

# Spatializing Sound

Ortbarkeit virtueller Soundquellen unterschiedlicher  
Audio Spatializer in Mixed Reality Applikationen

## Masterarbeit

Ausgeführt zum Zweck der Erlangung des akademischen Grades  
**Dipl.-Ing. für technisch-wissenschaftliche Berufe**

am Masterstudiengang Digital Media Production an der Fachhochschule  
St. Pölten, **Masterklasse Audio Design**

von:

**Manuel Schumach, B.Sc.**

mp211508

Betreuer:in: FH-Prof. Dipl.-Ing. (FH) Matthias Husinsky  
Zweitbetreuer:in: Mathis Nitschke

Wien, 13.01.2024

# Ehrenwörtliche Erklärung

Ich versichere, dass

- ich diese Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.

- ich dieses Thema bisher weder im Inland noch im Ausland einem Begutachter/einer Begutachterin zur Beurteilung oder in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Diese Arbeit stimmt mit der vom Begutachter bzw. der Begutachterin beurteilten Arbeit überein.

Wien, 13.01.2024.....

Ort, Datum

*Manuel Schumack*  
.....

Unterschrift

# Kurzfassung

Diese Masterarbeit erforscht die Spatialisierung von Klangquellen in der Mixed Reality (MR), insbesondere im Kontext der Unity Game Engine und die dafür verfügbaren Audio Spatializer. Das Einsatzgebiet birgt weitreichende Anwendungsmöglichkeiten, darunter Hörhilfen für blinde Menschen, Navigationssysteme, Radar- und Überwachungstechnik, Spracherkennung sowie Unterhaltungs- und Bildungsbereiche. Angesichts des wachsenden Mixed Reality Marktes, mit großen Investitionen von Branchengrößen wie Meta und Apple, ist die präzise Lokalisierung virtueller Klangquellen von entscheidender Bedeutung.

Die Arbeit konzentriert sich darauf, wie unterschiedliche Audio Spatializer die Lokalisierung von Klangquellen in Echtzeit beeinflussen. Hierbei spielen nicht nur technische Aspekte, sondern auch die Wahrnehmung und Immersion eine zentrale Rolle. Eine umfassende Literaturrecherche bildet die Grundlage für den Vergleich vier ausgewählter Spatializer, wobei vor Allem auf theoretische Konzepte des "Spatial Hearings" nach Jens Blauert eingegangen wird.

Der Kern der Forschung liegt in zwei Hauptfragen: Erstens, wie präzise lassen sich virtuelle Soundquellen mit unterschiedlichen Audio Spatializern in einer MR-Umgebung (=Mixed Reality) verorten? Um diese Frage zu beantworten, wird ein Hörtest durchgeführt, der die Distanzlokalisierung und die Lokalisierung in horizontaler Ebene an 17 Testpersonen testet. Testsamples werden dabei binaural über Kopfhörer wiedergegeben. Zweitens, sind im Jahr 2023 am Markt verfügbare Audio Spatializer zur Lokalisierung von virtuellen Klangquellen ohne weitere unterstützende Hard/Software wie personalisierte HRTFs in einer MR-Umgebung geeignet? Hierbei werden vier Spatializer anhand definierter Toleranzbereiche und der statistischen Auswertungen eingeordnet.

Die Ergebnisse zeigen, dass der Meta Spatializer tendenziell die präziseste Lokalisierung bietet, gefolgt vom DearVR Spatializer. Die Varianzanalyse zeigt jedoch keine signifikanten Unterschiede zwischen den Spatializern.

Die vorliegende Masterarbeit bietet entscheidende Einblicke für Entwickler:innen von Mixed Reality-Anwendungen, insbesondere hinsichtlich der Auswahl und Integration von Audio Spatializern. Die praxisorientierte Bewertung der Spatializer hebt die spezifischen Stärken und Schwächen hervor, was dazu beiträgt, die Herausforderungen und Potenziale dieser Technologien besser zu verstehen.

# Abstract

This master's thesis explores the spatialization of audio sources in Mixed Reality (MR), particularly within the context of the Unity Game Engine and the available Audio Spatializer tools for it. The research field holds diverse applications, including assistive technologies for blind individuals, navigation systems, radar and surveillance technology, speech recognition, as well as entertainment and educational domains. Given the growing Mixed Reality market, marked by substantial investments from industry giants like Meta and Apple, the precise localization of virtual sound sources is of paramount importance.

The study focuses on how different Audio Spatializers influence real-time localization of audio sources. Not only technical aspects but also perception and immersion play a crucial role in this investigation. A comprehensive literature review forms the basis for the selection and comparison of Spatializers, incorporating theoretical concepts from Jens Blauert's "Spatial Hearing."

The core of the research revolves around two primary questions: Firstly, how accurately can virtual sound sources be localized in an MR environment using different Audio Spatializers? To answer this question, a hearing test is conducted, assessing distance localization and horizontal plane localization on 17 participants. Secondly, are currently available Audio Spatializers, for the localization of virtual sound sources in an MR environment, suitable without additional supporting hardware/software, such as personalized Head-Related Transfer Functions (HRTFs)? Four Spatializers are compared based on defined tolerance ranges and statistical analyses.

The results indicate that the Meta Spatializer tends to provide the most precise localization, followed by the DearVR Spatializer. However, the variance analysis reveals no significant differences between the Spatializers.

The present master's thesis provides crucial insights for developers of mixed reality applications, especially concerning the selection and integration of audio spatializers. The practical evaluation of these spatializers highlights their specific strengths and weaknesses, contributing to a better understanding of the challenges and potentials associated with these technologies.

# Inhaltsverzeichnis

<b>Ehrenwörtliche Erklärung</b>	<b>II</b>
<b>Kurzfassung</b>	<b>III</b>
<b>Abstract</b>	<b>IV</b>
<b>Inhaltsverzeichnis</b>	<b>V</b>
<b>1 Einleitung</b>	<b>8</b>
1.1 Fragestellung und Hypothesen	9
1.2 Struktur der Arbeit	10
1.3 Terminologie	11
1.3.1 VR, AR, MR, XR	11
1.3.2 Lokalisierung	12
1.3.3 Spatial Audio	12
1.3.4 Virtuelle Klangquelle	13
1.4 Abkürzungen	14
<b>2 Hintergrund</b>	<b>15</b>
2.1 Lokalisierung	15
2.1.1 Lokalisierung in horizontaler Ebene	16
2.1.2 Lokalisierung in vertikaler Ebene (Medianebene)	17
2.1.3 Distanzerkennung	17
2.1.4 Lokalisierung mit zwei Lautsprechern	20
2.1.5 Testen der Lokalisierung (Ortsunschärfe)	20
2.1.6 Toleranzbereiche	24
2.1.7 HRTF	24
2.1.8 Lokalisierung durch Kopfbewegung	26
2.2 Audio für Mixed Reality	28
2.2.1 Auralisation	29
2.2.2 Wiedergabe	31
2.2.3 Headtracking	32
2.2.4 Immersion	34
2.2.5 Nachhall in virtuellen Räumen	34
2.2.6 Wahrnehmung virtueller Klangquellen in der MR	35
2.2.7 Herausforderungen in der Mixed Reality	36
2.3 Audio Spatializer	38
2.3.1 Game Engine	39

2.3.2	Audio Middleware	40
2.3.3	Meta Spatializer	42
2.3.4	DearVR Spatializer	45
2.3.5	Steam Audio	47
2.3.6	Project Acoustics	48
2.3.7	Spatializer in Unity	49
<b>3</b>	<b>Testen der Spatializer</b>	<b>50</b>
3.1	Vorstudie	50
3.1.1	Testdesign	50
3.1.2	Durchführung	51
3.1.3	Evaluation	52
3.1.4	Ergebnisse und Anpassungen	53
3.2	Testdesign	55
3.3	MR-Applikation	56
3.3.1	Applikationsdesign	56
3.3.2	Voraussetzungen	58
3.3.3	Unity Projekt	60
3.3.4	3D Model	63
3.3.5	Interface	64
3.3.6	Flussdiagramm	66
3.3.7	Implementierung	67
3.4	Auralisierung	70
3.4.1	Kunstkopfaufnahmen vor Ort	70
3.4.2	Raumakustische Messung	72
3.4.3	Raummessungen durch Audio Spatializer	73
3.4.4	Meta Spatializer	77
3.4.5	DearVR Spatializer	81
3.4.6	Steam Audio	85
3.4.7	Microsoft Acoustics	87
3.4.8	Build	90
3.5	Online-Fragebogen	91
3.5.1	Aufbau	91
3.6	Hörtest	93
<b>4</b>	<b>Evaluation</b>	<b>96</b>
4.1.1	Datenaufbereitung in Unity	97
4.2	Ergebnisse	104
4.2.1	Forschungsfrage 1	104
4.2.2	Forschungsfrage 2	112

4.2.3	Varianzanalyse	113
4.3	Weitere Ergebnisse	115
4.3.1	Online-Fragebogen	115
4.3.2	Lokalisierbarkeit auf Distanz	121
<b>5</b>	<b>Fazit</b>	<b>125</b>
5.1	Ausblick	126
	<b>Literaturverzeichnis</b>	<b>128</b>
	<b>Abbildungsverzeichnis</b>	<b>133</b>
	<b>Tabellenverzeichnis</b>	<b>137</b>
	<b>Anhang</b>	<b>138</b>
	UI Screenshots	138
	Questionnaire	143

# 1 Einleitung

Die Spatialisierung von Klangquellen in der Virtual (=VR), Augmented (=AR) und Mixed Reality (=MR) ist ein aktives Forschungsfeld mit vielen Anwendungsgebieten. (Villegas, 2015, S. 1) Eines der wichtigsten Einsatzgebiete ist dabei die Hörhilfe für blinde Personen, andere Anwendungsgebiete finden sich in Navigationssystemen, Radartechnik, Überwachungstechnik, Spracherkennung, im Militärbereich und der Unterhaltungsbranche. (Liaquat et al., 2021, S. 2)

Diese Arbeit beschäftigt sich mit der Spatialisierung von Klangquellen in der Mixed Reality. Die Erwartungen an den Mixed Reality Markt sind groß. Marktanalysen erwarten bis 2024 hohe Absätze im Augmented Reality Bereich, große Konzerne wie Meta und Apple investieren hohe Summen zur Entwicklung ihrer Mixed Reality Hard und Software. Entwickler:innen von Mixed Reality Anwendungen müssen sich auf die verwendete Hardware verlassen können um Applikationen im Unterhaltungsbereich, vor allem aber auch MR-Anwendungen zur Unterstützung von Menschen mit Behinderung oder etwa Educational Software umzusetzen. (*Mixed Reality Market Report | Size, Share, Growth & Trends (2022-27)*, o. J.)

Das von Blauert als „Spatial Hearing“ bezeichnete auditive Lokalisieren von Klangquellen ist dabei für die Immersion in Mixed Reality Applikationen von großer Bedeutung. (Roginska & Geluso, 2018, S. 5) In MR-Anwendungen müssen virtuelle Klangquellen in Echtzeit möglichst realistisch an die zuhörende Person übermittelt werden. Dies geschieht mithilfe von Audio Spatializern. Sie sind Anwendungen und Plugins die für DAWs (=Digital Audio Workstation), Entwicklungsumgebungen wie Unity<sup>1</sup> oder Unreal Engine<sup>2</sup> oder Audio-Middlewares wie FMod<sup>3</sup> oder Wwise<sup>4</sup> entwickelt werden. Ihre Hauptaufgabe ist es, den Weg einer Klangquelle zur hörenden Person zu berechnen und Parameter zur

---

<sup>1</sup> <https://unity.com/de>

<sup>2</sup> <https://www.unrealengine.com/de>

<sup>3</sup> <https://www.fmod.com/>

<sup>4</sup> <https://www.audiokinetic.com/en/products/wwise/>

Veränderung dieser Übertragung zur Verfügung zu stellen. Die Ausgabe kann dabei je nach Spatializer von der binauralen Wiedergabe mit Kopfhörern bis hin zur Mehrkanalwiedergabe reichen. (Technologies, o. J.-a)

Der Markt für Spatialisierungssoftware ist voller Produkte mit unterschiedlichen Versprechungen und versuchen den Ansprüchen der Entwickler:innen von Mixed Reality Applikationen gerecht zu werden. Diese beinhalten unter anderem eine plausible raumakustische Darstellung virtueller Räume, aber auch die akkurate Lokalisierung virtueller Klangquellen in ebendiesen. Die Entwickler:innen und Sound Designer:innen entscheiden sich während der Entwicklung für das geeignete Produkt. Die Arbeit soll eine Übersicht von Audio Spatializern für die Unity Game Engine vorstellen und diese in einem Hörtest gegenüberstellen.

## 1.1 Fragestellung und Hypothesen

### **Wie präzise lassen sich virtuelle Soundquellen mit unterschiedlichen Audio Spatializern in einer Mixed Reality Umgebung verorten?**

Diese Frage verdeutlicht den Vergleich der Audio Spatializer im Hinblick auf die Lokalisierbarkeit von virtuellen Klangquellen. Mithilfe eines Hörtests kann die Präzision durch die Variablen Distanz und Azimuthabweichung bemessen werden. Dieser Testansatz wurde zuvor von Jens Blauert aber auch von Roginska und Geluso beschrieben. (Siehe 2.1.5 Testen der Lokalisierung (Ortsunschärfe)) Der:die Testpartizipant:in platziert innerhalb einer Mixed Reality Applikation einen Marker dort, wo er:sie das virtuelle Klangobjekt vermutet. Dieser Test wird für alle vier zu testenden Spatializer wiederholt.

Hypothese:

Aufgrund der klanglichen Unterschiede in Bezug auf die Auralisation (wie Erstreflexionen oder Nachhall) einer virtuellen Klangquelle durch einen Audio Spatializer unterscheidet sich die Lokalisierung derselben Klangquelle, unabhängig ihrer spektralen und temporalen Eigenschaften, im Vergleich zu den anderen Spatializern.

### **Sind aktuell am Markt verfügbare Audio Spatializer zur Lokalisierung von virtuellen Klangquellen ohne weitere unterstützender Hard/Software wie personalisierte HRTFs (= head-related transfer functions) in einer Mixed Reality Umgebung geeignet?**

Diese Frage bezieht sich auf gewisse Toleranzbereiche der Variablen Distanz und Azimuthabweichung die die getesteten Spatializer einhalten sollen, um für

Applikationen mit hoher Anforderung an die Lokalisierbarkeit von virtuellen Klangquellen geeignet zu sein. Diese Toleranzbereiche werden durch eine Literaturrecherche definiert. (Siehe 2.1.6 Toleranzbereiche) Die gemessenen Werte aus dem Hörtest müssen sich innerhalb dieser befinden, um die Frage für alle Spatializer mit „Ja“ beantworten zu können.

Hypothese:

Manche Spatializer externalisieren virtuelle Klangquellen besser als andere. Dadurch ist eine im Vergleich bessere Lokalisierung gewährleistet und beeinflusst die Wahl des Audio Spatializers.

## 1.2 Struktur der Arbeit

Zunächst werden Grundbegriffe des Themengebiets „Audio in Mixed Reality“ erklärt.

Danach soll auf den theoretischen Hintergrund des von Blauert beschriebenen „Spatial Hearings“ eingegangen werden, um die Anforderungen an Audio Spatializer zu erörtern. Dieses Kapitel liefert ein grundlegendes Verständnis über die Lokalisierungsfähigkeit des menschlichen Gehörs. In diesem Kapitel soll der Toleranzbereich für die Lokalisierbarkeit von virtuellen Klangquellen definiert werden.

Im nachfolgenden Abschnitt wird eine technisch orientierte Erörterung im Kontext räumlichen Hörens in Mixed Reality Applikationen vorgenommen, mit dem Ziel, eine nahtlose Überleitung zu den Audio Spatializern zu schaffen. Hierbei liegt der Fokus auf den Themen Auralisation und der Wahrnehmung virtueller Klangquellen innerhalb der Mixed Reality. Durch eine detaillierte Betrachtung dieser Aspekte sollen grundlegende Zusammenhänge und technologische Herausforderungen verdeutlicht werden, die in der Entwicklung und Implementierung von Audio Spatializern eine zentrale Rolle spielen. Der Fokus liegt hierbei auf die binaurale Wiedergabe.

Daraufhin werden endlich Audio Spatializer vorgestellt, ihre grundlegende Funktion erläutert und hinsichtlich ihrer Features verglichen. Dieses Kapitel zeigt anhand der Dokumentation der verschiedenen Anbieter die Features und Arbeitsweise der einzelnen Spatializer. Dieses gesamte zweite Kapitel bildet die Basis für die Simulation der Klangkulisse im praktischen Teil der Arbeit.

Das anschließende Kapitel, Kapitel 3, widmet sich dem Hörtest, wobei das Testdesign und die entwickelten Mixed Reality Applikationen anhand der bereits

durchgeführten Vorstudie eingehend erläutert werden. Um eine adäquate Vergleichsbasis zwischen den akustischen Bedingungen, die durch die Audio Spatializer erzeugt werden, und die reale Raumakustik darstellen zu können, werden umfassende Raummessungen in jedem Raum durchgeführt und die resultierenden Daten präsentiert. In diesem Zusammenhang erfolgt eine detaillierte Beschreibung der Integration der Audio Spatializer in die Hörtest-Applikationen. Das Kapitel schließt mit einer umfassenden Darstellung der Durchführung des Hörtests, wodurch ein tiefer Einblick in die methodologischen Schritte und die angewandten Verfahren gewährt wird.

In der Evaluation werden die Ergebnisse des Tests anhand einer statistischen Auswertung präsentiert. Weiters sollen diese Auswertungen die für die Arbeit relevanten Fragestellungen beantworten.

Im Fazit wird ein kritischer Rückblick auf die Arbeit gegeben und ein Ausblick auf nicht abgebildete relevante Themengebiete gegeben werden oder auch Verbesserungen der Methodik vorgeschlagen.

## 1.3 Terminologie

### 1.3.1 VR, AR, MR, XR

Seit den 1960er Jahren wird Virtual Reality für ganz unterschiedliche Techniken (Hardware, wie Software) verwendet. Beispiele dafür sind Online-Virtual-Worlds (z. B. Second Life), Massively Multiplayer Online Role-Playing Games (MMORPGs) wie World of Warcraft<sup>5</sup>, chirurgische Simulatoren, Räume, in denen alle Wände mit Displays bedeckt sind (Cave Automatic Virtual Environments, CAVE<sup>6</sup>), sowie verschiedene Arten von Head-Mounted Displays (HMDs<sup>7</sup>). (Jensen & Konradsen, 2018, S. 2)

Virtual Reality (VR) wird von Stephan Schütze in zwei Ausprägungen beschrieben: Die klassische Virtuelle Realität, in welcher der:die Benutzer:in ein Headset aufsetzt. Dieses ersetzt typischerweise das gesamte Sichtfeld durch eine gerenderte virtuelle Welt. Diese kann entweder durch ein lineares Video oder einen

---

<sup>5</sup> <https://worldofwarcraft.blizzard.com/de-de/>

<sup>6</sup> [https://en.wikipedia.org/wiki/Cave\\_automatic\\_virtual\\_environment](https://en.wikipedia.org/wiki/Cave_automatic_virtual_environment)

<sup>7</sup> [https://en.wikipedia.org/wiki/Head-mounted\\_display](https://en.wikipedia.org/wiki/Head-mounted_display)

interaktiven 3D Raum dargestellt werden. Dabei vereinen sich neben der visuellen Repräsentation auch auditives, narratives oder auch haptisches Feedback.

Das 360 Grad Video wird als weitere Ausprägung der VR beschrieben. Das Video wird üblicherweise mit einem VR-Headset oder günstigen Headset Adaptern für das Smartphone konsumiert. (Schütze, 2018, S. 25–26)

Augmented oder Alternate Reality (AR) vermischt die echte Welt mit virtuellen Inhalten. Die echte Welt wird nicht ersetzt, sondern erweitert (augmentiert). Die Herausforderungen für AR-Audio liegen in der Vermischung zwischen dem, was wir in der echten Welt hören können und den virtuellen hinzugefügten Klangquellen. (Schütze, 2018, S. 27)

Im Rahmen der Definition nach Milgram, wird Mixed Reality als ein Umfeld verstanden, in dem reale und virtuelle Objekte gemeinsam auf einem einzigen Display präsentiert werden. Dies impliziert, dass Mixed Reality sich in einem Bereich zwischen den Extremen der reinen Realität und der reinen Virtualität auf diesem Kontinuum positioniert. Es schafft somit eine Umgebung, in der Elemente aus der realen Welt und virtuelle Elemente in einer integrierten Darstellung miteinander agieren können. (Milgram et al., 1995, S. 2)

### **1.3.2 Lokalisierung**

Die Lokalisierung wird in dieser Masterarbeit als die menschliche Fähigkeit bezeichnet, Soundquellen hinsichtlich ihrer Position zu verorten. Wenzel et al. beschreiben dabei den Vorgang der „auditory spatial perception“ als die menschliche Fähigkeit individuellen Klangquellen in einem 3D Raum zu lokalisieren, auch wenn mehrere simultane Klangquellen präsent sind. (Roginska & Geluso, 2018, S. 10)

### **1.3.3 Spatial Audio**

Spatial Audio bezeichnet Techniken und Formate zur Wiedergabe virtueller Klangquellen in einem speziellen Lautsprecher-setup. Sie beschäftigen sich mit der örtlichen Platzierung der Klangquellen und deren plausible Wiedergabe in Bezug auf den örtlichen Eindruck innerhalb dieser virtuellen Umgebungen. Man spricht hierbei auch von Spatialisierung. Die im virtuellen Raum platzierten Klangquellen werden dann vom Hörenden über ein Lautsprecher-setup oder Kopfhörer wahrgenommen und lokalisiert. (Schütze, 2018, S. 15–23)

### **1.3.4 Virtuelle Klangquelle**

Pulkki definiert eine virtuelle Klangquelle als ein auditives Objekt, das an einem Ort wahrgenommen wird, der keiner physischen Schallquelle entspricht. Typischerweise entsprechen die auditiven Hinweise („spatial cues“) virtueller Quellen nicht den Hinweisen echter Quellen. In der auditiven Forschung sind sie interessant, weil die Art und Weise, wie Hörer solche Hinweise wahrnehmen, menschliche Mechanismen für die räumliche Klangwahrnehmung widerspiegelt.

