCI)

Kopfhörer: , Art  (z.B. offen, geschlossen)

Abspielsoftware:  (z.B. VLC, Mplayer, Windows Mediaplayer, iTunes)

### Raum (Versuchsumgebung)

Raumgröße:  qm, Bodenbelag:  (z.B. Teppich, Laminat/Parkett, Linoleum, Fliesen)

Uhrzeit:  Uhr, Datum:

Monitorbreite:  cm, Abstand zum Monitor:  cm. **(Bitte die Abstandsanweisung beachten!)**

# Aufgabenteil

In den Aufgaben wird entweder nur ein Geräusch abgespielt oder dieses mit Darstellungen kombiniert. Jede Aufgabe wird nach einer Einblendung wiederholt. Nach der Aufgaben wird eingeblendet, dass das Video pausiert werden kann, um den Fragebogen auszufüllen. Anschließend bitte das Video fortsetzen.

## Abschnitt 1: Einfaches Richtungs- & Entfernungshören

In diesem Abschnitt werden verschiedene Geräusche abgespielt, die dann Objekten auf dem Bildschirm zugeordnet werden sollen. Es gibt erst Aufgaben zum Richtungshören und dann zum Entfernungshören. Bei dem Richtungshören wird eine Geräuschquelle verwendet und einem der beiden dargestellten Objekten zugeordnet. Bei dem Entfernungshören werden zwei Geräusche abgespielt und anschließend soll angegeben werden, welches Geräusch weniger weit entfernt ist. Um die Ausrichtung auf den Monitor zu erleichtern ist die Bildschirmmitte markiert.

**Bitte auf die Mitte des Monitors zentrieren und den Kopf nicht auf die Geräuschquelle ausrichten. Gleichzeitig ist es wichtig, dass der angegebene Abstand zum Monitor eingehalten wird.**

### Richtungshören

Versuch 1: Aktive Tonquelle

A                       B

Versuch 2: Aktive Tonquelle

A                       B

Versuch 3: Aktive Tonquelle

A                       B

Versuch 4: Aktive Tonquelle

A                       B

### Entfernungshören

Versuch 5: Nähere Tonquelle

A                       B

Versuch 6: Nähere Tonquelle

A                       B

Versuch 7: Nähere Tonquelle

A                       B

Versuch 8: Nähere Tonquelle

A                       B

Versuch 9: Nähere Tonquelle

A                       B

Versuch 10: Nähere Tonquelle

A                       B

## Abschnitt 2: Fortgeschrittenes Richtungshören

In diesem Testfall wird eine Klingel abgespielt und anschließend soll deren Richtung angegeben werden. Dazu wird auf dem Monitor ein Raster angezeigt. Zu Beginn der Aufgabe richtet man sich zentral auf den Monitor aus und schließt dann die Augen. Anschließend hört man das Geräusch und zeigt mit dem Finger in die Richtung. Mit der anderen Hand kann man das Video pausieren. Abschließend wird die Gradzahl abgelesen und in den Fragebogen eingetragen. Falls ein Geräusch nicht zugeordnet werden kann, dann bitte einfach ein 'X' eingeben.

**Bitte auf die Mitte des Monitors zentrieren und den Kopf nicht auf die Geräuschquelle ausrichten. Gleichzeitig ist es wichtig, dass der angegebene Abstand zum Monitor eingehalten wird.**

Versuch 1:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 2:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 3:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 4:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 5:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 6:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 7:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 8:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 9:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

Versuch 10:  (Bitte abgelesenen Winkel oder ‚X‘ eintragen.)

### Abschnitt 3: Bewegliche Tonquellen

In diesem Testfall stehen bewegliche Tonquellen und deren Wahrnehmung im Vordergrund. Die Tonquelle steht still, bewegt sich aus dem Bereich vor dem Kopf zum Bereich hinter dem Kopf oder von links nach rechts. Bei den Bewegungen ist jeweils auch die invertierte Richtung möglich.

Diese Bewegungsrichtung soll für die Testfälle angegeben werden. Zusätzlich wird nach der Externalisierung und der Natürlichkeit gefragt. Bei der Natürlichkeit bitte eintragen, ob das Geräusch einen natürlichen d.h. nicht künstlichen oder ungewohnten Eindruck erzeugt hat. Unter Externalisierung wird im Audibereich verstanden, dass ein Geräusch außerhalb des Kopfes wahrgenommen wird. Diesen Eindruck auch bitte für jeden Testfall festhalten.