Wenn die auditiven Hinweise einer virtuellen Klangquelle, denen einer kleinformatigen echten Quelle entsprechen würden, könnte die virtuelle Quelle als "punktförmig" oder "scharf" beschrieben werden. Typischerweise gibt es jedoch Abweichungen in den Hinweisen virtueller Quellen in verschiedenen Frequenzbändern. Eine virtuelle Quelle kann dann als "verbreitet" wahrgenommen werden, das heißt sie ist nicht mehr punktförmig, und die wahrgenommene Größe einer virtuellen Quelle ist größer. In einigen Fällen kann die virtuelle Quelle als "diffus" wahrgenommen werden, deren Richtung dann undefiniert ist. In einigen Fällen erscheint sie innerhalb des Kopfes des Hörers. (Im Kopf Lokalisiertheit – Siehe 2.1 Lokalisierung) (Pulkki, 2001, S. 8)

## 1.4 Abkürzungen

AR	Augmented Reality
BRIR	Binaural Room Impulse Response
CPU	Central Processing Unit
DAW	Digital Audio Workstation
DRIR	Directional Room Impulse Response
DSP	Digital Sound Processor
FPS	Frames Per Second
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HMD	Head-mounted Display
HRTF	Head Related Transfer Function
IDE	Integrated Development Environment
IID	Interaural Intensity Difference
ITD	Interaural Level Difference
3D	Three Dimensional
MR	Mixed Reality
PCVR	Personal Computer Virtual Reality
R/D	Reverberant-to-direct Sound Ratio
SDK	Software Development Kit
VR	Virtual Reality

## 2 Hintergrund

### 2.1 Lokalisierung

Da der Hörtest dieser Masterarbeit die Lokalisierbarkeit von virtuellen Klangquellen testet, ist es notwendig die Lokalisierung des menschlichen Hörens zu verstehen. Dabei greife ich hauptsächlich auf das Werk „Spatial Hearing: The psychophysics of human sound localization“ von Jens Blauert aus dem Jahr 1996 zurück. Dieses verweist selbst auf viele Studien und Hörtests, welche in Bezug auf die Lokalisierbarkeit von Klangquellen in der Vergangenheit durchgeführt wurde. Es dient als Standardreferenz in vielen wissenschaftlichen Ressourcen und soll auch in dieser Masterarbeit als Fundament zur Theorie der Lokalisierung dienen.

Die Lokalisierung oder räumliche Wahrnehmung („Spatial Hearing“ oder „Räumliches Hören“) (Blauert, 1996, S. 38)) von Klangquellen wird zumeist über ein Koordinatensystem mit den Dimensionen des Azimuth (horizontale Richtung), der Elevation (Höhenabbildung) und Distanz (Tiefenabbildung) dargestellt. Im Mittelpunkt des Koordinatensystems befindet sich der Kopf der hörenden Person und blickt vorwärts mit  $0^\circ$  Azimuth und  $0^\circ$  Elevation.

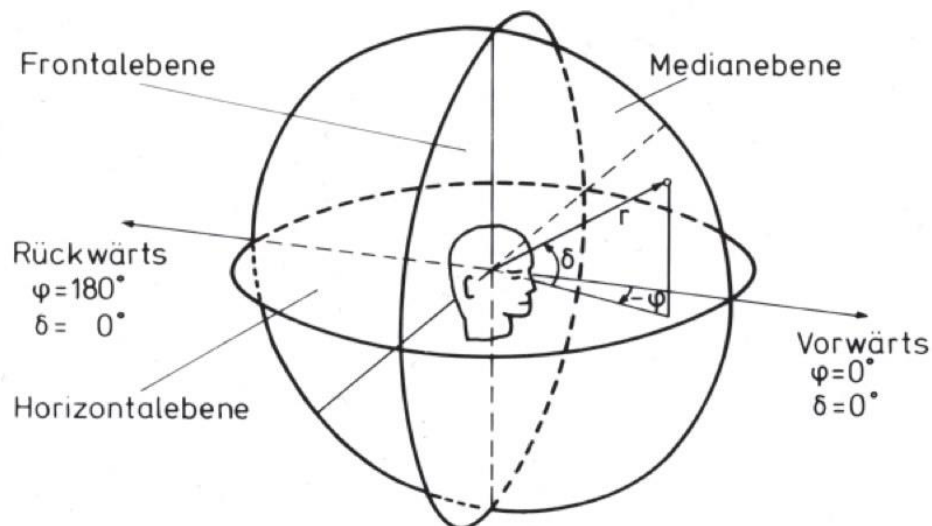


Abbildung 1: Das kopfbezogene Koordinatensystem nach Blauert (Blauert, 1996, S. 11)

## 2 Hintergrund

---

Das Gehör nutzt bestimmte Merkmale der Schallsignale, um die Position und Ausdehnung von Hörereignissen zu bestimmen. Es gibt zwei Arten von "Ohrsignal-Merkmalssklassen": die monauralen und die interauralen Ohrsignalmerkmale. Für Hörversuche werden Beschallungsarten von Blauert und Braasch in monotonisch, ditonisch und dichotisch unterteilt. Wird nur ein Ohr beschallt spricht man von monotonisch. Werden beide Ohren identisch beschallt spricht man von ditonisch. Werden beide Ohren mit unterschiedlichen Signalen beschallt handelt es sich um die dichotische Beschallungsart.

Neben dem Gehör helfen auch andere Sinne wie Seh- und Tastsinn sowie der Lage-Kraft-Richtungssinn des zentralen Nervensystems, Hörereignisse zu bilden. Die Registrierung der Kopfhaltung ist besonders wichtig, da Kopfbewegungen die Ohrsignale in spezifischer Weise ändern und zusätzliche Informationen für das Gehör liefern können. Diese "Peilbewegungen" können jedoch nur bei Schallereignissen länger als etwa 200ms durchgeführt werden. Verantwortlich dafür ist die "Trägheit des Kopfes". Auch die kognitive und emotionale Verfassung eines Hörenden spielt eine Rolle bei der Lokalisierung von Schallquellen. (Weinzierl, 2008, S. 88–89)

### 2.1.1 Lokalisierung in horizontaler Ebene

Die Lokalisation einer Schallquelle hängt von drei Hinweisen ab: interauralen Zeitdifferenzen (ITD), interauralen Pegeldifferenzen (ILD) und monauralen spektralen Hinweisen. Wenn sich eine Schallquelle rechts von einer hörenden Person befindet, trifft die Schallwelle zuerst auf das rechte Ohr und wandert dann einen zusätzlichen Weg um den Kopf herum zum linken Ohr. Die Zeit, die das Signal dabei benötigt, wird ITD genannt. Bei niedrigen Frequenzen (ungefähr 1,5 kHz oder weniger) ist diese Differenz der wichtigste Hinweis für die horizontale Lokalisation. (Middlebrooks, 2015, S. 100–103)

Der Kopf wirft einen akustischen Schatten für Schallwellen mit einer kürzeren Wellenlänge als die Abmessungen des Kopfes. Dadurch ist der Schallpegel am Ohr, das näher an der Schallquelle liegt. Die resultierende interaurale Pegeldifferenz (ILD) kann potenziell als Hinweis auf die Position der Schallquelle dienen. Umgekehrt sind Schallwellen mit Frequenzen unterhalb von ca. 1 kHz Wellenlängen, die größer sind als der Durchmesser des Kopfes, vernachlässigbar. (Middlebrooks, 2015, S. 102)

Die Filterung des einfallenden Signals durch die Pinnae (Ohrmuschel) wird oft als "monaurale spektrale Hinweise" bezeichnet - monaural, weil die erforderliche spektrale Analyse an jedem Ohr individuell erfolgen kann. Spektrale Hinweise

stehen beiden Ohren einer normal hörenden Person zur Verfügung, sie sind also bei der Lokalisierung von  $0^\circ$  und  $180^\circ$  auf der horizontalen Hörebene liegenden Klangquellen erforderlich. Die erfolgreiche Verwendung von spektralen Form-Hinweisen für die Lokalisierung erfordert, dass die Zuhörer mit den Filtereigenschaften ihrer eigenen Ohren vertraut sind und dass die Schallquelle relativ breit im Spektrum ist. (Middlebrooks, 2015, S. 103)

### 2.1.2 Lokalisierung in vertikaler Ebene (Medianebene)

Blauert und Braasch beschreiben für die Lokalisierung auf vertikaler Ebene die sogenannte „Lateralisation“ wie folgt: Im offenen Schallfeld ist es nicht möglich, interaurale Ohrsignalmerkmale einzeln zu untersuchen. Jedoch können diese Merkmale durch dichotische Kopfhörerbeschallung erzeugt werden. In der Regel wird bei solchen Experimenten die Versuchsperson gebeten, nur die seitliche Verschiebung des Hörereignisses von der Medianebene zu schätzen, anstatt den genauen Ort des Hörereignisses zu identifizieren. Aus versuchsmethodischen Gründen wird hierbei zwischen interauralen Pegel und Zeitdifferenzen unterschieden.

Die interauralen Pegeldifferenzen führen über das gesamte Hörspektrum zu seitlichen Hörereignisauslenkungen. Sie sind also zur Verortung eines Signals auf vertikaler Ebene über das gesamte Hörfrequenzband verantwortlich. Die interauralen Zeitdifferenzen von Sinustönen hingegen können nur bis etwa 1,6kHz ausgewertet werden. Darüber kommt es zur Lateralisierung der Hüllkurven und nicht der „Feinstruktur“ des Signals.

Außerdem kann es bei widersprüchlichen interauralen Merkmalen zum Zerfall des Hörereignisses kommen. Tiefe Signalanteile werden woanders verortet als hohe Anteile. (Weinzierl, 2008, S. 96–98)

### 2.1.3 Distanzerkennung

Die Übertragungsfunktionen von Klangquellen ändern sich stark mit der Entfernung und somit auch das Frequenzspektrum des Signals. Dieser Effekt ist auch als binaurale Parallaxe bekannt. (Weinzierl, 2008, S. 98)

Als Entfernungshören wird das Erkennen und Lokalisieren von Klangquellen in der Distanz zur hörenden Person verstanden. Von Blauert und Braasch unterteilen im Handbuch der Audiotechnik das Entfernungshören in vier Entfernungsbereiche:

Bei „sehr geringen Schallquellenentfernungen“ mit einem Abstand zum Kopf von weniger als 25cm steigt die interaurale Pegeldifferenz aufgrund der Schirmwirkung

## 2 Hintergrund

---

des Kopfes. Auch der Pegel des Signals steigt naturgemäß mit der Annäherung. (Weinzierl, 2008, S. 98)

Bei „mittleren Entfernungsbereichen“ zwischen 25cm und 15 Meter tritt die Änderung des Signalpegels bei Punktschallquellen im freien Schallfeld durch 6dB Abfall pro Abstandsverdopplung auf (für Linienquellen 3dB/Abstandsverdopplung). (Weinzierl, 2008, S. 26) Bei bekannten Quellen ist das Gehör dabei in der Lage die Entfernung abzuschätzen. Das Intensitätsverhältnis zwischen dem Direktschall und dem reflektierten Schall bleibt in diesem Abstandsbereich ein relevantes Merkmal zur auditiven Distanzerkennung. Wichtige Hauptmerkmale der Entfernungsbeurteilung sind der Schallquellenpegel und das Intensitätsverhältnis von Direktschall zu reflektiertem Schall. Die Gewichtung dieser Merkmale hängt stark von der Einfallsrichtung und anderen Schalleigenschaften ab. (Weinzierl, 2008, S. 98–99)

Bei „großen Entfernungen“ über 15 Meter kommt es durch den langen Übertragungsweg zu einer „frequenzabhängige Zusatzdämpfung“. Diese wird mit steigender Frequenz intensiviert. Daher klingen etwa fernes Donnerrollen oder Musikkapellen dumpf. Die Entfernung, aus der wir ein Hörereignis wahrnehmen, und die geschätzte Entfernung der Schallquelle sind jedoch nicht dasselbe. Der akustische Horizont des Menschen liegt bei etwa 15 Metern. Wir können zum Beispiel Donnerrollen aus dieser Entfernung wahrnehmen, aber aufgrund von Erfahrungen und visuellen Eindrücken auf größere Entfernungen schließen. (Weinzierl, 2008, S. 99)

Die vierte Unterscheidung, die „Im-Kopf-Lokalisiertheit“, beschreibt das Phänomen, dass ein Hörereignis im Kopf entsteht sobald zwei Schallquellen gleiche oder ähnliche Signale abstrahlen. Dies tritt vor allem bei der Nutzung von Kopfhörern auf. Mit geeigneten Entzerrfiltern lässt sich dieser Effekt jedoch vermeiden. (Weinzierl, 2008, S. 99)

Wenzel et al. beschreibt, dass beim Hören von Musik über Kopfhörer häufig das Klangbild ausschließlich innerhalb oder am Rande des Kopfes erscheint, anstatt dass es außerhalb des Zuhörers zu hören ist. Dieses Phänomen wird als "Inside-the-Head Locatedness" (IHL) oder wie von Blauert beschrieben „In-Kopf-Lokalisiertheit“ bezeichnet und kann als "Externalisierungsfehler" des Systems betrachtet werden. Dieser Fehler tritt insbesondere bei binauralen und 3D-Klangsimulationen über Kopfhörer auf. Wenn sich eine Schallquelle direkt vor dem:der Zuhörer:in befindet sind die binauralen Signale fast ident. Dies entspricht einer diotischen Klangdarstellung, bei der beiden Ohren der gleiche oder fast der gleiche Klang präsentiert wird. Dies ist ein künstlicher Zustand im Vergleich zur

## 2 Hintergrund

---

Wahrnehmung von Schallquellen außerhalb eines Zuhörers, was zu dem kognitiven Schluss führen lässt, dass der Klang im Körperinneren oder in der Nähe des Körpers entsteht, ähnlich der eigenen Stimme. (Roginska & Geluso, 2018, S. 29)

Eine Studie in Bezug auf die 3D-Audio-Simulation von Sprachstimuli von Begault et al. hat gezeigt, dass die Einbeziehung von Nachhall oder Head-Tracking in eine Simulation dazu beiträgt, den internalisierten Klang (die In-Kopf-Lokalisiertheit), zu verringern. (Begault et al., 2000)

Wenzel et al. berichtet außerdem, dass die Distanzinformationen ebenfalls abhängig von der Erwartung oder Vertrautheit einer Schallquelle sind. Insbesondere wenn es sich dabei um Sprache handelt. Selbst bei nicht-räumlichen Signalen handelt man etwas über den Abstand und den räumlichen Kontext erraten. Durch die Verwendung von 3D-Sound-Techniken und der Auralisierung durch einen Hall, kann die Immersion einer 3D Simulation (bspw. VR, AR, ...) sowie die Qualität bei der Darstellung unterschiedlicher virtueller Umgebungen erheblich gesteigert werden.

Bei der Implementierung von Distanzinformationen eines 3D Soundsystem in einer virtuellen Umgebung spielt die kognitive Assoziation der Schallquelle eine gewichtige Rolle. Ist die Schallquelle komplett synthetisch oder unbekannt, so benötigt der:die Hörer:in möglicherweise mehr Zeit sich mit den Veränderungen der Lautstärke oder anderer räumlicher Hinweise ("Spatial Cues") die bei unterschiedlichen Entfernungen auftreten, vertraut zu machen. Hörtests bestätigen, dass es Hörer:innen sehr leicht fällt die Entfernung von Sprache zu bewerten. (Roginska & Geluso, 2018, S. 25)

In Bezug auf Umgebungen mit viel Nachhall beschreibt Wenzel et al., dass sich das Verhältnis zwischen direktem und indirektem Schall in Abhängigkeit von der Distanz zwischen Schallquelle und Zuhörer ändert. Die Reduktion des Schalldrucks nach dem "inverse square law" tritt im akustischen "Nahfeld" einer Schallquelle auf, wo der Pegel des direkten Schalls deutlich höher ist als der des indirekten Schalls. Mit zunehmender Entfernung besteht der Schalldruck bei einem Hörenden immer mehr aus indirektem sowie direktem Schall. An einem bestimmten Punkt erreicht die Schallquelle eine "critical distance" (auch als "Nachhall-Radius" oder "Nachhall-Entfernung" bezeichnet), an der der Pegel des direkten und reflektierten Schalls gleich ist. An Orten jenseits der kritischen Entfernung ist der Gesamtpegel tendenziell gleich, da der Schall bei dem:der Empfänger:in hauptsächlich aus indirektem Schall besteht. (Roginska & Geluso, 2018, S. 27)

### 2.1.4 Lokalisierung mit zwei Lautsprechern

Bei der „Summenlokalisierung“ beschreiben Blauert und Braasch, dass breitbandige Signale wie Sprache oder Musik bei der Wiedergabe durch zwei Lautsprecher durchwegs weniger scharf lokalisierbar sind als bei Einzelschallquellen im freien Schallfeld. Das Gehör bildet aus beiden Ohrsignalen ein Hörereignis. Trifft das Signal vom zweiten Lautsprecher verzögert auf die zuhörende Person ein, so hat dies kaum Einfluss auf die lokalisierte Hörereignisrichtung. (Weinzierl, 2008, S. 101–103)

Dies ist auch als Präzedenzeffekt oder auch Gesetz der ersten Wellenfront bekannt. Das Gesetz der ersten Wellenfront beschreibt den Effekt, dass die Lokalisation von Schallquellen hauptsächlich durch die Richtung des direkten Schalls bestimmt wird, obwohl früh einsetzende Reflexionen aus anderen Richtungen kurz nach dem direkten Schall eintreffen. Wenn Reflexionen später eintreffen oder einen bestimmten Energiepegel überschreiten, werden sie als eigenständiges "sound event" wahrgenommen. Dies wird durch die Echschwelle ("echo threshold") bestimmt. Die Echschwelle kann jedoch nicht als fester Zeitwert angenommen werden, sondern variiert je nach verschiedenen physikalischen Parametern, der Klangquelle und dem (räumlichen) Kontext. (Neidhardt et al., 2022, S. 6)

### 2.1.5 Testen der Lokalisierung (Ortsunschärfe)

Wird die Lokalisierungsfähigkeit von Zuhörern getestet, so wird zumeist die Position einer Schallquelle und deren Unterscheidbarkeit überprüft. Der Grad der Genauigkeit hängt dabei von der Präzision ab, mit welcher die getesteten Personen in der Lage sind Positionen von Klangquellen zu erkennen und zu differenzieren. (Roginska & Geluso, 2018, S. 17)

In diversen Studien zur Lokalisierung und Wahrnehmung von Klangquellen konnten üblicherweise diverse Fehler und Faktoren zur fehlerhaften Einschätzung der Positionen bei der Lokalisierung gefunden werden. Blauert beschreibt 1996 den „Localization Blur“. Diese "Ortsunschärfe" bezieht sich auf die kleinste Veränderung in bestimmten Merkmalen eines Klangereignisses oder eines anderen Ereignisses, die ausreicht, um eine Änderung in der Position des Hörereignisses (z. B. in der Richtung oder Entfernung) zu bewirken. Es ist also eine Eigenschaft der Fähigkeit, Geräusche im Raum zu lokalisieren. Ein Beispiel für Ortsunschärfe ist, wenn sich die Schallquelle seitlich verschiebt und die Richtung des Hörereignisses unscharf wird oder wenn bestimmte

## 2 Hintergrund

---

Frequenzänderungen im Schallsignal zu einer Unschärfe in der Entfernung des Hörereignisses führen. (Blauert, 1996, S. 37)

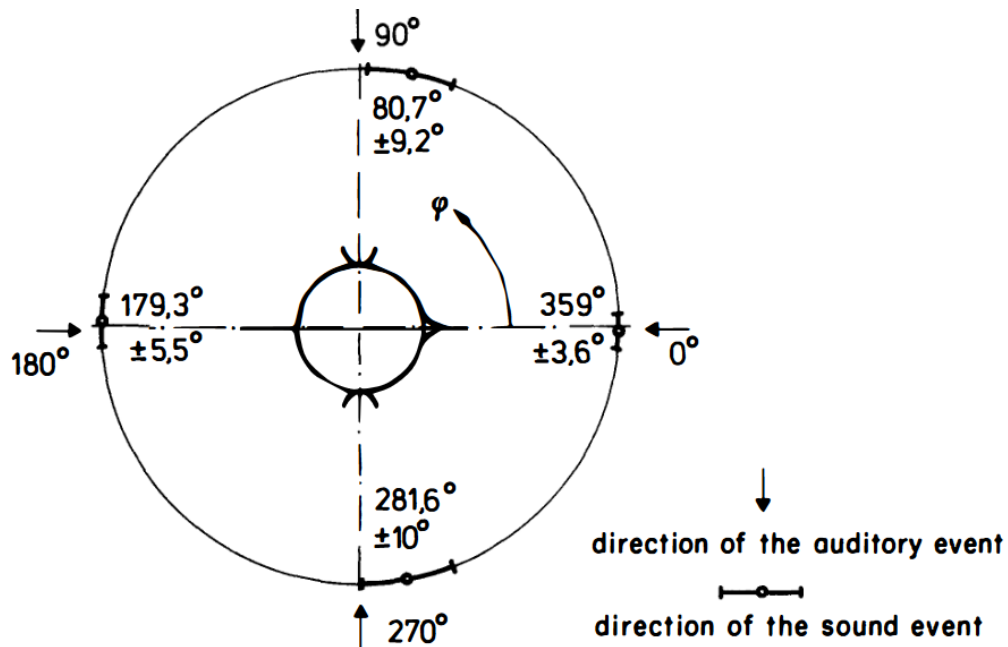


Abbildung 2: Blauert beschreibt die Position der einfallenden Klangquelle als „auditory event“ und die wahrgenommene Richtung als „sound event“. Hier die Ortsunschärfe für die horizontale Hörebene. (Blauert, 1996, S. 41)

Blauert nimmt den maximalen Grad der Lokalisierungsfähigkeit als minimale Ortsunschärfe (=„Localization Blur“) unter optimalen Bedingungen an. Forschungsergebnisse haben gezeigt, dass die Region mit der präzisesten räumlichen Wahrnehmung in oder in der Nähe der Vorwärtsrichtung liegt und dass innerhalb dieser Region eine seitliche Verschiebung der Schallquelle am leichtesten zu einer Veränderung der Position des auditiven Ereignisses führt. (Blauert, 1996, S. 38) Der „Localization Blur“ Fehler in der örtlichen Wahrnehmung kann sich dabei zwischen 5° und maximal 20° bewegen. (Roginska & Geluso, 2018, S. 17) Für ein Signal, welches von vorne auf eine hörende Person eintrifft, kann der Localization Blur auf horizontaler Ebene bei einem Sprachsignal auf 1,5° genau lokalisiert werden. Die Lokalisationsunschärfe nimmt außerdem bei größerer Distanz zum Hörenden zu. Weiters beeinflusst auch die Frequenz des Signals die Lokalisationsunschärfe in unterschiedlichen Richtungen. (Blauert, 1996, S. 42)

Abbildung 2 zeigt die Ergebnisse zur Ortsunschärfe von mehreren Hörtests mit bis zu 900 untrainierten Partizipant:innen. Die Lokalisation wurde auf horizontaler Hörebene (Azimuth) getestet. Ein 100ms langes weißes Rauschen wurde dabei

## 2 Hintergrund

den Testsubjekten aus unterschiedlichen Richtungen vorgespielt. Die Teilnehmer:innen sollten daraufhin einen beweglichen Lautsprecher auf die Position des gehörten „auditory events“ bewegen. Zu sehen ist, dass für ein solches Signal eine Ortsunschärfe bei  $90^\circ$  und  $270^\circ$  bei etwa  $10^\circ$  liegt. Ein Signal von vorne ( $0^\circ$ ) wird mit einer Ortsunschärfe von  $3,5^\circ$  am besten lokalisiert, jenes von hinten ( $180^\circ$ ) bei etwa  $5,5^\circ$  Ortsunschärfe. (Blauert, 1996, S. 41)

Für reine Sinustöne genauso wie für Breitbandsignale gilt, dass die Ortsunschärfe mit der Entfernung der Schallquelle aus der Vorwärtsrichtung ( $0^\circ$ ) zunimmt. Darüber hinaus gibt es eine starke Abhängigkeit von der Frequenz des Signals. Dabei wurden je nach Signal unterschiedliche Ortsunschärfen bei diversen Hörtests gemessen. (Blauert, 1996, S. 42)

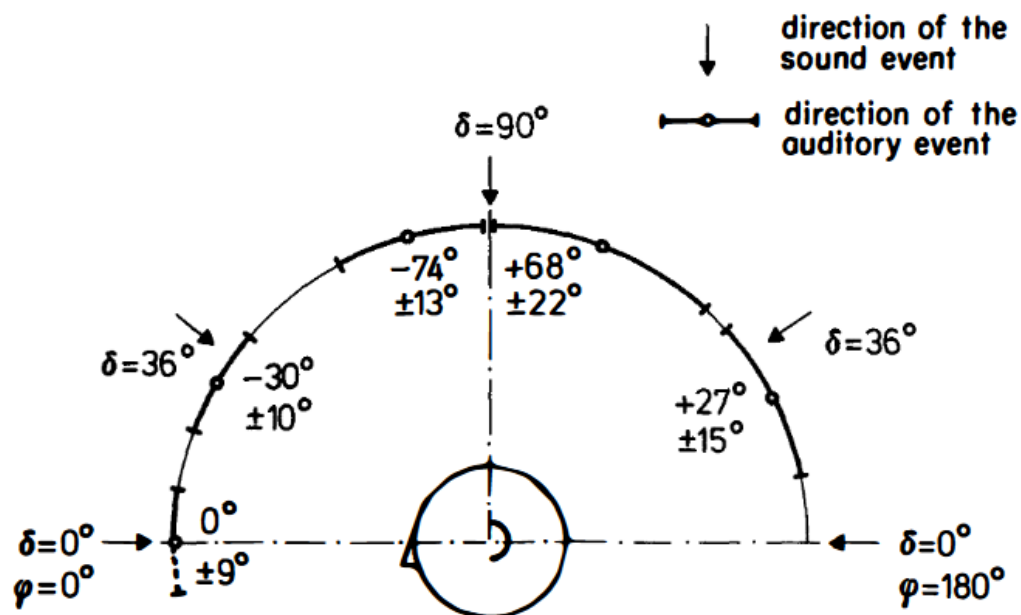


Abbildung 3: Ortsunschärfe auf vertikaler Hörebene (Blauert, 1996, S. 44)

Das Richtungshören in der vertikalen Ebene (auch Medianebene) unterscheidet sich grundlegend von dem in der horizontalen Ebene. Wenn sich die Schallquelle in der Medianebene befindet, sind die Signale in beiden Ohren zunächst identisch. Interaurale Signalunterschiede stehen daher selten zur Verfügung, um bei der Lokalisation der Signale zu helfen. Die Ortsunschärfe bei dem Elevationswinkels  $0$  der Klangquelle in Vorwärtsrichtung beträgt ungefähr  $17^\circ$  für kontinuierliche Sprache durch eine unbekannte Person, etwa  $9^\circ$  für kontinuierliche Sprache durch eine bekannte Person und  $4^\circ$  für weißes Rauschen. In Abbildung 3 ist die Lokalisierungsunschärfe für kontinuierliche Sprache durch eine bekannte Person bei mehreren anderen Einfallswinkeln in der Medianebene zu sehen.

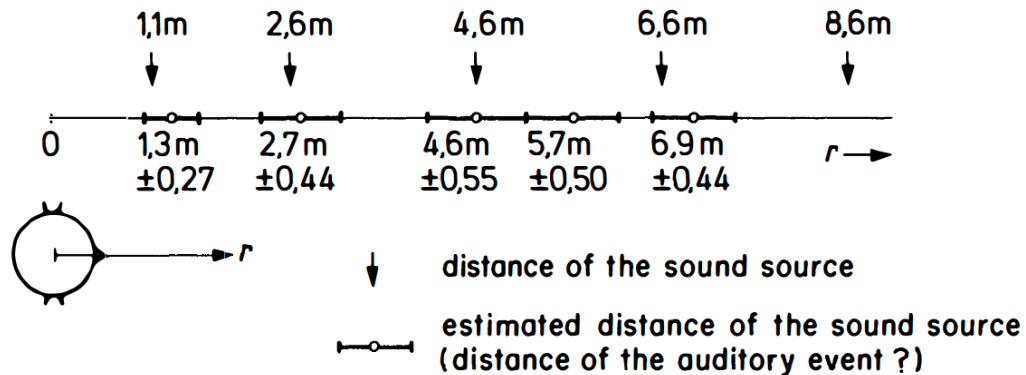


Abbildung 4: Ortsunschärfe bei der Distanzerkennung (Blauert, 1996, S. 47)

Für die Distanzerkennung einer Klangquelle spielt die Vertrautheit der Testteilnehmer:innen mit dem Signal eine wichtige Rolle. Vertrauten Signalen wie die menschliche Sprache können bei normaler Lautstärke kann die Entfernung des auditiven Events recht gut eingeschätzt werden. Es treten jedoch schnell Diskrepanzen auf, wenn sich das Signal beispielsweise in der vorgetragenen Sprechart oder der üblichen Lautstärke unterscheiden. So wird beispielsweise ein Flüstern, welches bei einer Entfernung von 9m zur hörenden Person wiedergegeben wird, dennoch als recht Nahe (< 3m) wahrgenommen. (Blauert, 1996, S. 45–46)

Abbildung 4 zeigt die Ergebnisse eines von Haustein 1969 durchgeführten Hörtests mit Impulsklängen. Blauert beschreibt die Ergebnisse als annähernd genauso gut wie diejenigen, die durch einen Hörtest mit normaler Sprache erzielt wurden. Als Voraussetzung musste das Testsignal jedoch dem Probanden vor dem eigentlichen Hörversuch aus verschiedenen Entfernungen gründlich vorgeführt werden. Die Versuchspersonen wurden gefragt die Entfernung der Schallquellen abzuschätzen. (Blauert, 1996, S. 46)

Für das Entfernungshören, aber auch beim Lokalisieren auf der Medianebene (vertikale Hörebene) kommt es zu einer Verbesserung der Lokalisierung über die Zeit. Klänge die der Versuchsperson bekannt sind, werden besser verortet als unbekannte Klangquellen. Ebenfalls schlechter werden schmalbandige Signale im Vergleich zu breitbandigen (zum Beispiel die menschliche Stimme) verortet. Eine weitere zeitabhängige Eigenschaft des räumlichen Hörens ist die „Persistenz“ oder Trägheit. Darunter wird verstanden, dass Klangquellen, die sich schnell bewegen erst verzögert dort wahrgenommen werden, wo sie sich tatsächlich befinden.

### 2.1.6 Toleranzbereiche

Aus dem von Blauert beschriebenen „Localization Blur“ lässt sich für diese Masterarbeit folgender Toleranzbereich definieren, in welchem eine virtuelle Klangquelle „korrekt“ lokalisiert werden darf. Da bei dem Hörtest Sprache vorgespielt wird, es sich also ein breitbandiges, der hörenden Person bekanntes Signal handelt, nehme ich für die Distanzerkennung eine maximale Distanzdifferenz von  $\pm 0,55\text{m}$  an (Siehe Abbildung 4). Für die horizontale Hörebene nehme ich eine Toleranz für die Winkeldifferenz von  $\pm 3,6^\circ$  an. Weil der:die Testpartizipant:in den Kopf frei bewegen kann, muss nur ein Toleranzbereich für „von vorne“ definiert werden (siehe Abbildung 2).

### 2.1.7 HRTF

Roginska bezeichnet HRTFs (= head related transfer function) als einzigartigen Fingerabdruck („unique fingerprint“) des menschlichen auditiven Systems zur Lokalisierung von Klangquellen. Die Schallwellen, welche von einer Klangquelle ausgesendet werden, werden beim Transfer zum Trommelfell einer hörenden Person durch deren Kopf, Torso und Pinnae gefiltert. Dies führt zu einer richtungsabhängigen Färbung des Klangs. Da sich die Anatomie von Mensch zu Mensch unterscheidet, beschreiben die HRTFs die einzigartige Ausprägung der entsprechenden Filterung und klanglichen Färbung. Zusammen mit der physischen Lokalisierung führt dies zu einer weiteren wichtigen Komponente der Klangquellenlokalisierung. (Roginska & Geluso, 2018, S. 88)

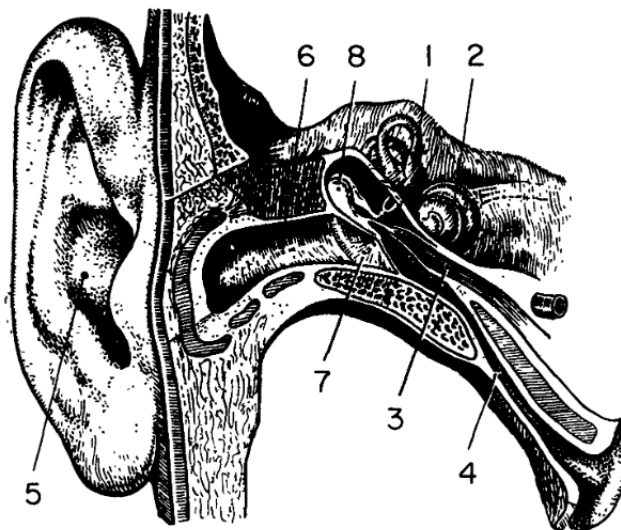


Abbildung 5: Ein Querschnitt des Ohres. „1: Semicircular canals. 2: Cochlea. 3: Eardrum-tensioning muscles. 4: Eustachian tube. 5: Cavum conchae. 6: External auditory canal. 7: Eardrum. 8: Hammer.“ (Blauert, 1996, S. 52)

## 2 Hintergrund

---

Der offene Teil des äußeren Gehörgangs führt in das Cavum conchae. Diese ist der größte Hohlraum der Ohrmuschel. Diese ist also unmittelbar vor dem Gehörgang in das auditorische System eingebunden. Akustisch fungiert die Ohrmuschel als linearer Filter, dessen Übertragungsfunktion von der Richtung und Entfernung der Schallquelle abhängt. Abhängig von der Richtung und Entfernung werden einfallende Schallsignale verändert. Somit „codiert“ die Ohrmuschel räumliche Eigenschaften des Schallfeldes in zeitliche und spektrale Eigenschaften um. Die akustische Wirkung der Ohrmuschel beruht auf verschiedenen physikalischen Phänomenen wie Reflexion, Abschattung, Dispersion, Beugung, Interferenz und Resonanz. Blauert betont, dass darin die Wichtigkeit für das räumliche Hören liegt. (Blauert, 1996, S. 63)

HRTFs können sich stark voneinander unterscheiden. Hörtests, die die Lokalisierung von Klangquellen mit unterschiedlichen HRTFs testeten, haben gezeigt, dass die Lokalisierung durch HRTFs eines „guten Lokalisierers“ sogar besser werden kann. Jedoch konnten auch gegenteilige und gleichbleibende Effekte beobachtet werden. (Roginska & Geluso, 2018, S. 90)

Die Verwendung nicht individualisierter Head-Related Transfer Functions (HRTFs) stellt eine bedeutende Herausforderung bei der Implementierung von 3D-Soundsystemen dar, da die spektralen Charakteristika der HRTFs zwischen individuellen Personen erheblich variieren. Empirische Untersuchungen haben deutlich gezeigt, dass die Präzision der Lokalisierung virtueller Klangquellen signifikant abnimmt, wenn die Wiedergabe durch alternative HRTFs erfolgt. Untersuchungen haben gezeigt, dass bei Personen mit guter Höhenlokalisierung eine Beeinträchtigung dieser Fähigkeit vorliegt, wenn mit HRTFs einer schlechter lokalisierenden Person hört. (Begault, 1994, S. 57–58)

Eine offensichtliche Variation der Leistung zwischen Testsubjekten wird besonders bei nicht individualisierten HRTFs deutlich. Untersuchungen zur Lokalisation von sprachlichen Stimuli zeigten unterschiedliche Ergebnisse abhängig von den Subjekten. Einige Subjekte lokalisierten relativ genau, während andere ihre Urteile eher in Richtung der Ebene, die 90 Grad zwischen links und rechts verläuft, verschoben. Weitere Experimente zeigten ähnliche Ergebnisse, bei denen die Urteile eines Subjekts in ähnlicher Weise verschoben wurden, aber in Richtung der Ebene, die 0 und 180 Grad verbindet. (Begault, 1994, S. 58)

### 2.1.7.1 SOFA

Das Spatially Oriented Format for Acoustics (SOFA) ist ein Dateiformat, das speziell für die Lokalisierung von Klangquellen im dreidimensionalen Raum

## 2 Hintergrund

---

entwickelt wurde. Es ermöglicht die präzise Wiedergabe von Audioinhalten mit räumlicher Tiefe und Positionierung, indem es Informationen über HRTFs einer Person (oder eines Kunstkopfes wie KEMAR<sup>8</sup>) aber auch Binaurale (BRIR = binarual room impulse response) und räumliche (DRIR = directional room impulse response) Raumimpulsantworten.

Eine SOFA-Datei wird üblicherweise wie folgt produziert: In einem schalltoten Raum oder einem Raum werden Erregungssignale erzeugt. Mikrofone zeichnen die eingehenden Signale auf. Die Messung wird wiederholt, wobei die räumliche Position der Erregungsquelle relativ zum Hörer variiert wird, indem die Position des Hörers, der Schallquelle oder beider in verschiedenen Dimensionen verändert wird. Binaurale HRTF-Messaufbauten verwenden lediglich zwei Mikrofone zur Aufzeichnung der Signale für das linke und rechte Ohr. Hingegen können HRTF/DRIR-Messungen auch mehrere Mikrofone berücksichtigen, beispielsweise drei Mikrofone pro Kopfseite in Hörhilfen, zehn Mikrofone in einer Array-Struktur angeordnet in verschiedenen Richtungen und Abständen von der Mitte, ein mehrkanaliges Mikrofonarray um die Zuhörer herum in einem reziproken HRTF-Messsystem, mehrkanalige Mikrofonarrays zur Messung von DRIRs [12] oder verschiedene Mikrofonpositionen in einem Raum, zum Beispiel für Akustikmessungen in Konzertsälen. Diese Informationen ermöglichen es, den Klang entsprechend den räumlichen Gegebenheiten korrekt zu positionieren und eine immersive Klangumgebung zu schaffen.

Durch die Verwendung des SOFA-Formats wird eine standardisierte und interoperable Methode für die Lokalisierung von Klangquellen bereitgestellt. Diese kann von verschiedenen Audiosystemen unterstützt werden. Es erleichtert die Integration von HRTFs und die präzise Positionierung von Audioinhalten, um eine realistische Klangwiedergabe zu ermöglichen. (*General information on SOFA - Sofaconventions*, o. J.)

### 2.1.8 Lokalisierung durch Kopfbewegung

Die Übertragungsfunktionen der äußeren Ohrmuscheln hängen von der Position der Schallquelle ab. Das bedeutet, dass die linearen Verzerrungen, denen Schallsignale auf dem Weg zum Trommelfell ausgesetzt sind, von der Richtung und Entfernung der Schallquelle in Bezug auf den Kopf des Hörers abhängen. Wenn sich der Kopf des Hörers relativ zur Schallquelle bewegt, werden sich die monauralen und interauralen Merkmale des Signals am Trommelfell in einer

---

<sup>8</sup> <http://kemar.us/>

## 2 Hintergrund

---

bestimmten Weise verändern. Theorien des räumlichen Hörens, die die Beziehung zwischen der Position des auditiven Ereignisses und den Veränderungen der Eingangssignale im Ohr während Kopfbewegungen beschreiben, werden als Bewegungstheorien oder motorische Theorien ("motional or motoric theories") bezeichnet.

Bewegungstheorien beschreiben auch die Veränderungen anderer Merkmale auditiver Ereignisse während einer Kopfbewegung, insbesondere Merkmale wie die Lautstärke und/oder die Klangfarbe eines Signals. Zum Beispiel können markante freiwillige Such- und Orientierungsbewegungen des Kopfes bei Personen beobachtet werden, die auf einem Ohr taub sind. Diese Bewegungen dienen offenbar hauptsächlich dazu, die Verwendung der (trainierbaren) Fähigkeit zu erleichtern, die Position der Schallquelle anhand der genannten Veränderungen des auditiven Ereignisses zu ermitteln. (Blauert, 1996, S. 178)

Blauert definiert dabei zwei Arten der Kopfbewegungen für das Lokalisieren einer Schallquelle:

1. Eine mehr oder weniger unbewusste, spontane (reflexartige) Bewegung des Kopfes in Richtung des auditiven Ereignisses und somit in Richtung der wahrscheinlichsten Position der Schallquelle. Die biologische Bedeutung solcher Bewegungen ist offensichtlich, erfordert jedoch eine mehr oder weniger präzise lokalisiertes auditives Event, bevor die Bewegung beginnt. Durch Bewegungen des Kopfes nimmt die Lokalisationsunschärfe ab, da das Klangobjekt in den Bereich des schärfsten Hörens gebracht wird.