#### Versuch 1

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

#### Versuch 2

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

#### Versuch 3

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

#### Versuch 4

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

#### Versuch 5

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 6

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 7

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 8

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 9

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 10

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Versuch 11

Bewegungsrichtung	<input type="radio"/> links→rechts	<input type="radio"/> rechts→links	<input type="radio"/> vorne→hinten	<input type="radio"/> hinten→vorne	<input type="radio"/> keine Bewegung
Natürlichkeit	<input type="radio"/> natürlich	<input type="radio"/> eher natürlich	<input type="radio"/> neutral	<input type="radio"/> eher künstlich	<input type="radio"/> künstlich
Externalisierung	<input type="radio"/> außerhalb	<input type="radio"/> eher außerhalb	<input type="radio"/> keine Zuordnung	<input type="radio"/> eher innerhalb	<input type="radio"/> innerhalb

## Abschnitt 4: Bestimmung des Auflösungsvermögens

Dieser Test dient der Erforschung der Fähigkeit zur Unterscheidung der Entfernung von zwei Tonquellen. Dazu wird zweimal das selbe Geräusch abgespielt. Anschließend soll angegeben werden, ob diese sich an der gleichen oder an unterschiedlichen Positionen befinden.

Auflösungsvermögen: 1

identisch     unterschiedlich

Auflösungsvermögen: 2

identisch     unterschiedlich

Auflösungsvermögen: 3

identisch     unterschiedlich

Auflösungsvermögen: 4

identisch     unterschiedlich

Auflösungsvermögen: 5

identisch     unterschiedlich

Auflösungsvermögen: 6

identisch     unterschiedlich

Auflösungsvermögen: 7

identisch     unterschiedlich

Auflösungsvermögen: 8

identisch     unterschiedlich

Auflösungsvermögen: 9

identisch     unterschiedlich

Auflösungsvermögen: 10

identisch     unterschiedlich

Auflösungsvermögen: 11

identisch     unterschiedlich

Auflösungsvermögen: 12

identisch     unterschiedlich

Auflösungsvermögen: 13

identisch     unterschiedlich

Auflösungsvermögen: 14

identisch     unterschiedlich

Auflösungsvermögen: 15

identisch     unterschiedlich

Auflösungsvermögen: 16

identisch     unterschiedlich

## Abschnitt 5: Richtungs-/ Distanzhören mit Referenzpunkten

Bei diesem Test wird dreimal das gleiche Geräusch abgespielt. Das erste und das letzte Geräusch spannen einen Bereich auf in dem das mittlere Geräusch eingeordnet werden soll. Man gibt auf der Skala an, ob sich das Geräusch eher an der Position des ersten Geräusches, hier als 'A' oder an der des letzten Geräusches, hier als 'B' bezeichnet, befindet. Für die **ersten sechs Testfälle** wird die Klingel **einmal pro Position** geläutet, für die **restlichen Testfälle** wird diese **dreimal pro Position** geläutet.

Referenzpunkte: 1

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 2

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 3

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 4

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 5

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 6

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 7

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 8

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 9

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 10

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 11

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 12

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 13

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 14

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 15

A       eher A       in der Mitte       eher B       B

Referenzpunkte: 16

A       eher A       in der Mitte       eher B       B

## Anmerkungen

## Erklärung zur selbstständigen Bearbeitung einer Abschlussarbeit

Gemäß der Allgemeinen Prüfungs- und Studienordnung ist zusammen mit der Abschlussarbeit eine schriftliche Erklärung abzugeben, in der der Studierende bestätigt, dass die Abschlussarbeit „– bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit [(§ 18 Abs. 1 APSO-TI-BM bzw. § 21 Abs. 1 APSO-INGI)] – ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt wurden. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich zu machen.“

*Quelle: § 16 Abs. 5 APSO-TI-BM bzw. § 15 Abs. 6 APSO-INGI*

## Erklärung zur selbstständigen Bearbeitung der Arbeit

Hiermit versichere ich,

Name: \_\_\_\_\_

Vorname: \_\_\_\_\_

dass ich die vorliegende Masterarbeit – bzw. bei einer Gruppenarbeit die entsprechend gekennzeichneten Teile der Arbeit – mit dem Thema:

### **Distanzmodellierung für 3D-Audio zum Einsatz in binauralem Echtzeit-VR**

ohne fremde Hilfe selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

\_\_\_\_\_  
Ort                      Datum                      Unterschrift im Original