2. Such- und Orientierungsbewegungen werden mehr oder weniger bewusst unternommen. Ihr Ziel besteht darin, weitere Informationen zu sammeln, um eine endgültige Position für ein auditives Ereignis zu ermitteln, das anfangs nicht genau lokalisiert ist. In der Regel wird die Lokalisation des auditiven Ereignisses während der Kopfbewegung genauer. In einigen Fällen ändert das auditive Ereignis seine Position auf spezifische Weise, zum Beispiel von vorne nach hinten oder von vorne nach oben vorne. (Blauert, 1996, S. 179)

Wie wichtig die Kopfbewegung für die Lokalisation von Klangquellen ist, zeigt auch folgendes Phänomen: Ist der Kopf fixiert kann dies bei relativ schmalbandigen Signalen dazu führen, dass die auditiven Ereignisse nicht in der Richtung des Schalleintritts ("auditory event") oder auch nur in dessen Nähe erscheinen, sondern systematisch in einer mehr oder weniger achsensymmetrischen Richtung zur Achse der Ohren (eine Linie, die beide Ohren verbindet) liegen. In der horizontalen Ebene kann ein Einfallswinkel von 30° einem auditiven Ereignis bei

150° entsprechen. Wenn die Person ihren Kopf frei bewegen kann und die Signaldauer ausreichend lang ist, um Such- und Orientierungsbewegungen zuzulassen, treten Effekte wie die gerade beschriebenen fast nie auf. Nach ausreichenden Kopfbewegungen erscheint das auditive Ereignis immer in Richtung der Schallquelle. In allen Fällen, in denen die Lokalisierung zunächst ungenau oder verschwommen ist, gewinnen diese Kopfbewegungen eine große Bedeutung. (Blauert, 1996, S. 43–44)

## 2.2 Audio für Mixed Reality

Vorländer charakterisiert im Jahr 2016 die essenziellen Komponenten "Virtueller auditive Displays" der virtuellen Realität (VR). Diese Displays dienen der Darstellung sowie simultanen Wahrnehmung von Realität und ihren physikalischen Eigenschaften in einer interaktiven, computergenerierten virtuellen Umgebung. Eine wesentliche Charakteristik besteht darin, dass sie multimodal gestaltet sind und somit sowohl visuelle als auch auditive Informationen bereitstellen. Zusätzlich besteht die Möglichkeit, weitere Sinne wie Kraftfeedback (Haptik) oder den Tastsinn in die Mensch-Maschine-Schnittstelle zu integrieren. Dies eröffnet Anwendungsbereiche, beispielsweise in architektonischen Kontexten, wie virtuellen Rundgängen durch Gebäudekomplexe, bei denen auditive Informationen ebenso wie visuelle Informationen entscheidend sind, um den Gesamteindruck zu verstärken. Dieses Kapitel befasst sich mit der Technischen Umsetzung dieses virtuellen auditiven Displays, von Vorländer auch als Auralisation beschrieben. (Vorländer, 2016, S. 2)

Gegenwärtig stellen Game Engines die prävalenteste Lösung zur Konstruktion virtueller Szenarien dar, wobei ebenfalls Werkzeuge für die räumliche Darstellung von 3D-Audio existieren. Ungeachtet der individuellen Bemühungen jeder Game Engine, die Leistung ihrer Audio-Engines zu optimieren und die Qualität der Klangwiedergabe zu steigern, existieren gleichzeitig spezifische Tools, die sich explizit auf VR-Audio fokussieren und dabei präzise Ziele verfolgen. In vielen Fällen manifestiert sich dies durch die Bereitstellung von Plugin-Schnittstellen, die speziell für Audio-Spatializer konzipiert sind. (Firat et al., 2022, S. 3)

Wie Garner erläutert, sind Audio Source Development Kits (SDKs), die von VR-Firmen produziert werden (z. B. Oculus Audio SDK<sup>9</sup>, VR Works Audio SDK<sup>10</sup>, 3D

---

<sup>9</sup> <https://developer.oculus.com/documentation/unity/unity-audio/>

<sup>10</sup> <https://developer.nvidia.com/vrworks/vrworks-audio>

## 2 Hintergrund

---

TuneIn Toolkit<sup>11</sup>), Drittanbieter-Plugins (z. B. Steam Audio<sup>12</sup>) und Middleware-Tools, die separate Klangobjekte und -ereignisse erstellen können (z. B. FMOD<sup>13</sup>, Wwise - Wave Works Interactive Sound Engine von Audiokinetic<sup>14</sup>), die drei Hauptoptionen. Es ist auch erwähnenswert, dass es andere Plug-ins gibt, wie Real Space 3D Audio<sup>15</sup>, Csound<sup>16</sup>, Slab3d<sup>17</sup>, Anaglyph<sup>18</sup>, OpenAL Soft<sup>19</sup>, SoLoud<sup>20</sup>, Google Resonance Audio<sup>21</sup>, EVERTims<sup>22</sup>, die von der VR-Audio-Community produziert werden, jedoch unterschiedliche Ziele haben. (Garner, 2018)

Dieses Kapitel und auch die Vorstellung der Audio Spatializer bezieht sich auf Audio Spatializer in Game Engines. Wie unter 2.3 Audio Spatializer beschrieben verarbeiten diese virtuelle Klangquellen in Echtzeit und sind in der Lage sie „spatialisiert“ über ein Mixed Reality Gerät auszugeben.

### 2.2.1 Auralisation

Vorländer beschreibt den Prozess der Auralisierung wie folgt: Sie umfasst die Aufteilung in die Prozesse der Klanggenerierung, Klangausbreitung und Klangwiedergabe. In der Abbildung wird das diskrete Quellsignal  $s(n)$  als "dry" bezeichnet. Das resultierende Signal nach der Klangausbreitung in Räumen,  $g(n)$ , enthält die Eigenschaften sowohl der Klangquelle als auch des Klangausbreitungs- oder Übertragungssystems. Die Eigenschaften eines linearen und zeitinvarianten Klangausbreitungssystems wird durch die Impulsantwort des Systems,  $h(n)$ , beschrieben. Das Klangsignal an der Empfangsposition wird dann durch die Faltung des ursprünglichen "dry" Klangsignals mit der Impulsantwort erzielt (die Impulsantwort wird normalerweise durch einen digitalen linearen Filter dargestellt). (Vorländer, 2016, S. 3)

---

<sup>11</sup> <https://www.3d-tune-in.eu/toolkit-developers>

<sup>12</sup> <https://valvesoftware.github.io/steam-audio/>

<sup>13</sup> <https://www.fmod.com/>

<sup>14</sup> <https://www.audiokinetic.com/en/products/wwise/>

<sup>15</sup> <https://www.ceva-ip.com/product/ceva-realspace/>

<sup>16</sup> <https://csound.com/>

<sup>17</sup> <https://slab3d.sourceforge.net/>

<sup>18</sup> <http://anaglyph.dalembert.upmc.fr/>

<sup>19</sup> <https://github.com/kcat/openal-soft>

<sup>20</sup> <https://solhsa.com/soloud/>

<sup>21</sup> <https://resonance-audio.github.io/resonance-audio/>

<sup>22</sup> <https://evertims.github.io/>

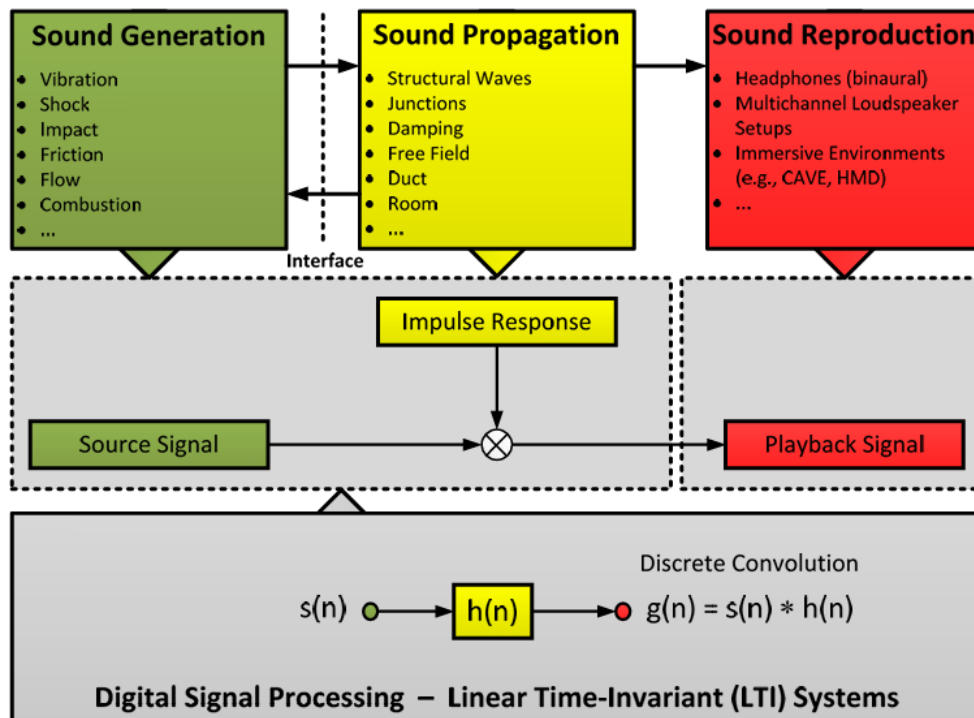


Abbildung 6: Die Blöcke Klanggeneration, Ausbreitung und Wiedergabe der Auralisation nach Vorländer (Vorländer, 2016, S. 3)

Psychoakustisch besteht die Impulsantwort des Übertragungsweges (IR = Impulse Response) aus drei Teilen: dem Direktschall, die Erstreflexionen (Early Reflections) und der späten (oder diffuse) Nachhallzeit. Diese sind für das menschliche Gehör jedoch nicht gleichwertig und unterliegen entsprechend der Wahrnehmung des Klangs individuellen Anforderungen. Geringe Abweichungen in zeitlicher oder spektraler Domäne zwischen Direktschall und Erstreflexionen führen beispielsweise zu einer Änderung der subjektiven Lokalisierung der Klangquelle. Der späte Teil des Impulse Response wiederum wird mit einer viel geringeren zeitlichen Auflösung bewertet. Hier spielt die Gesamtintensität durch Beugung und gestreute Reflexionen in einem bestimmten Zeitfenster energetisch eine Rolle. (Vorländer, 2016, S. 4)

Die Integration dieser Auralisierungskette in ein Mixed-Reality-System stellt eine Herausforderung in Bezug auf die erforderliche minimale Latenz dar. In den verschiedenen Systemen einer Mixed-Reality-Applikation, darunter Tracking, Audio-Hardware, Signalverarbeitung und Audiowiedergabe, wird die zulässige maximale Berechnungszeit sowohl für Akustiksimulationen als auch das Rendering (Wiedergabe) reduziert. Diese Herausforderung wird verstärkt durch die begrenzte Rechenleistung der meisten MR-Headsets. Die Notwendigkeit einer

## 2 Hintergrund

---

Echtzeitverarbeitung erzwingt daher eine erhebliche Verringerung der Komplexität, um den zeitlichen Einschränkungen und den Ressourcenbeschränkungen der MR-Headsets gerecht zu werden. (Vorländer, 2016, S. 5)

Die Echtzeitverarbeitung basiert zumeist auf partitionierter Blockfaltung und verarbeitet die monauralen Audiosignale der virtuellen Quellen mit diesen Filtern. Für jeden Empfänger werden die Signale benachbarter Schallwege summiert. Wenn sich die Schallausbreitung ändert (z. B. Bewegung oder Rotation des Zuhörers), wird die raumakustische Simulation erneut durchgeführt und die Filter oder Teile davon ausgetauscht. Die ungleichmäßige Filterpartitionierung wird gewählt, um die erforderlichen Aktualisierungsraten der Filter für die Anwendung zu unterstützen. (Wefers, 2015)

### **2.2.2 Wiedergabe**

#### *2.2.2.1 Mono*

Wird ein einziger Lautsprecher zur Wiedergabe einer Audioaufnahme verwendet, so entsteht naturgemäß keine Räumlichkeit. Ist in der Aufnahme jedoch ausreichend Rauminformationen enthalten, so führt dies wenigstens zu einer Wahrnehmung in der Tiefe und der Größe des Raumes. Unter Rauminformationen versteht man hierbei Wandreflektionen, Raumnachhall und durch Abstände bestimmte Filterungen der Klangquelle. Das erreichte Klangbild im Sinne einer Spatialisierung wird bei monophonen Systemen jedoch stets nur auf diesen einen Lautsprecher begrenzt bleiben. (Roginska & Geluso, 2018, S. 63–64)

#### *2.2.2.2 Stereo*

Die Spatialisierung einer Klangquelle ist auch in einem Stereosystem möglich. Diese werden akustisch durch eine Kombination aus Richtungs-, Entfernungs-, Spektral-, Phasen- und Umgebungsinformationen aufgenommen. Alternativ können sie auch elektronisch mit Hilfe von Stereo- oder Mitte-Seiten-Aufnahme in Verbindung mit spektralen, zeitlichen, phasen- und Nachhallbearbeitung erstellt werden. Diese Effekte basieren auf der binauralen Wahrnehmung des Klangs durch die hörende Person und der Illusion von virtuellen Quellen und Räumen, die durch Lautsprecher erzeugt werden. Es ist nicht erforderlich, den Zuhörer physisch mit akustischen Quellen zu umgeben. Aufgrund der geringen Anzahl an Lautsprechern müssen immersive Effekte immer auch auf psychoakustische Prinzipien beruhen, um wahrgenommen zu werden. (Roginska & Geluso, 2018, S. 86)

### 2.2.2.3 *Binaural Audio*

Der Unterscheid zwischen Stereo und Binauralen Systemen kann durch die entstehende Hörperspektive verdeutlicht werden. In Stereosystemen wird versucht die zu spezialisierende Klangquelle in den zwei Lautsprechern der hörenden Person wiederzugeben. Wohingegen bei Binauralen Systemen der Hörende über die Wiedergabe zur Klangquelle oder dem „virtuellen Raum“ gebracht wird. (Roginska & Geluso, 2018, S. 63)

In der realen als auch virtuellen Hörumgebung ist das menschliche Gehör von seinen analytischen und psychoakustischen Fähigkeiten abhängig. Sie sind dafür verantwortlich, dass der:die Hörer:in die Klangquellen verarbeiten kann. Um eine reale Umgebung virtuell zu reproduzieren, gibt es unterschiedliche technische Verfahren. Binaural Audio nutzt die physiologischen und psychoakustischen Gegebenheiten des natürlichen menschlichen Hörens um eine virtuelle Hörumgebung über Kopfhörer (aber auch zwei Lautsprecher, die Kopfhörer emulieren) zu reproduzieren. Dies resultiert in einer "First-Person"-Hörperspektive, in der die Klangquellen so repräsentiert werden, als befände sich der:die Hörer:in physisch am Aufnahmeort. (Roginska & Geluso, 2018, S. 2–3)

Bei der Binauralen Synthese wird versucht ein binaurales Signal zu erzeugen, welches eine Schallquelle an einem bestimmten Ort um die hörende Person simulieren soll. Wie in dem Kapitel HRTFs beschrieben, enthalten die HRTFs des linken und rechten Ohres dabei die entsprechenden räumlichen Hinweise, um eine Schallquelle korrekt zu lokalisieren. Werden diese HRTFs nun auf ein Tonsignal gerechnet, können zwei Signale für linkes und rechtes Ohr erzeugt werden. Die Schallquelle erscheint nun an dem vorgesehenen Ort, wird jedoch rein über ein linkes und rechtes Ohrsignal abgebildet. (Roginska & Geluso, 2018, S. 101)

### 2.2.3 **Headtracking**

In der realen Welt kommt es zu ständigen visuellen als auch akustischen Veränderungen der Umgebung. Die relative Position einer Schallquelle kann auf zwei Arten modifiziert werden: Durch die Änderung der Position der Schallquelle oder durch die Rotation und Bewegung der zuhörenden Person. Die Lokalisation einer Schallquelle wird durch diese Bewegungen und Rotationen genauer.

Durch die Bewegung des Kopfes erhält der:die Zuhörer:in viele weitere Informationen zu einer Schallquelle. Da die Lokalisationsgenauigkeit bei 0° Azimuth am höchsten ist, kann durch die Drehung des Kopfes in Richtung der Schallquelle eine bessere Lokalisierung gewährleistet werden. Außerdem spielt

## 2 Hintergrund

---

die Kopfbewegung eine dominante Rolle bei der Auflösung von Verwirrungen bei mehreren Schallquellen, die sich im "Cone of Confusion" befinden.

Ein weiteres Problem kann durch die Kopfrotation gelöst werden: Befindet sich die Schallquelle direkt vor bzw. hinter der hörenden Person ( $0^\circ$  oder  $180^\circ$  Azimuth) so sind interauralen Zeitdifferenzen und Pegeldifferenzen (ITD und ILD) in beiden Fällen gleich groß. Wird der Kopf nach rechts gedreht so verkürzt sich der Weg am beim von hinten einfallendem Schall am rechten Ohr und verlängert sich am linken. Die Quelle kann nun seitlich verortet werden. Befindet sich die Quelle bei  $0^\circ$  und der Kopf wird nach rechts bewegt so verkürzt sich der zurückgelegte Weg am linken Ohr. (Roginska & Geluso, 2018, S. 102)

In einem generischen Mixed Reality-Szenario kann eine Person frei durch die Szene navigieren und über den "Viewport" bestimmen, der basierend auf seiner Kopfposition und Rotation angezeigt werden soll. Dabei wird zwischen drei oder sechs Freiheitsgraden (3-DoF oder 6-DoF (=Degrees of Freedom)) unterscheiden. Bei 3-DoF ist die hörende Person virtuell in der Mitte einer Kugel positioniert. Die Kopfrotationen sind dabei die einzige Form der Interaktion. Rotationen des Kopfes können als „Pitch-Yaw-Roll“-Bewegungen abgebildet werden. Bei 6-DoF kann außerdem die Position im Raum (X-Y-Z Position) verändert werden. (Rossi et al., 2021, S. 1) Durch die Änderung des physischen Abstands zwischen der hörenden Person und der Schallquelle ändern sich die ITDs und ILDs entsprechend. Bei einer Positionsänderung geschieht außerdem eine Beeinflussung der Beziehung des Empfängers einer Schallquelle zum Raum. Dies führt zu einer Rekonstruktion des Direktschalls und der „Early Reflections“. Diese tragen sehr zur Lokalisierbarkeit einer Schallquelle innerhalb eines Raumes bei. (Roginska & Geluso, 2018, S. 103).

Ohne zusätzliche Verarbeitung führen synthetisierte binaurale Signale, die über Kopfhörer abgespielt werden, zu einer statischen Hörerfahrung. Dabei werden die Bewegungen des Kopfes nicht berücksichtigt und die akustische Szene bleibt starr. Dies führt zu einer weniger natürlichen Hörerfahrung und kann zur Entstehung von "In-Head-Localization" (=IHL) führen. Um eine natürlichere Hörumgebung und eine erhöhte Immersion zu schaffen, muss die Kopfrotation und Position des Hörers berücksichtigt werden. Dadurch verändert die virtuelle Schallquelle ihren Platz innerhalb der Mixed Reality nicht, wenn die hörende Person den Kopf bewegt. Um eine dynamische virtuelle Umgebung zu schaffen, muss die relative Position der virtuellen Schallquelle zum Hörer entsprechend ihrer Position verändert werden. (Roginska & Geluso, 2018, S. 104)

### 2.2.4 Immersion

Aufgrund der ganzheitlichen Erfahrung trägt besonders Audio zur immersiven (= Erfahrung eines Benutzers von Mixed Reality Applikationen bei. Akustisch entsteht Immersion dabei durch die völlige Einhüllung von Klängen aus allen Richtungen, was als unausweichliche Konsequenz des natürlichen menschlichen Hörens im Medium Luft bezeichnet wird. Auch in anechoischen Kammern ist der eigene Körper immer noch hörbar. Im Allgemeinen wird unter Immersion im Audiobereich die psychische Wahrnehmung der Einhüllung von spezifischen Soundquellen und der Umgebungsgeräusche (ambient sound) verstanden. (Roginska & Geluso, 2018, S. 5)

Als Beispiel beschreibt Roginska und Geluso die New York Central Station in welche eine natürliche immersive Audioumgebung. In ihr wirkt das Zusammenspiel von nahen gerichteten Klängen wie Gesprächen und Schritten oder weiter entfernten Klängen wie Bahnansagen mit der einhüllenden Nachhallcharakteristik der Station. Die Kombination dieser Klänge und Klangsphären definiert die immersive klangliche Wahrnehmung.

Eine virtuelle akustische Umgebung wird über Lautsprecher oder Kopfhörer wiedergegeben. Das Ziel immersiven Sounds ist dabei die realistische Abbildung einer realen Klangumgebung im virtuellen Raum. (Roginska & Geluso, 2018, S. 1–2)

### 2.2.5 Nachhall in virtuellen Räumen

In geschlossenen, reflektierenden Räumen treffen nach dem direkten Schall einer Schallquelle auch Reflexionen von Boden, Wänden und Decke auf uns. Diese Reflexionen nehmen mit der Zeit exponentiell zu, wobei die Intensität abnimmt. Die frühen Reflexionen erzeugen eine räumliche Wahrnehmung, während die späteren Reflexionen als Nachhall wahrgenommen werden. Uns erreichen Reflexionen aus allen Richtungen, was zu einer räumlichen Umhüllung führt. Unsere Wahrnehmung reduziert jedoch die Raumempfindung durch Fokussierung der Aufmerksamkeit. Dies erleichtert die technische Reproduktion und virtuelle Erzeugung von Räumlichkeit. Bereits eine einzelne Reflexion kann eine Raumempfindung vermitteln, und ein einfaches Stereosignal kann einen realistischen virtuellen Raum erzeugen. Diese Konzepte gelten auch für die simulierte räumliche Umhüllung. Eine gute Stereoaufnahme kann das Gefühl vermitteln, dass sich hinter den Lautsprechern ein Raum öffnet, mit Tiefe, Breite, Höhe und Oberflächenbeschaffenheiten des Aufnahme Raums. (Görne, 2017, S. 164)

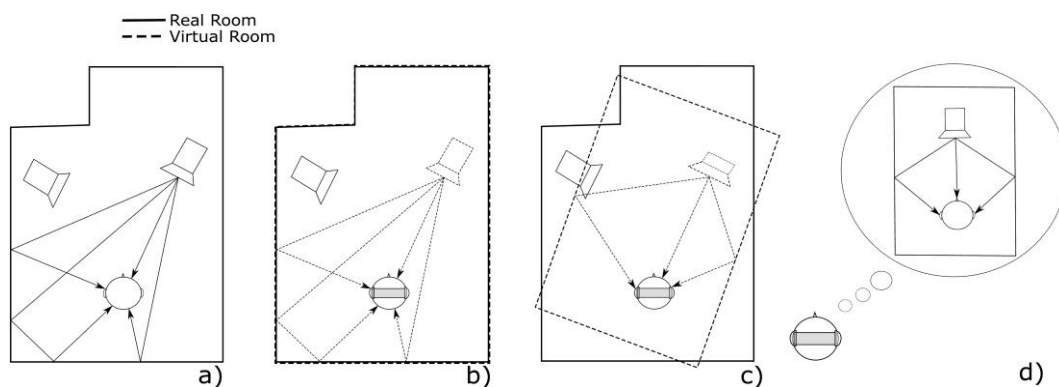
## 2 Hintergrund

---

Im Sounddesign gibt es zwei Motivationen für die Erzeugung des virtuellen Raums: die realistische Darstellung des Ortsklangs in der virtuellen Realität und die Verleihung einer metaphorischen Räumlichkeit an Klangobjekte. Die emotionale Wirkung des virtuellen Raums hängt hauptsächlich von seiner Größe ab und den damit verbundenen Assoziationen. Große Räume können den Menschen klein und ehrfürchtig fühlen lassen, während kleine Räume eine intime Atmosphäre erzeugen. Der Klang großer Räume zeichnet sich durch späte Reflexionen und langen Nachhall aus, typischerweise mehrere Sekunden. Bei geringem Nachhall können möglicherweise Echos wahrgenommen werden. (Görne, 2017, S. 166)

### 2.2.6 Wahrnehmung virtueller Klangquellen in der MR

Neidhardt et al. beschäftigen sich in ihrem 2022 veröffentlichtem Research Article mit der Frage nach der Wahrnehmung virtueller Klangquellen in einer „Auditory Augmented Reality“ (=AAR).



*Abbildung 7: Das Konzept des "perceptual matching" von Raumakustik für AAR umfasst: a) die originalgetreue Wiedergabe einer Klangquelle in einem Raum, b) einen physisch perfekt abgestimmten virtuellen Raum, welche in der Praxis jedoch schwer umzusetzen ist, c) einen physisch geringfügig unterschiedlichen, aber wahrnehmungsmäßig angepassten Raum, d) die Erwartungen des Zuhörers, die möglicherweise von dem ursprünglichen Klangfeld abweichen könnte. (Neidhardt et al., 2022, S. 3)*

In den meisten Fällen ist dem Benutzer eines Mixed Reality Devices der Raum, in dem er sich befindet, akustisch bekannt. Dies führt zu einer gewissen Erwartungshaltung wie virtuelle Klangquellen in diesem Raum wahrgenommen werden sollen. Diese müssen eine ähnliche Nachhallzeit wie reale Klangobjekte haben, dürfen keine auffällige Klangfärbung aufweisen und sollten an Positionen lokalisiert werden, die in Bezug auf ihren Inhalt sinnvoll sind. Darüber hinaus ist bekannt, dass der Nachhall die Lokalisation beeinflusst, insbesondere in Bezug auf die Entfernung. Eine angemessene Anpassung der Raumakustik ist daher

erforderlich, um eine korrekte Lokalisation zu gewährleisten. Darüber hinaus erwartet der Benutzer, dass sich das virtuelle Objekt beim Bewegen des Kopfes oder Ändern der Position im Raum wie ein echtes Objekt verhält. Im Zusammenhang mit der binauralen Synthese bezeichnen Werner et al. dies als den Raumdivergenzeffekt ("room divergence effect"). Der deutlichste Höreindruck dabei ist ein Verlust der räumlichen Ortung von virtuellen Klangquellen oder der "collapse of the externalization of auditory events". Dies gilt für alle Bemühungen zur Auralisierung unter Verwendung von binauralen Technologien. Der Effekt ist in einer AAR-Umgebung am stärksten ausgeprägt. (Neidhardt et al., 2022, S. 5)

Ein wichtiger Hinweis auf die Entfernung einer Klangquelle, ist das Verhältnis von Nachhall zu Direktschall, das als "R/D-Ratio" bezeichnet wird. Um dieses R/D-Ratio zu implementieren kann das Nachhallfeld einfach zum Direktschall zugemischt werden. Dieser Prozess allein hinterlässt jedoch keine eindeutigen Hinweise über die Intensität und die räumlichen Ausprägungen und somit kein eindeutiges Urteil über die Entfernung einer virtuellen Schallquelle. Das Verhältnis zwischen Entfernung und räumlichen Ausprägungen, basierend auf Intensität und Nachhall, wird immer gemeinsam über die Techniken der Raummodellierung und Auralisierung gebildet. (Begault, 1994, S. 138)

Begault beschreibt auch, dass die Implementierung von Nachhall innerhalb eines 3D-Soundsystems aus vielen Gründen wünschenswert ist. So gibt der Nachhall nicht nur Aufschluss über die Entfernung einer Klangquelle. Grundsätzlich kann ein 3D-Soundsystem ohne Nachhall nur virtuelle akustische Simulationen von Schallquellen in einer hallfreien Umgebung erzeugen. Die Verwendung von Nachhall geht jedoch über die Simulation hinaus. Da das natürliche räumliche Hören normalerweise mit Nachhall stattfindet, kann virtuelles räumliches Hören ohne Nachhall unrealistisch erscheinen (obwohl das gleiche Ergebnis durch grob implementierte Nachhalleffekte erzielt werden kann). In der Praxis der Tonaufzeichnung ist es zum Beispiel fast schon eine Verpflichtung, natürliche und/oder synthetischen Nachhall zu manipulieren, um Musikinstrumente echter klingen zu lassen. Beim Hören über Kopfhörer trägt Nachhall dazu bei, dass sich Schallquellen außerhalb des Kopfes befinden. Er vermittelt auch ein Gefühl für den umgebenden Kontext des Hörers und der Schallquelle und kann dem Hörer ein Bild von der Ausdehnung der Schallquelle geben. (Begault, 1994, S. 138–139)

### **2.2.7 Herausforderungen in der Mixed Reality**

Anhand einiger Beispiele beschreiben McGill et al. die „acoustic transparency“ (akustische Transparenz) als die Vereinigung von virtuellen Klangquellen mit denen der Realität. Wiedergabesysteme wie klassische Kopfhörer, Hearables und

## 2 Hintergrund

---

Earables blockieren dabei naturgemäß einfallenden Schall der Umgebung. Für den Consumerbereich wurde daher zunächst passive akustische Transparenz geschaffen. Dank entsprechender Bauweisen können teils virtuelle und echte Realität in die Gehörgänge gelangen, ebenso wurden Bone Conduction Systeme (z.B. ADHEAR<sup>23</sup>, Alpha 2<sup>24</sup>, Baha Connect<sup>25</sup>) consumer-freundlich. Aktive Noise Cancelling Systeme (z.B. Bose<sup>26</sup>, Sony<sup>27</sup>, Apple<sup>28</sup>) erlauben das Ausblenden der Realität mithilfe von Filterung kann dabei auch zum Beispiel Sprache von außen verstärken. Wie auch schon in dieser Masterarbeit beschrieben (siehe Headtracking) können mithilfe von eingebauten IMUs (=inertial measurement unit) in den Wiedergabesystemen Kopfbewegungen erfasst werden und gewisse Klangquellen wie zum Beispiel Stimmen besser externalisiert werden. Dies führt wiederum zu einer besseren Immersion bei der hörenden Person. In naher Zukunft sollen sich Benutzer:innen also in einem "personal, private (though shareable through software), acoustically transparent, highfidelity and always available audio space" befinden. (McGill et al., 2020, S. 1)

Neben den Wiedergabesystemen gibt es auch Herausforderungen in Bezug auf die Erwartungshaltung eines virtuellen Klangobjekts in einer Mixed Reality Umgebung: Wenn virtuelle Elemente bekannte reale Klangobjekte imitieren, hat eine hörende Person bestimmte Erwartungen bezüglich ihrer akustischen Eigenschaften, die auf seinen alltäglichen Hörerfahrungen beruhen. Dazu gehören beispielsweise ihre Größe, Form, Richtwirkung und Verhalten. Je nach Hörerfahrung in Bezug auf den wiedergegebenen Inhalt können diese Erwartungen, die auch als interne Referenz bezeichnet werden, sehr präzise, aber auch recht vage oder sogar falsch sein. Neidhardt et al. beschreiben beispielsweise einen Experten für Autos, welcher möglicherweise detailliertere Erwartungen an bestimmte Autotypen hat. Wenn das virtuelle Auto vor dem Zuhörer erstellt wird und in 6DOF erkundet, werden kann, werden Automechaniker sehr wahrscheinlich kritischere Zuhörer sein, da sie sich häufig um Autos umgeben und im Rahmen ihrer täglichen Aktivitäten viel auf verschiedene Geräusche achten. Im Gegensatz dazu kann eine grobe Annäherung die Erwartungen für

---

<sup>23</sup> <https://www.medel.com/hearing-solutions/bone-conduction-system>

<sup>24</sup> <https://www.medtronic.com/us-en/healthcare-professionals/products/ear-nose-throat/hearing-systems/alpha-2-mpo-eplus.html>

<sup>25</sup> <https://www1.cochlear.com/home/products-and-accessories/baha-system/baha-implants/baha-connect-system>

<sup>26</sup> [https://www.bose.at/de\\_at/products/headphones/noise\\_cancelling\\_headphones.html](https://www.bose.at/de_at/products/headphones/noise_cancelling_headphones.html)

<sup>27</sup> <https://www.sony.at/electronics/noise-cancelling-kopfhoerer>

<sup>28</sup> <https://support.apple.com/de-at/HT210643>

Personen, die sich nicht für den Klang von Autos interessieren, perfekt erfüllen. Ähnlich achten Raumakustiker und Toningenieure in ihrem Alltag mehr auf die Details der Raumakustik und bemerken eher Ungenauigkeiten in der reproduzierten Raumklangumgebung. (Neidhardt et al., 2022, S. 5–6)

Dies bestätigen auch Werner et al. und zeigen auf, dass die plausible Erzeugung virtueller Klangobjekte in einen realen Raum die Hauptherausforderung für AAR (Audio Augmented Reality) ist. Das menschliche Gehirn fungiert als ein hervorragender Mustererkennungsalgorithmus, um Klangobjekte zu erkennen und zu lokalisieren. Dabei werden die erlernten und somit erwarteten akustischen Hinweise mit den tatsächlichen Hinweisen aus der realen Umgebung verglichen. Darüber hinaus müssen die präsentierten Klangobjekte diesen Erwartungen entsprechen. Eine zu große Abweichung der akustischen Eigenschaften zwischen dem realen Raum und der virtuell reproduzierten Umgebung führt zu einer kognitiven Diskrepanz und somit zu einem Verlust der Plausibilität der gesamten auditiven Szene. (Werner et al., 2021, S. 3)

### 2.3 Audio Spatializer

Vereinfacht gesagt verarbeiten Audio Spatializer die Informationen zu einer Klangquelle bezüglich ihrer Position im virtuellen Raum und den räumlichen Eigenschaften von ebendiesem, um die Klangquelle für die hörende Person räumlich korrekt wiederzugeben. Wie schon beschrieben ist dieser räumliche Klang entscheidend für das immersive Erleben. Es liefert starke Hinweise zu der 3D Umgebung in welcher sich der:die Benutzer:in befindet. Entwickler:innen von Audio Spatializern wie die des Meta Spatializers bieten in ihren Dokumentationen<sup>29</sup> einen Einstieg in die Thematik der 3D Audio Spatialisierung. Wie auch in dieser Masterarbeit beschrieben werden die Hauptkomponenten der Spatialisierung als „Direction“ und „Distance“ identifiziert.

Die „Directional Spatialization“ wird über die HRTFs ermittelt. Die meisten „HRTF-based spatialization implementations“ (wie Audio Spatializer) nutzen dabei zumeist eine von wenigen öffentlich zur Verfügung stehenden HRTF-Datensätzen, also generalisierte HRTFs. Diese wurden entweder mit einer Reihe an menschlichen Testsubjekten oder mithilfe eines synthetischen Kopfmodells

---

<sup>29</sup> <https://developer.oculus.com/resources/audio-intro-spatialization/>

## 2 Hintergrund

---

erzeugt. Beispiele dieser Datenbanken sind die IRCAM HRTF Database<sup>30</sup> oder die KEMAR Database<sup>31</sup>.

Je nach Einfallsrichtung einer Klangquelle werden Filter auf das Schallereignis berechnet. Diese Berechnung erfolgt in den meisten Fällen durch eine Faltung in der Zeit- oder Frequenzdomäne. Siehe auch 2.2.1 Auralisation.

Meta beschreibt außerdem ein Distanz Model, welches verwendet wird, um der hörenden Person eine Einschätzung zur Entfernung der Klangquelle geben zu können. Dabei spielen folgende Punkte eine wichtige Rolle: Lautstärke, Initial Time Delay, Direktschall/Nachhall, Motion Parallax, Air Absorption. (*3D Audio Spatialization* | Oculus Developers, o. J.)

### 2.3.1 Game Engine

Da Audio Spatializer Audiosignale verarbeiten und an ein Ausgabesystem weitergeben, sind sie Teil der Auralisationskette (siehe 2.2.1 Auralisation). Toftedahl und Engström beschreiben die Game Engine folgendermaßen: „From a software architecture perspective, a game engine is a complex system of intertwined layers relating to hardware and other software.“ Die Komplexität einer Game Engine lässt sich auch mit dem Zusammenspiel und der großen Anzahl an Komponenten beschreiben. Neben Komponenten wie „graphic rendering functions, collision detection systems, physics simulations and animation systems“ zählt auch die Audio Engine, also auch die beschriebene Auralisationskette, dazu. (Toftedahl & Engström, 2019, S. 8)

Die Applikation für den Hörtest dieser Masterarbeit wird in Unity 2023.1.20.f1 programmiert. Unity ist eine Game Engine, die von Unity Technologies entwickelt wurde. Sie ermöglicht die Erstellung von 2D- und 3D-Spielen für verschiedene Plattformen wie PC, Konsolen, mobile Geräte und VR-Headsets. Die Engine bietet einen visuellen Editor, unterstützt verschiedene Programmiersprachen wie C# und JavaScript und enthält umfangreiche Grafikfunktionen wie Beleuchtung und Schattierung. Unity ermöglicht die plattformübergreifende Entwicklung und verfügt über eine aktive Entwicklergemeinschaft sowie einen Asset Store, der vorgefertigte Assets und Erweiterungen bereitstellt. Es wird sowohl von Indie-

---

<sup>30</sup> <http://recherche.ircam.fr/equipes/salles/listen/>

<sup>31</sup> <https://sound.media.mit.edu/resources/KEMAR.html>

## 2 Hintergrund

---

Entwicklern als auch von großen Studios genutzt. (*Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine*, o. J.)

Eine weitere namhafte Game Engine ist die Unreal Engine. Die Unreal Engine ist eine Game Engine, entwickelt von Epic Games. Sie ermöglicht die Erstellung von 2D- und 3D-Spielen für verschiedene Plattformen wie PC, Konsolen, mobile Geräte und Virtual Reality. Die Engine bietet fortschrittliche Grafikfunktionen, Animationstools und Unterstützung für verschiedene Programmiersprachen wie C++. Sie ermöglicht die Implementierung von Audio, künstlicher Intelligenz und Multiplayer-Funktionalität. Die Unreal Engine ist für ihre visuellen Effekte bekannt und wird sowohl von Indie-Entwicklern als auch von großen Studios verwendet. (*Unreal Engine | Das stärkste Werkzeug für 3D-Echtzeit-Entwicklung*, o. J.)

Weitere Game Engines existieren, werden allerdings in weit weniger veröffentlichte Projekten verwendet als die Unreal Engine oder Unity. Große „Triple A“ Game Studios entwickeln meist ihre eigenen Game Engines, Indie Game Entwickler und Projekte im wissenschaftlichen Kontext greifen zumeist auf die frei zur Verfügung stehenden Game Engines zurück. Beispiele für Game Engines sind Source, Cryengine, Gamebryo, IW, Anvil, id Tech, Essence oder Clausewitz. (Toftedahl & Engström, 2019, S. 6–8)

### 2.3.2 Audio Middleware

Der Sound Design Prozess für ein Videospiel, welches eine Game Engine der oben genannten Game Engines nutzt, lässt sich von Mizutani in sechs Schritten zusammenfassen (Mizutani, 2017, S. 1):

1. Sound Designer testen die vom Programmierer integrierten Sounds und erkennen eine notwendige Anpassung.
2. Komponisten oder Sound Designer manipulieren, erweitern oder erstellen neue Sounds und erzeugen ein entsprechendes Audiosample als Datei.
3. Sound Designer erklären den Programmierer:innen wo, wann und wie dieses Sample abzuspielen ist.
4. Die Programmierer:innen implementieren das Sound Sample im Spielecode.
5. Eine neue Version des Spiels steht zur Verfügung.
6. Bei 1. fortfahren.

Der „feedback circle“ bei diesem klassischen Sound Design Prozess ist sehr lange, wobei mindestens zwei voneinander unabhängige Teams benötigt werden. Sound Designer sind von der vollständigen Implementierung der Sounds und einer

## 2 Hintergrund

---

angelieferten Testversion abhängig. Weiters haben Programmierer:innen zumeist keine bis wenig Erfahrung im Feld des Sound Designs und Sound Designer:innen oftmals unzureichend Erfahrung im Bereich der Programmierung. Interdisziplinäre Unterschiede können die Kommunikation behindern und weitere Feedbackschleifen hervorrufen.

Diese Probleme sind nicht ausschließlich auf den Sound Design Prozess beschränkt. Sie betreffen auch Graphic Designer oder 3D-Modellierer:innen. Um dem entgegenzuwirken werden zumeist so genannte Middlewares verwendet. Diese unterstützen die Sound Designer bei der Integration und Manipulation von Sounds, mithilfe einer eigenen GUI (=Graphical user interface). In dieser können alle Sound Samples verwaltet werden.

Hier sollen kurz zwei wichtige Game Audio Middlewares vorgestellt werden. FMOD und Wwise werden als „commercial off-the-shelf solutions (=COTS)“ bezeichnet und können für Unity und Unreal Engine verwendet werden. (Mizutani, 2017, S. 1)

### 2.3.2.1 Wwise

Wwise ist eine professionelle Audio-Entwicklungsumgebung und Middleware, die von Audiokinetic entwickelt wurde. Sie wird in der Videospiele- und Unterhaltungsindustrie eingesetzt, um hochwertigen und interaktiven Sound für Spiele, Filme und andere interaktive Medien zu ermöglichen.

Wwise bietet eine umfangreiche Palette an Tools und Funktionen für das Audiodesign und die Integration von Sound in Projekte. Es ermöglicht die Erstellung und Bearbeitung von Soundeffekten, Musikkompositionen und Dialogen. Die Middleware ermöglicht es Entwicklern, komplexe Soundlandschaften zu erstellen, indem sie Echtzeit-Audio-Rendering, 3D-Audio-Positionierung und Effekte nutzen.

Die Software unterstützt eine Vielzahl von Plattformen und kann in verschiedene Entwicklungsumgebungen integriert werden. Sie ermöglicht die nahtlose Zusammenarbeit zwischen Sounddesignern und Entwicklern, indem sie einen effizienten Workflow und Tools zur Zusammenarbeit bietet.

Wwise bietet außerdem fortschrittliche Funktionen wie die Integration von Physiksimulationen für Soundeffekte und die Unterstützung von VR- und AR-Technologien. (*Wwise | Audiokinetic*, o. J.)

### 2.3.2.2 FMOD

FMOD ist eine professionelle Middleware für interaktive Audioentwicklung, die von Firelight Technologies entwickelt wurde. Sie wird in der Videospiele- und Unterhaltungsindustrie eingesetzt, um hochwertigen und immersiven Sound in Spielen, Filmen und anderen interaktiven Medien zu ermöglichen.

FMOD bietet eine breite Palette von Tools und Funktionen für das Audiodesign und die Integration von Sound in Projekte. Es ermöglicht die Erstellung und Bearbeitung von Soundeffekten, Musikkompositionen und Dialogen. Die Middleware unterstützt Echtzeit-Audio-Rendering, 3D-Audio-Positionierung, Effekte und umfangreiche Mixing-Möglichkeiten.

Die Software ist plattformübergreifend und kann in verschiedenen Entwicklungsumgebungen eingesetzt werden. Sie bietet flexible Integrationsoptionen und ermöglicht eine nahtlose Zusammenarbeit zwischen Sounddesignern und Entwicklern.

FMOD unterstützt fortschrittliche Funktionen wie die Integration von Physiksimulationen für Soundeffekte, die Unterstützung von VR- und AR-Technologien sowie die Integration von Live-Performance-Tools. (FMOD, o. J.)

### 2.3.3 Meta Spatializer

Der Meta Spatializer ist Teil der Meta XR Audio SDK (=Software Development Kit) welche im Zuge der Presence Platform am 7. Februar 2023 veröffentlicht wurde. Dieser gilt als Weiterentwicklung des Oculus Spatializers. Die Audio SDK wurde speziell für Entwickler:innen geschaffen, die auch ohne Vorkenntnisse im Sound Design, schnell ein akustisch immersives XR-Erlebnis schaffen wollen. Dabei bietet die Audio SDK Unterstützung für "almost any standalone mobile VR device, as well as PCVR (SteamVR, etc.) and devices from other manufacturers". (Gildersleeve, 2023)

Der Meta Spatializer ist somit Teil der Meta XR Audio SDK. Meta beschreibt in der Dokumentation zum Spatializer folgende Features:

#### **Position**

Die virtuelle Klangquellen werden an ihre vorgesehenen Positionen platziert. Diese müssen monophon sein, damit der Spatializer sie binauralisiert wiedergeben kann. HRTFs helfen bei der Spatialisierung mithilfe der „C/C++ SDK and plugins“.

### **Nahfeld Rendering**

Das Nahfeld Rendering wird für Klangquellen die näher als 1m zum Kopf der hörenden Person liegen verwendet. Dabei werden die Effekte der akustischen Beugung miteinbezogen. Somit soll eine realistischere Repräsentation des Klangs gewährleistet werden.

### **Raum Akustik**

Räumliche Hinweise wie die Erstreflexionen („early reflections“) oder der Nachhall („late reverberation“) werden berechnet, um die Distanz einer Klangquelle einschätzen zu können.

### **Ambisonics**

Ambisonics ist ein Mehrkanal-Audioformat, das ein dreidimensionales Klangfeld repräsentiert. Man kann es sich als eine Art Himmelskuppel für Audio vorstellen, mit der hörenden Person im Zentrum. Es ist eine rechenintensive Methode, um eine vorab gerenderte oder live aufgenommene Szene wiederzugeben. Der Kompromiss besteht darin, dass ambisonische Klänge weniger räumliche Klarheit bieten und mehr Verwischungen aufweisen als punktuellenbasierte HRTF-verarbeitete Klangquellen. Meta empfiehlt die Verwendung von Ambisonics für nicht-diegetische Klänge wie Musik und Atmosphäre. Meta hat eine verbesserte Methode zur binauralen Wiedergabe von Ambisonics entwickelt, die auf der Dekodierung von sphärischen Harmoniken basiert. Im Vergleich zu gängigen Wiedergabemethoden mit virtuellen Lautsprechern bietet dieser Ansatz eine flachere Frequenzantwort, bessere Ortung, mehr Klarheit des Klangs und einen geringeren Ressourcenverbrauch. Das Meta XR Audio SDK unterstützt Ambisonics erster Ordnung im AmbiX (ACN/SN3D)-Format.

### **Dämpfung und Reflexionen**

Wichtige Merkmale bei der Simulation von Entfernung sind Reflexionen. Um korrekte Entfernungsanweisungen zu liefern, ist es entscheidend, ein natürliches Gleichgewicht zwischen dem direkten Signalweg und den Reflexionen herzustellen. Dabei muss die Dämpfung aller Reflexionswege berücksichtigt werden, die vom Raummodell reflektiert werden.

Für eine Klangquelle, die sich nahe beim Hörer befindet, sind die Reflexionen kaum hörbar, und der Klang wird vom direkten Signal dominiert. Ein weiter entfernter Klang enthält mehr Nachhall und die Reflexionen sind typischerweise fast genauso laut wie das direkte Signal. Dies wird alles vom Spatializer verwaltet. Es kann jedoch eine Distanz-Dämpfungskurve ("Attenuation curve") pro

Klangquelle festgelegt werden. Der Meta Spatializer wird jedoch das korrekte Verhältnis zwischen direktem und nachhallendem Signal beibehalten.

### **Volumetric Sounds**

Volumetric Sounds ist ein Feature, welches es erlaubt Klangquellen eine gewisse Größe zuzuordnen. Standardmäßig ist eine virtuelle Klangquelle unendlich klein und strahlt in alle Richtungen gleichmäßig ab. Um einen größeren Klang darzustellen, wurde von Oculus Research ein Prozess entwickelt, welches Projektionen einer Klangquelle anhand der Entfernung und dem Radius berechnet. Dabei werden räumliche Filter während der Laufzeit konstruiert. Ist der Klang sehr weit entfernt, ist die Projektion klein und hört sich wie eine Punktschallquelle an. Befindet er sich näher bei der hörenden Person, wirkt sie größer und der Klang breitet sich weiter aus. Somit können physikalisch korrekte Klangquellen mit guter Performance repräsentiert werden. Befindet sich die hörende Person innerhalb des Radius einer Klangquelle, wird sie vom Klang eingehüllt. „It smoothly blends from spatial to surround as the listener approaches the center of the sound.“ (Stirling, o. J.)

### **Abstrahlverhalten**

Der Meta Spatializer bietet die Möglichkeit Klangquellen nicht nur omnidirektional abstrahlen zu lassen. So können zum Beispiel Stimmen in eine bestimmte Richtung abstrahlen. Diese klingen leiser und gedämpfter, wenn sich die sprechende Person der hörenden abwendet.

### **HRTF Intensität**

Mit den „HRTF intensity parameter controls“ kann die HRTF-Färbung auf eine virtuelle Klangquelle bestimmt werden. „This has the effect of reducing the change to the timbre or “coloration” caused by HRTF“, wird jedoch als reduzierte Externalisation wahrgenommen. Zusätzlich wird die Lokalisierungsgenauigkeit gemindert. Es wird empfohlen, diese Einstellungen nur zu kreativen Zwecken zu manipulieren.

### **Binauralisierung**

Der Meta Spatializer binauralisiert die virtuellen Klangquellen, was bedeutet, dass die Ausgabe über ein Stereosystem bereitgestellt wird. Daher ist die Ausgabe über Kopfhörer oder die eingebauten Lautsprecher des HMDs erforderlich. Diese werden von Meta für ein gutes „spatial image“ empfohlen. Zusätzlich bietet Meta einen Guide für „listening devices“ in welchem behauptet wird, dass die Verwendung von Kopfhörern in Zukunft der Standard für Audio in VR-Geräten sein

## 2 Hintergrund

---

wird. Dies wird damit begründet, dass die Ausgabe „better isolation, privacy, portability, and spatialization“ bietet. „For modern VR, especially with head tracking and user movement, speaker arrays do not provide an adequate audio experience.“ (*Listening Devices | Oculus Developers*, o. J.)

### **Performance**

Für jede Echtzeit Anwendung muss eine gute Performance gewährleistet werden. Der Meta Spatializer wurde entsprechend optimiert. Gleichzeitig verweist Meta auf einen gewissen Overhead bei der räumlichen Darstellung von Klängen im Vergleich zu traditionellen 3d-Panning-Methoden. Eine Beeinflussung der Bildrate (FPS = Frames per seconds) ist nicht zu befürchten, da das Audiosystem auf einem separaten Thread verarbeitet wird.

Weiters wird darauf hingewiesen, alle virtuellen Klangquellen über das Meta XR Audio SDK zu verarbeiten, da es ansonsten zu Lokalisierungsproblemen (Front/Back Ambiguity, Elevation Ambiguity, Hard Panning) kommen kann. (*Meta XR Audio SDK Features: Unity | Oculus Developers*, o. J.)

### **2.3.4 DearVR Spatializer**

Der von Dear Reality entwickelte Audio Spatializer ist Teil einiger Produkte zur Produktion von „immersive Audio“. Die Kernfunktionalität des Spatializers ist dabei in Form des „DearVR Core“ in allen Produkten („Plugins“) verbaut. DearVR PRO<sup>32</sup> wird für die Erstellung von immersiven Soundproduktionen innerhalb einer DAW (=Digital Audio Workstation) verwendet. Es existiert auch ein eigenes Plugin für die Produktion von immersiver Musik, dem DearVR MUSIC.<sup>33</sup>

Eine direkte Integration in Unity ist mit dem DearVR Unity Plugin möglich. Es wird als Unity Asset zur Verfügung gestellt. DearVR Unity beinhaltet drei „processing units“, welche für die Spatialisierung der virtuellen Klangquellen notwendig sind: Position, Reflektion und Nachhall.

Die Positionsverarbeitungseinheit verarbeitet Distanz, Rotation und Höhe einer Klangquelle relativ zur Hörposition. Sie verwandelt ein Mono Klangereignis in ein 3D Sound Objekt.

---

<sup>32</sup> <https://www.dear-reality.com/products/dearvr-pro>

<sup>33</sup> <https://www.dear-reality.com/products/dearvr-music>

## 2 Hintergrund

---

Abhängig von der Position einer hörenden Person in einem Raum und seinem Abstand zu den Grenzen variiert die Zeit, die eine Reflexion benötigt, um an ihre Ohren zu gelangen. Reflexionen variieren auch je nach Position der Schallquelle im Raum. Erstreflexionen werden dabei so modelliert, als würden Wände und Gegenstände im Raum sie wie ein Spiegel reflektieren. Die dearVR Auralisierung erzeugt Reflexionsmuster, abhängig von der Position der hörenden Person und der Position der Klangquelle. Wenn sich die hörende Person oder ein Klangobjekt bewegt, berechnet die Engine die Eigenschaften der Reflexionen in Echtzeit neu.

Aufgrund der hohen Komplexität einer Echtzeit Nachhallsimulation bietet DearVR Unity ein Reverb Plugin. Dieses Plugin beinhaltet 45 Presets, welche aus unterschiedlichen Räumen und Lokalitäten erstellt wurden. Diese Presets können in ihrer Größe und durch einen Filter manipuliert werden. (Dear Reality, 2020)

Weitere Features des Spatializers werden in der offiziellen Dokumentation erläutert und beinhalten unter anderen:

### **Absorption**

Um die Absorption von Schallwellen durch Wände oder geschlossene Türen und Fenster zu simulieren, können absorbierende Objekte einem Unity Layer zugewiesen werden.

### **Beugung**

Bei der Beugung treffen die Schallwellen einer Klangquelle nicht direkt auf die hörende Person ein. Sie werden durch Gegenstände im Raum (zum Beispiel eine Säule) blockiert. Dabei wird hauptsächlich das Direktsignal beeinflusst.

### **Performance Mode**

Mit Unity werden "spatialized" Audio Sources verarbeitet, egal ob sie gerade einen Klang von sich geben oder nicht. Der DearVR Performance Mode erlaubt es diese Verarbeitung zu umgehen und somit die Rechenlast zu vermindern.

### **Automatische Raumanalyse**

Um die Auralisation des in Unity vorliegenden 3D Raums vorzunehmen, bietet dieses Feature eine automatische Raumanalyse. Dabei werden die Dimensionen des Raums automatisch erfasst. Sie können auch manuell eingetragen werden.

### 2.3.5 Steam Audio

Steam Audio<sup>34</sup> ist ein weiterer Audio Spatializer der Spieleplattform Steam. Die verschiedenen Features zur Darstellung von Klangquellen in einem dreidimensionalen Raum werden in der Dokumentation wie folgt dargestellt:

#### **Positionierung**

Steam Audio modelliert die horizontale und vertikale Richtung einer virtuellen Klangquelle relativ zur hörenden Person.

#### **Ambisonic**

Steam Audio ermöglicht die räumliche Darstellung von Ambisonic-Audio-Clips, indem sie basierend auf der Ausrichtung der hörenden Person in der virtuellen Welt gedreht werden. Dies ist besonders nützlich, um feste räumliche Klangumgebungen zu erzeugen.

#### **Benutzerdefinierten HRTFs mit SOFA-Dateien**

Neben den integrierten HRTFs kann Steam Audio punktbasierte und Ambisonic-Quellen mit beliebigen HRTFs räumlich darstellen. Diese Dateien müssen im SOFA-Format vorliegen.

#### **Absorption**

Steam Audio kann die direkte Verdeckung des Direktschalls durch feste Objekte modellieren. Auch teilweise Absorption wird durch entsprechende Filterung des Klangs simuliert.

#### **Erstreflexionen**

Steam Audio erzeugt Erstreflexionen durch die Reflexion des Klangs an festen Objekten. Dadurch können verschiedene Arten von Umgebungseffekten wie Nachhall, Flatterechos, Schallverdeckung durch Gebäude erzeugt werden.

#### **Beugung**

Steam Audio kann modellieren, wie der Schall von der Quelle zur hörenden Person entlang mehrerer Pfade propagiert wird. Dies ermöglicht eine korrekte räumliche Darstellung von Schall, der den:der Zuhörer:in nach dem Durchlaufen von Fluren, um Ecken, durch Türen und Fenster usw. erreicht.

---

<sup>34</sup> <https://valvesoftware.github.io/steam-audio/>

### **Automatische Raumanalyse**

Steam Audio analysiert Größe, Form, Layout und Materialeigenschaften von Räumen und Objekten in der Szene. Auf Grundlage dieser Informationen berechnet es automatisch Absorption, Reflexionen und andere Umgebungseffekte.

### **Nachhall und Faltungshall**

Mit Steam Audio müssen Effektfiler nicht manuell im gesamten Szenario platziert und angepasst werden. Dabei kann eine Echtzeit Berechnung des Halls oder eine vorberechnete Variante des Halls verwendet werden, bei dem die akustischen Eigenschaften der Umgebung basierend auf physikalischen Prinzipien in der gesamten Szene berechnet werden. Weiters besteht die Möglichkeit der Faltungshallberechnung. Dieser wird besonders für Außenbereiche empfohlen, da der parametrische Hall in diesen Situationen einige Einschränkung hat.

### **GPU-Nutzung**

Zur Berechnung der Echtzeit-Umgebungseffekte nutzt Steam Audio neben der CPU (=central processing unit) auch die GPU (=graphics processing unit) der Graphikkarte. Dabei wird darauf geachtet, dass die "frame rate stability" priorisiert wird. (*Steam Audio Unity Integration — Steam Audio Unity Integration documentation*, o. J.)

### **2.3.6 Project Acoustics**

Mit dem Project Acoustics<sup>35</sup> Spatializer verfolgt Microsoft einen Ansatz der „wave-based simulation“. Im Gegensatz zu der häufig eingesetzten „ray-based simulation“ wird die Akustik eines Raums so nahtloser, genauer und vorhersehbarer. Die Simulation des Nachhalls, Erstreflexionen oder Absorption werden dem:der Sound Designer:in dennoch als veränderbarer Parameter angeboten. Die Erkenntnisse zu diesem Konzept wurden aus dem „Project Triton“ integriert. Features umfassen die Beugung, Absorption, das „Portaling“ und den Nachhall. („Project Triton - Immersive Sound Propagation“, o. J.)

„Portaling“ ist die Umleitung einer virtuellen Klangquelle in einem Raum durch eine Öffnung, dem Portal.

---

<sup>35</sup> <https://learn.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>

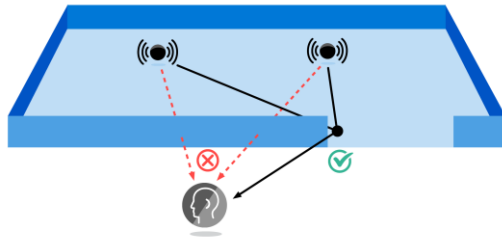


Abbildung 8: Schallwellen "durch die Wand" welche aus der Quellrichtung gehört werden sind inkorrekt. Offene Portale (Türen, Fenster,...) leiten den klang realistisch um. („Project Triton - Immersive Sound Propagation“, o. J.)

Da die „wave-based simulation“ eine äußerst intensive Berechnung anhand der Raumgeometrie und Materiale durchführt, können diese Simulationen nicht in Echtzeit stattfinden. Stattdessen wird die zuvor berechnete Simulation in die Szene gespeichert („bake“). (kevinasg, 2023)

### 2.3.7 Spatializer in Unity

Es zeigt sich, dass jeder Spatializer viele Einstellmöglichkeiten zur akustischen Repräsentation eines Raumes und seiner virtuellen Klangquellen bieten. So werden APIs für Unity angeboten (DearVR, Steam Audio, Meta) um die Parameter einer Audio Source in Unity zu erweitern oder überschreiben. Manche Parameter beziehen sich auf den Raum selbst, wie der Nachhall oder die automatische Raumanalyse (DearVR Reverb Plugin oder Project Acoustics eingebaute Simulation) oder auch auf die Hörposition und der hörenden Person selbst wie beispielsweise die Verwendung individualisierbarer HRTFs (Steam Audio).

Jeder in dieser Masterarbeit vorgestellte Spatializer bietet eine Möglichkeit zur Integration in Unity. Dabei muss der Spatializer integriert und in den Projekteinstellungen ausgewählt werden. Alle Anbieter liefern entsprechende Dokumentation. Manche Spatializer können auch in einer Audio-Middleware wie Wwise oder FMOD integriert werden. Dadurch ändert sich auch der Workflow und teilweise auch die Möglichkeiten in Bezug auf das Sound Design. Die Verwendung von Wwise erlaubt beispielsweise die Einstellung des Abstrahlverhalten. (*Simulating Directivity Using Cone-Shaped Boundaries*, o. J.) Als reines Unity Spatializer Plugin bietet von den hier vorgestellten Anbietern jedoch nur der Meta Spatializer dieses Feature.

Um den Hörtest fair gestalten zu können, werden ausschließlich die direkten Integrationen eines Spatializers in Unity zur klanglichen Gestaltung des Raumes und der virtuellen Klangquellen genutzt.

## 3 Testen der Spatializer

### 3.1 Vorstudie

Im Rahmen einer vom Autor der Arbeit durchgeführten Vorstudie im Frühjahr 2023 konnten bereits erste Erfahrungen für diesen Hörtest gesammelt werden. In Kooperation mit Dear Reality wurde mit dem DearVR Spatializer Erkenntnisse zur Lokalisierung von virtuellen Klangquellen in der Mixed Reality für diesen Spatializer geliefert.

Die Studie untersuchte dabei einerseits die Unterschiede bei der Lokalisierung zwischen den extra auralen Lautsprechern der Meta Quest 2 und verwendeten Studiokopfhörern. Wichtiger für die Masterarbeit ist die Fragestellung nach der Lokalisierung mit dem DearVR Spatializer. Die Ergebnisse sollen nachfolgend dargestellt werden.

Die erstellte Mixed Reality Applikation und der durchgeführte Hörtest dient als Vorbild für diese Masterarbeit. Testdesign, Implementierung, Ablauf und Auswertung konnten aus den Ergebnissen der Vorstudie profitieren.

#### 3.1.1 Testdesign

Im Zuge der Auswertung wurden die Nullhypothesen mithilfe zweier t-Tests auf ein signifikantes Ergebnis getestet. Dafür sind drei Populationen aus dem Hörtest zu gewinnen. Jedes Datenset (Population) bestand dabei aus 36 Samples. Diese beschreiben eine Position im zweidimensionalen Raum als X und Y Koordinate. Da die Elevation der Soundquelle nicht getestet wurde, ist ein Vergleich der Z-Achse hinfällig.

Unabhängige Variablen waren die im virtuellen Raum platzierten Soundquellen. Die Position wurde dabei in Distanz und Winkeldistanz zur Testperson angegeben. Abhängige Variablen waren einerseits die von dem:der Hörenden platzierten Marker zu einem korrespondierenden Sample in Distanz und Winkeldistanz, sowie ein weiterer Datensatz, der beim erneuten Testdurchlauf mit offenen Studiokopfhörern entstand. Insgesamt wurden 8 Personen allen Konditionen (mit und ohne Kopfhörer) ausgesetzt, mit jeweils 36 Samples. Dies entsprach einem "full factorial repeated within subject design". Es handelte sich um 72 (36x2) Stimuli.

### 3 Testen der Spatializer

---

Verglichen wurden also die getesteten Positionen der Samples mit den vom Hörenden vermuteten Positionen mit und ohne Kopfhörer auf einer zweidimensionalen Ebene angegeben als Distanz und Winkeldistanz zur Testperson. Eine weitere Kontrollvariable bzw. abhängige Variable, welche jedoch nicht für die zu testende Hypothese relevant war, stellte die Geschwindigkeit dar, mit welcher die Testperson ein Sample bewertete.

Ein Testdurchlauf benötigt etwa 5-10 Minuten, was pro Sample einer Dauer von etwa 8-16 Sekunden entsprach. Insgesamt durchlief eine Testperson den Test zweimal (mit und ohne Kopfhörer), was einer Dauer von 10-20 Minuten entsprach.

Der Test fand an der Fachhochschule St. Pölten am 23.01.2023 statt. Aufgrund einer Kooperation mit dem Health Department war es möglich, ein bereits vorhandenes 3D Modell des Motion Capturing Labors für die Entwicklung zu verwenden. Dieses Labor wurde ausgesucht, da es eine mittlere Größe (Tiefe: ~17m, Breite: ~8m, Höhe: ~3m) besitzt und recht weitläufig ist. Es gibt wenige Gegenstände, durch welche es zu einer Okklusion oder Obstruktion kommen konnte. Das Motion Capturing Lab eignete sich daher für die Realisierung einer glaubhaften Wiedergabe der Soundquellen in der Mixed Reality.

#### 3.1.2 Durchführung

Die 8 Testpersonen durchliefen dabei den Test hintereinander. Pro Person wurden etwa 8 Minuten pro Testlauf benötigt, jede Person durchlief jeweils den Test mit Kopfhörer (AKG K240, 55 Ohm) direkt am VR-Headset angeschlossen und ohne Kopfhörer (extraaural). Das verwendete VR-Headset ist die Meta Quest 2. Von 11:30 bis 15:30 wurden so insgesamt 16 Datensets gesammelt. Während des Tests sind lediglich der Versuchsleiter und die Testperson im Raum. Dabei wurde der Lärmpegel so gering wie möglich gehalten. Gespräche über Unklarheiten während des Tests waren zulässig, da sie das Testergebnis nur positiv beeinflussen konnten.

Beim Testen wurde vermehrt angemerkt, dass die Kopfhörer Ausgabe trotz maximaler Lautstärke mit den AKG K240 zu leise sei.

Bei einigen Usern wurde das 3D Modell nach der Erstsynchronisierung des virtuellen Raums mit dem realen Raum um einige Zentimeter versetzt. Dies geschieht beim Umsetzen der Meta Quest 2 auf die Testperson. Das System kann das 3D-Modell nicht immer korrekt platzieren. Die Problematik äußert sich durch den versetzten Ring, in welchem die Testperson zu stehen hat. Erneutes

Synchronisieren ist dann durch den Versuchsleiter von Nöten. Eine Person ohne jegliche Vorkenntnis mit VR war mit den Meta Quest Controllern überfordert.

#### 3.1.3 Evaluation

##### 3.1.3.1 Externe offene Kopfhörer vs. Extraaurale Lautsprecher

Der erste t-Test prüfte, ob ein signifikanter Unterschied zwischen den Distanzdifferenzen und Winkeldifferenzen innerhalb den Gruppen mit und ohne Kopfhörer vorhanden ist.

H0: Es gibt keine Unterschiede beim Lokalisieren von Soundquellen mit externen offenen Studiokopfhörern zur Lokalisierung von Soundquellen mit extraauralen Lautsprechern.

Die Varianzhomogenität mittels eines Levene-Tests hat ergeben, dass bei den Distanzdifferenzen ( $p = .013$ ) als auch den Winkeldifferenzen ( $p = .016$ ) keine Gleichheit der Varianzen angenommen werden kann. Daher werden die Ergebnisse des unabhängigen t-Test nicht berücksichtigt. Der Welsh-Test jedoch zeigt, dass es einen signifikanten Unterschied der durchschnittlichen Distanzdifferenzen zwischen den Ergebnissen ohne Kopfhörer ( $M = 1.145$ ,  $SD = .526$ ) und jenen mit Kopfhörer ( $M = 1.559$ ,  $SD = .845$ ) gibt,  $t(58.571) = -2.496$ ,  $p = .015$ . Die Unterschiede haben einen mittelgroßen Effekt,  $d = -.59$ . Weiters zeigt ein Welsh-Test einen nicht signifikanten Unterschied der durchschnittlichen Winkeldifferenzen zwischen den Ergebnissen ohne Kopfhörer ( $M = 14.44$ ,  $SD = 13.49$ ) und jenen mit Kopfhörer ( $M = 16.06$ ,  $SD = 7.52$ ),  $t(54.83) = -0.626$ ,  $p = .534$ .

Die Nullhypothese ist somit in Teilen widerlegt. Virtuelle Soundquellen können ohne Kopfhörer in der Distanz genauer verortet werden als mit Kopfhörer.

##### 3.1.3.2 Lokalisierung

Der zweite t-Test prüfte, ob ein signifikanter Unterschied zwischen den Distanzdifferenzen und Winkeldifferenzen innerhalb der Gruppen 0 + 1 (ohne + mit Kopfhörer) und einer Kontrollgruppe 2 (Differenz = 0) vorhanden ist. Es wird angenommen, dass die Kontrollgruppe alle Soundquellen perfekt verorten konnte.

H0: Die Soundquellen werden punktgenau verortet.

Die Varianzhomogenität mittels eines Levene-Tests hat ergeben, dass in beiden Fällen (Distanzdifferenz und Winkeldifferenz) keine Gleichheit der Varianzen vorhanden ist. ( $p = <.001$ ) Ein Welsh-Test zeigt in beiden Variablen einen signifikanten Unterschied zwischen der Gruppe ohne + mit Kopfhörer für die

### 3 Testen der Spatializer

---

Distanzdifferenzen ( $M = 1.352$ ,  $SD = 0.729$ ) und Winkeldifferenzen ( $M = 15.25$ ,  $SD = 10.875$ ) und dem Ursprung ( $M = 0.$ ,  $SD = 0.$ ). Welsh-Test der Distanzdifferenzen,  $t(71) = 15.733$ ,  $p = <.001$ . Welsh-Test der Winkeldifferenzen,  $t(71) = 11.899$ ,  $p = <.001$ . Die Effektgröße ist in beiden Fällen weit über 1 und somit sehr groß. Distanzdifferenz  $d = 2.622$ , Winkeldifferenz  $d = 1.983$ .

Die Nullhypothese ist somit widerlegt. Virtuelle Soundquellen wurden nicht punktgenau verortet.

#### 3.1.4 Ergebnisse und Anpassungen

Die Ergebnisse der t-Tests zeigten nun also, dass die Distanzeinschätzung einer Soundquelle mit dem DearVR Spatializer und den im Experiment gesetzten Parametern ohne Kopfhörer präziser ist. Im Durchschnitt kann von etwa 40cm Präzisionsgewinnung bei der Wiedergabe ohne Kopfhörer ausgegangen werden, wenn die eingebauten Lautsprecher der Meta Quest 2 mit den AKG K240 Kopfhörer verglichen werden. Diese Erkenntnis wird auch durch Eindrücke der Testpersonen eindeutiger. Viele merkten an, dass die Kopfhörer zu leise sind und es daher schwerer war die Schallquellen zu verorten.

Bei den Winkeldifferenzen konnte keine eindeutige Aussage getroffen werden da der Welsh-Test nicht signifikante Unterschiede erkennen lässt, jedoch zeigen die erhobenen Werte interessanterweise eine um 2 Grad bessere Ortung im Azimut ohne Kopfhörer. Die Werte sind jedoch bei einer Standardabweichung von 7-8 Grad mit Kopfhörer weniger gestreut (14-15 Grad). Die Hypothese „Die Ortbarkeit der Soundquelle mit externen offenen Studiokopfhörern ist präziser als mit den integrierten (extraauralen) Lautsprechern der Meta Quest Pro.“ muss somit in Teilen verworfen werden. Das Ergebnis für die Winkeldifferenzen würde diesen Teil der Hypothese bestätigen, der Test ist jedoch nicht signifikant. Das Ergebnis für die Distanzdifferenzen ist signifikant und verwirft die Hypothese ins Gegenteil: „Die Distanz einer Soundquelle mit externen offenen Studiokopfhörern lässt sich weniger präzise verorten als mit den integrierten (extraauralen) Lautsprechern der Meta Quest Pro.“

Die Positionen der Soundquellen konnten nicht punktgenau verortet werden (Frage 2), interessant ist dabei die Abweichung der Differenzen. Eine Schallquelle wurde durchschnittlich um 1.3 Meter in der Distanz verfehlt. Es zeigten sich Unterschiede zwischen nahen und fernen Schallquellen, die in dieser Testung jedoch nicht berücksichtigt wurden. Testpersonen gaben an, dass entferntere Objekte schwerer zu verorten waren als jene, die direkt vor ihr erklangen. Die durchschnittliche Abweichung der Winkeldifferenzen betrug etwa 15 Grad. Die

### 3 Testen der Spatializer

---

Alternativhypothese bestätigt sich, dies beantwortet die Fragestellung folgendermaßen: Eine verortete Soundquelle wurde mit dem DearVR Spatializer insgesamt im Durchschnitt um etwa 1.3 Meter in der Distanz und circa 15 Grad im Azimut verfehlt. Dieses Ergebnis ist auch außerhalb der Toleranzbereiche, welche in dieser Masterarbeit bereits definiert wurden. (siehe 2.1.6 Toleranzbereiche)

Der Hörtest konnte die Forschungsfrage 2 für diese Vorstudie beantworten, jedoch lässt sie Forschungsfrage 1 in Teilen offen. Um den Hörtest zu verbessern und die Fragestellungen der Masterarbeit zu beantworten wurden einige Anpassungen vorgenommen:

Wichtig ist die verwendete Wiedergabe-Hardware. Da alle Dokumentationen der Audio Spatializer empfehlen das verarbeitete Audiosignal mit Kopfhörer wiederzugeben (siehe 2.3 Audio Spatializer), sind geschlossene Kopfhörer mit einer noch geringeren Impedanz besser geeignet, dass sie eine höhere Lautstärke für die Testdurchläufe gewährleisten.

Auch der Testablauf sollte eine „Warm-Up“ Phase für die Testperson beinhalten, in welcher die grundlegenden Funktionen präsentiert werden und das Setzen der Marker geübt werden kann. Außerdem kann sich der:die Partizipant:in an die vorgespielte Stimme gewöhnen. Dadurch sollen versehentliche Fehleingaben und verhindert und die Lokalisierungsfähigkeit verbessert werden. Um den Raum stets im Nullpunkt zu halten und keine Verschiebung beim Umsetzen der Brille zu riskieren ist es von Nöten, mit den Spatial Anchors der Meta Unity API<sup>36</sup> zu arbeiten. Dadurch kann garantiert werden, dass das Modell sich zwischen keinem der Testpersonen verschiebt. Die Datenaufbereitung könnte bereits in dem Testprogramm stattfinden. Werte wurden nun lediglich als Vector gespeichert. Von zusätzlichem Interesse ist eine direkte Auswertung mit graphischer Darstellung der Soundquellen und die Abweichungen pro Sample. Eine höhere Anzahl an Testpersonen (oder Messwiederholungen) ist erforderlich.

---

<sup>36</sup> <https://developer.oculus.com/documentation/unity/unity-spatial-anchors-overview/>

## 3.2 Testdesign

Um vier Audio Spatializer für Mixed Reality Applikationen in ihrer Lokalisierung virtueller Klangquellen miteinander vergleichen zu können, werden vier Mixed Reality Applikation erstellt.

Das Testdesign ist dabei dem der Vorstudie angelehnt. (Siehe 3.1.1 Testdesign)

10-20 Personen werden dabei die vier Mixed Reality Applikationen vorgeführt. Diese sollen eine definierte Anzahl an Samples (virtuellen Klangquellen) der die durch die Audio Spatializer auralisiert wurden in den Variablen Azimuth und Entfernung zum Nullpunkt schätzen. Die Werte werden von der Applikation lokal hinterlegt und können daraufhin vom Versuchsleiter mithilfe einer einfaktoriellen Anova<sup>37</sup> miteinander verglichen und auf Signifikanz geprüft werden. Die Testpersonen werden mithilfe eines Online-Fragebogens durch den Hörtest geleitet. Nach einer Pseudonymisierung mithilfe einer durch die Testleitung festgelegte Teilnehmeridentifikation, welche von der Testperson im Online-Fragebogen zur Zuweisung einzugeben ist, folgt eine grundsätzliche demographische Datenerhebung zur Person. Da alle Testpersonen jeden Audio Spatializer getestet ist jede Testperson in jeder der vier Gruppen (vier Audio Spatializer) zu finden. Sie werden nach jeder Applikation gebeten im Online-Fragebogen den Gesamteindruck zu Lokalisierbarkeit und Auralisierung der virtuellen Klangquellen der soeben getesteten Applikation auf einer Skala zu hinterlassen.

Ein Datensatz (Population) besteht aus der geschätzten X und Y Position im virtuellen Raum zur korrespondierenden virtuellen Klangquelle pro Audio Spatializer. Aus dieser Position und der Position der Testperson im Raum lässt sich die geschätzte Distanz sowie der Azimuth errechnen. Eine perfekte Lokalisierung eines Samples ist also dann von der Testperson erreicht, wenn die X und Y Position mit der X und Y Position der virtuellen Klangquelle übereinstimmt. Zusätzlich besteht eine Population aus den im Online-Fragebogen erhobenen Daten, um eine Verbindung zwischen dem persönlichen Eindruck der Applikation auf die Testteilnehmer:innen und den gemessenen Daten herzustellen.

Insgesamt werden der Testperson pro Applikation 21 Samples (virtuelle Klangquellen) vorgeführt, die sie zu verorten hat. Dies ergibt bei vier Applikationen

---

<sup>37</sup> <https://www.ibm.com/docs/de/spss-statistics/25.0.0?topic=features-one-way-anova>

84 (21x4) Stimuli und entspricht einem "full factorial repeated within subject design"<sup>38</sup>.

Pro Sample werden nach kurzer Eingewöhnungsphase maximal 10 Sekunden zur Lokalisierung benötigt. Dies entspricht bei 21 Samples pro Applikation einer Testdauer von maximal 5 Minuten inklusive Applikationswechsel. Die gesamte Testzeit pro Person wird auf maximal 30-45 Minuten inklusive Fragebogen geschätzt.

Da der virtuelle Raum ein Abbild des Motion Capturing Lab der FH St.Pölten ist, kann der Test auch nur dort durchgeführt werden. Die Testpersonen wurden zu für sie geeignete Zeiten dorthin eingeladen.

## 3.3 MR-Applikation

### 3.3.1 Applikationsdesign

Der Ablauf der Applikation ist in Abbildung 9 skizziert. Nach dem Start der Applikation muss der Versuchsleiter das virtuelle 3D Modell mit dem reellen Raum örtlich synchronisieren. Diese „MR-Synchronisation“ ist relevant für die korrekte Position der virtuellen Klangquellen im Raum. Dabei wird versucht so präzise wie möglich durch Rotation und Zentrierung das Modell mit der echten Räumlichkeit zu verschmelzen. Es besteht die Möglichkeit zwischen dem 3D Modell und der Passthrough-Ansicht zu wechseln, um den Prozess zu vereinfachen. Gelingt dies, wird mithilfe sogenannter „Spatial Anchor“ die Raumposition durch die Meta Quest 3 abgespeichert. Beim Öffnen der Applikation ist dann ein erneutes Synchronisieren nicht mehr nötig. Die Position des Raumes wird als Spatial Anchor geladen und automatisch auf das 3D Modell gesetzt. Ist dies abgeschlossen, kann die Meta Quest 3 der Testperson übergeben werden.

Jedes gehörte Sample ist nun von der Testperson auditiv zu verorten und mithilfe eines Markers soll die geschätzte Position gespeichert werden. Der:die User:in befindet sich dabei im Passthrough-Modus. Das "Passthrough"-Feature erlaubt es dem:er User:in seine reale physische Umgebung in Echtzeit zu sehen. Mittels der Sensoren und Kameras an VR-Headset reproduziert Passthrough das, was der Nutzer visuell wahrnehmen würde, wenn er unmittelbar durch die Vorderseite seines Headsets in seine tatsächliche Umgebung schauen könnte. Die Testperson

---

<sup>38</sup> <https://bookdown.org/ndphillips/YaRrr/full-factorial-between-subjects-anova.html>

### 3 Testen der Spatializer

---

ist in der Lage sich, um die eigene Achse zu bewegen, jedoch angehalten sich örtlich nicht zu bewegen bzw. innerhalb eines halben Meter Radius zu bleiben. Nachdem die Testperson alle 21 Samples gehört und verortet hat, wird das Testergebnis als JSON-Datei lokal auf dem VR-Headset gespeichert. Es handelt sich um eine 6DoF Anwendung, auch wenn die Begrenzung auf den halben Meter Bewegungsradius eine Fixierung auf den Null Punkt für die spätere Auswertung erlaubt. Dafür wird zusätzlich die Kopfposition zum Zeitpunkt der Markierung im Raum gespeichert.

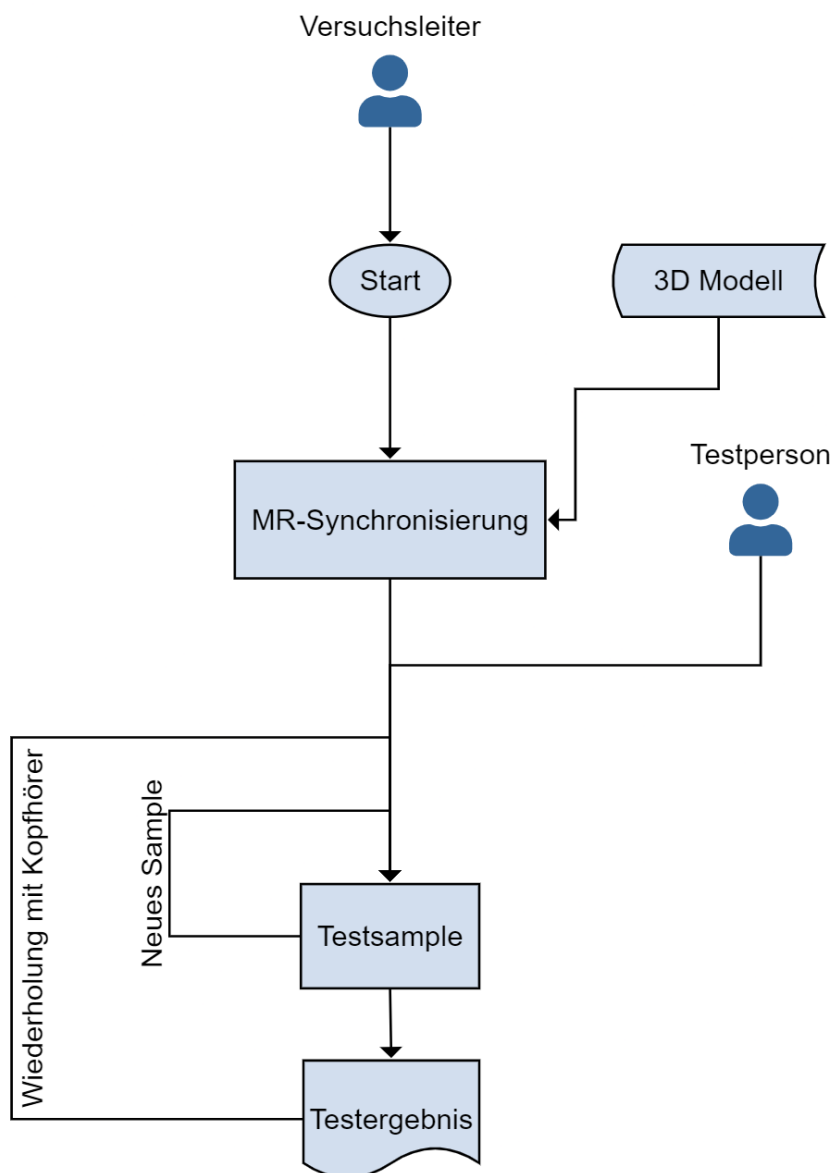


Abbildung 9: Ablauf einer Testapplikation

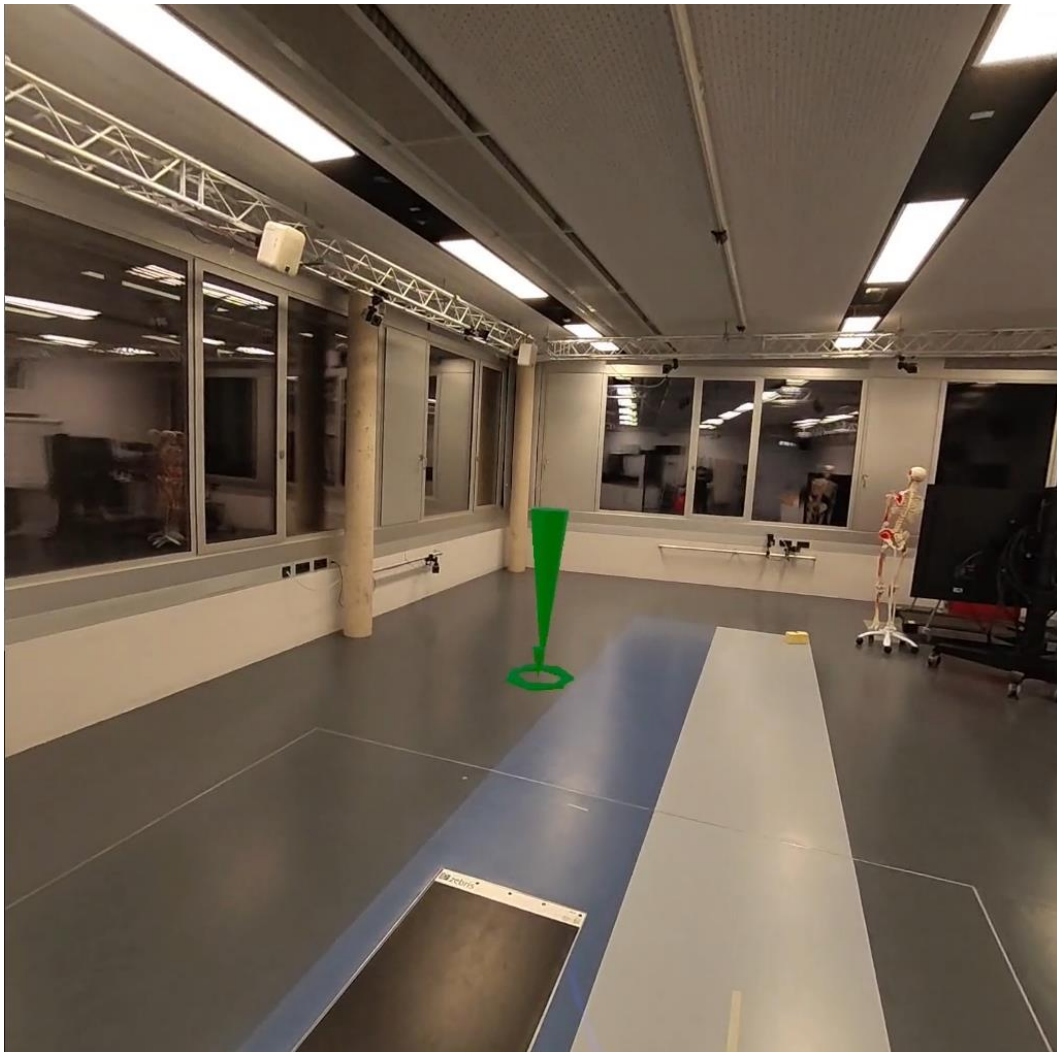


Abbildung 10: Platzieren eines Markers (grün) während eines Testdurchlaufs durch die Testperson. Der Passthrough-Modus ist aktiv.

#### 3.3.2 Voraussetzungen

Die Mixed Reality Applikation wird in Unity<sup>39</sup> erstellt. Die Entwicklung und der Hörtest selbst findet mit einem Meta Quest 3 Headset statt. Am 1. November 2023 wurde von Unity das neue OpenXR: Meta package in der Version 1.0.0 veröffentlicht. (*Cross-Platform Mixed Reality Development on Meta Quest 3*, o. J.) Dieses basiert auf Unitys AR Foundation Package und dem OpenXR plug-in und bietet eine API zur Entwicklung von Mixed Reality Applikationen auf unterschiedlichen Plattformen wie Meta Quest, HoloLens oder iOS und Android.

---

<sup>39</sup> <https://unity.com/de>

### 3 Testen der Spatializer

---

Zusätzlich dient es als Framework für unterschiedlichste Anforderungen an MR-Applikationen wie beispielsweise die Interaktion mithilfe der Hände (Hand Tracking) oder das Erkennen von Oberflächen (Plane Detection) und das Passthrough Feature. Außerdem ermöglicht es den Zugriff zu den „Spatial Anchors“. Diese helfen in MR und AR-Applikationen in der realen Welt platzierte Objekte zu „verankern“. Verankerte Objekte werden dadurch immer korrekt an der entsprechenden Stelle angezeigt. Das Unity AR Foundation Package erlaubt jedoch nur eine temporäre Speicherung des Spatial Anchors während der Laufzeit der Applikation. Da die Testapplikationen während des Tests gewechselt werden, der synchronisierte virtuelle Raum aber stets an derselben Stelle platziert werden soll, ist ein persistenter Spatial Anchor notwendig der über die Laufzeit hinweg auf dem VR-Headset hinterlegt ist.

Bis das OpenXR: Meta package diese Funktionalität bietet, musste für das Spatial Anchor Feature auf das Meta XR SDK in Kombination mit dem Oculus XR Plugin<sup>40</sup> zurückgegriffen werden. Dieses nutzt die „Shared Spatial Anchor“, welche in der physischen Umgebung hinterlegt und abgerufen werden können. (*Use Shared Spatial Anchors: Unity | Oculus Developers*, o. J.)

Die Unity Version zur Entwicklung der Hörtest Applikationen ist 2023.1.20f1. Unity wurde auf einem Windows PC installiert. Um den Code zu schreiben und zu debuggen wurde die Visual Studio 2022 Community Edition IDE (=Integrated Development Environment)<sup>41</sup> verwendet.

Während der Entwicklung werden Audiosignale direkt aus Unity über ein Zoom UAC-2<sup>42</sup> und DT 990 Pro mit 250 Ohm von beyerdynamic Kopfhörern<sup>43</sup> ausgegeben.

Da der Hörtest im Motion Capturing Lab der FH St.Pölten stattfand wurde außerdem ein maßstabsgetreues 3D Modell benötigt, welches die Wände, Decke, Fußböden und statische Möbel und Hindernisse (Tische, Säulen, ...) des Raumes abbilden. Dieses wurde von wissenschaftlichen Mitarbeiter:innen der FH St.Pölten zur Verfügung gestellt und liegt als Unity Prefab vor.

Weiters wurde eine Audiodatei für den Hörtest mit der Stimme des Autors der Arbeit produziert. Diese diente in Unity als virtuelle Klangquellen und sollte im

---

<sup>40</sup> <https://developer.oculus.com/documentation/unity/unity-xr-plugin/>

<sup>41</sup> <https://visualstudio.microsoft.com/de/vs/community/>

<sup>42</sup> <https://zoomcorp.com/en/jp/audio-interface/audio-interfaces/uac-2/>

<sup>43</sup> <https://www.beyerdynamic.de/dt-990-pro.html>

Hörtest lokalisiert werden. Dabei wurde der Satz „Dies ist ein Hörtest. Bitte hören Sie sorgfältig zu und versuchen sie die Stimme bestmöglich zu lokalisieren.“ mithilfe eines AKG C214<sup>44</sup> Großmembran Kondensatormikrophons in der DAW Reaper<sup>45</sup> aufgenommen. Das Sample wurde durch einen Equalizer gefiltert, um das Signal akustisch zu säubern und die für die Sprache relevanten Signalanteile hervorzuheben. Exportiert wird die.wav Datei mit 48kHz Abtastrate und 16 Bit pro Sample.

#### 3.3.3 Unity Projekt

In Unity wurde ein Projekt angelegt. Dieses wurde zunächst Schritt für Schritt nach der offiziellen von Meta veröffentlichten Ressource konfiguriert. (*Oculus-Entwicklerzentrum | Integration der Unity-Spiele-Engine für VR*, o. J.)

Das gesamte Projekt ist mit der Versionierungsverwaltungssoftware Git öffentlich auf dem Github<sup>46</sup> abrufbar.

Zunächst wurde eine neue Unity Scene unter dem Ordner Scenes erstellt. Diese diente für die vier Applikationen als Scene-Template und beinhaltet alle Ressourcen wie das 3D Modell, UI oder die Samples in Form von Unity Prefabs. Das Unity Prefab-System gestattet die Erstellung, Konfiguration und Speicherung eines GameObjects in Form eines wiederverwendbaren Assets, welches sämtliche seiner Komponenten, Eigenschaftswerte und untergeordneten GameObjects umfasst. Diese Prefab-Assets fungierten als Vorlagen, anhand derer neue Instanzen in der Szene erzeugt werden konnten. (Technologies, o. J.-b)

Die Meta XR Core SDK<sup>47</sup> umfasst zahlreiche essenzielle Funktionen, die von den Meta XR-Headsets bereitgestellt werden, darunter die Kamerakonfiguration, Controller, Hand-Tracking und die für diese Applikation relevanten Spatial Anchor. Dieses wird in das Projekt integriert. Daraufhin wurde das Oculus XR Plugin über das XR-Plugin Management in den Projekteinstellungen installiert. Nun wurden im Oculus Project Setup Tool alle ausstehenden Issues und empfohlenen Änderungen angewendet.

---

<sup>44</sup> <https://at.akg.com/condenser-microphones/C214.html>

<sup>45</sup> <https://www.reaper.fm/>

<sup>46</sup> <https://github.com/EvoZ1310/SpatSound>

<sup>47</sup> <https://developer.oculus.com/downloads/package/meta-xr-core-sdk/>

### 3 Testen der Spatializer

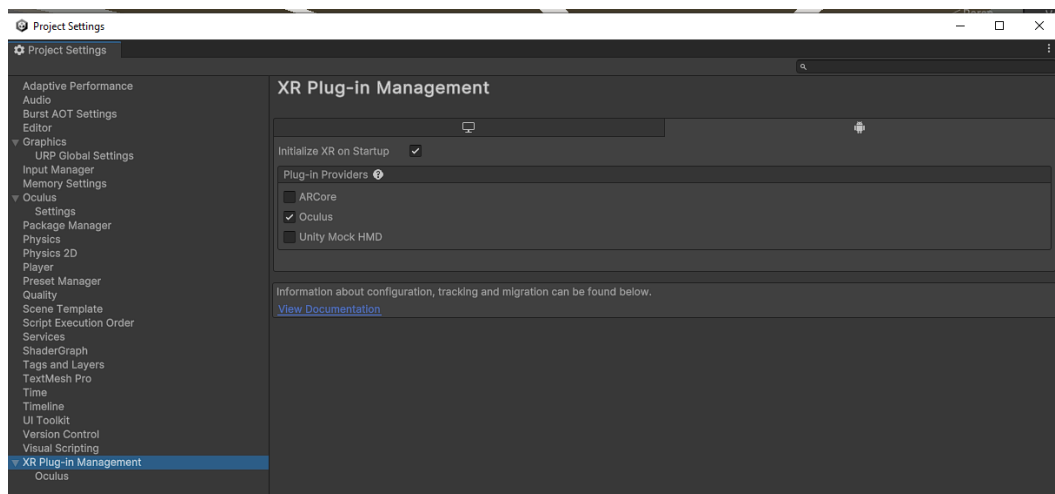


Abbildung 11: Installation des Oculus XR Plugins

Nun konnte das OVRCameraRig Prefab aus den Packages in die Scene gezogen werden. Es enthält neben der Camera (CenterEyeAnchor) ein Model der rechten und linken Controller, welche über die echten Controller während der Laufzeit geladen wird. Außerdem beinhaltet es die OVRManager Komponente, welche essenziell für die Entwicklung auf Meta Quest Geräten ist. Hier werden Endgeräte (Quest 2, Quest Pro, Quest 3) gewählt sowie Features aber auch Experimentelle Features der Quest an und abgeschaltet.

Der Tracking Origin Type wurde auf Floor Level gesetzt, um den Nullpunkt der virtuellen Welt auf Fußbodenhöhe zu setzen. Für das verwenden der Spatial Anchor wurde der Anchor Support, Shared Spatial Anchor Support und Scene Support eingeschaltet. Um Passthrough zu ermöglichen, wurde auch der Passthrough Support eingeschaltet sowie die Checkbox bei "Insight Passthrough" > „Enable Passthrough“ aktiviert.

### 3 Testen der Spatializer

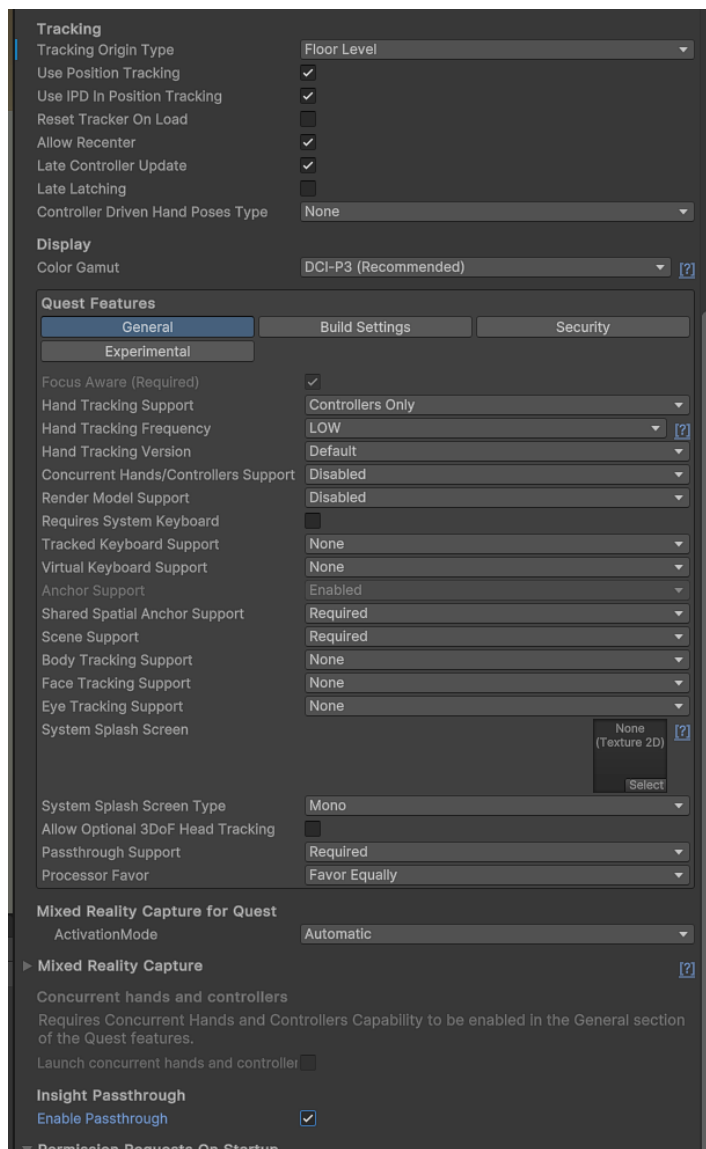


Abbildung 12: Konfiguration der OVR Manager Komponente am OVR Camera Rig GameObject.

Dem GameObject wurde außerdem die OVR Passthrough Layer Komponente hinzugefügt. Dies ermöglicht die Passthrough Funktion. Um während der Entwicklung über Quest Link diese Funktion nutzen zu können wurde Passthrough nach der offiziellen Dokumentation konfiguriert. (*Get Started with Passthrough | Oculus Developers*, o. J.)

Der `AudioListener` auf dem `CenterEyeAnchor` GameObject dient der Audioausgabe. So empfängt es die zur Auralisierung notwendigen Daten der virtuellen Klangquellen, verarbeitet sie durch den Audio Spatializer und gibt sie über die Lautsprecher des Endgerätes wieder.

## 3 Testen der Spatializer

### 3.3.4 3D Model

Das 3D Modell des Motion Capturing Labs liegt als .unitypackage unter dem Namen „RemoCapLab.unitypackage“ vor. Dieses kann mit einem Doppelklick in das Projekt importiert werden. Danach wurde für Texturen, Materials sowie Meshes ein Ordner angelegt und die importierten Dateien in die entsprechenden Ordner kopiert. Die Materialien können jedoch nicht sofort verwendet werden, da sie mit der URP (=Universal Render Pipeline) arbeiten. Daher muss die URP über den Package Manger im Projekt nachinstalliert werden. Für die globalen Render und Qualitätseinstellungen muss außerdem ein Universal Render Pipeline Asset angelegt und als Scriptable Render Pipeline Settings in den Graphikeinstellungen der Projekteinstellungen eingetragen werden. (*Installing the Universal Render Pipeline into an existing Project | Universal RP | 17.0.1, o. J.*)

Das 3D Modell wurde in der Szene so platziert, dass der Nullpunkt der Szene der Mitte der ersten Fläche der Kraft-Mess-Platte im Gangbereich des Raumes entspricht. Dies ist auch der Bereich, in welchem sich die Testperson während des Hörtests aufhalten soll.

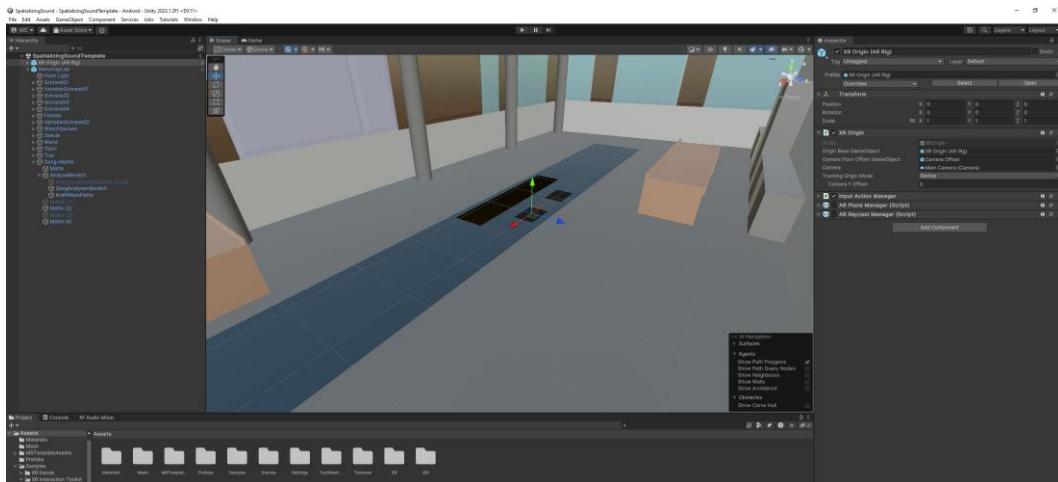


Abbildung 13: Platzierung des Motion Capturing Labs 3D Modells in der Unity Szene

Da dieses 3D Modell nur zum Abgleich mit der realen Welt dient und die Audio Spatializer die Dimensionen des Raumes zur Auralisierung nutzen, wird noch ein „unsichtbares“ 3D Modell benötigt. Dabei wurden alle Elemente des Raumes mit einem unsichtbaren Material versehen. In diesem unsichtbaren Raum befanden sich die Testpersonen während des Tests. Durch den Passthrough Modus konnte sie die reelle Welt sehen, welche durch die unsichtbaren Wände des 3D Modells angezeigt wird. Diese Verschmelzung von der sichtbaren realen Welt und dem 3D

### 3 Testen der Spatializer

---

Model erzeugt die Mixed Reality. Der mit dem unsichtbaren Material versehene Raum wurde ebenfalls als Prefab abgespeichert.

Weiters wurde die vorproduzierte Audio Datei „hoertestvocal.wav“, welches an den unterschiedlichen Positionen im Raum verortet werden soll, in den Ordner Audio importiert.

Daraufhin wurden im gesamten Raum insgesamt 21 Positionen für die Testsamples zufällig verteilt. Zunächst wurde dafür eine gleichmäßige Verteilung der Samples im Raum mit 2m Abstand zueinander gewählt. Jede Position wurde jedoch vom Versuchsleiter leicht angepasst, um eine Systematik in der Platzierung der Positionen durch alleiniges Hören der Testperson nicht zu erkennen. Das Testsample wird in Unity als Prefab „SoundSource“ abgespeichert. Dabei beinhaltet es eine `AudioSource` mit der `hoertestvocal.wav` als `Audioclip` integriert. Der Output ist zu diesem Zeitpunkt noch nicht gesetzt, da dieser pro Audio Spatializer unterschiedlich behandelt wird. In Abbildung 14 wird jede Position als gelbe Rechtecke in der Vogelperspektive gekennzeichnet:

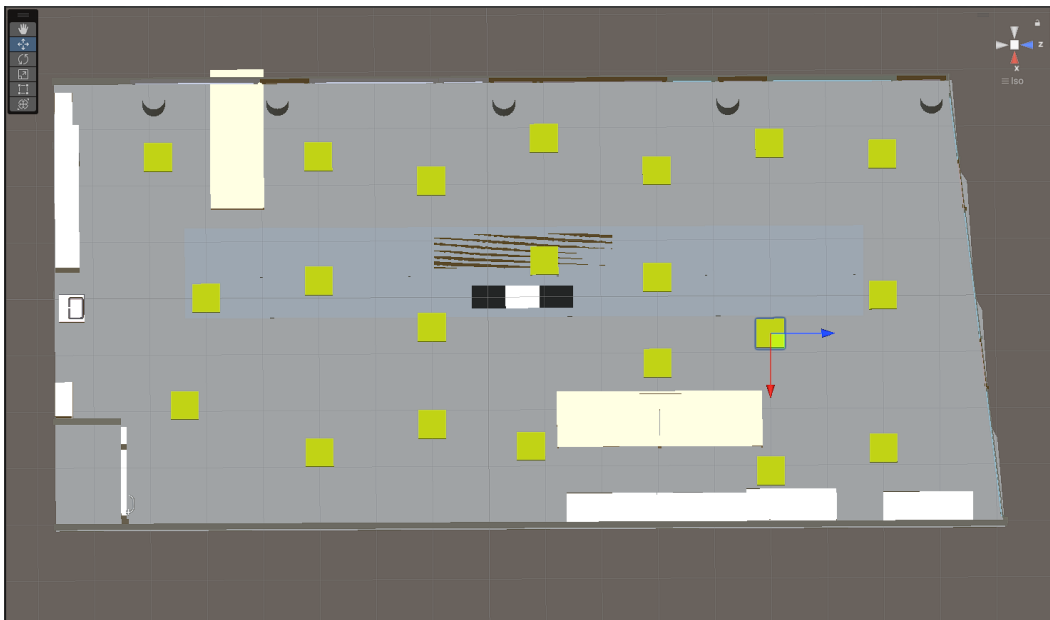


Abbildung 14: Positionen der zu verortenden Samples im 3D Model

#### 3.3.5 Interface

Alle Eingaben des Users werden mit den Meta Quest 3 Controllern durchgeführt. Die 3D Modelle der Controller zur Interaktion wurden über das Prefab `OVRInteraction` aus der Meta SDK integriert. Darin wurde das

### 3 Testen der Spatializer

---

OVRControllers sowie ControllerRayInteractor Prefabs eingefügt, um mit der UI über RayInteraction interagieren zu können.

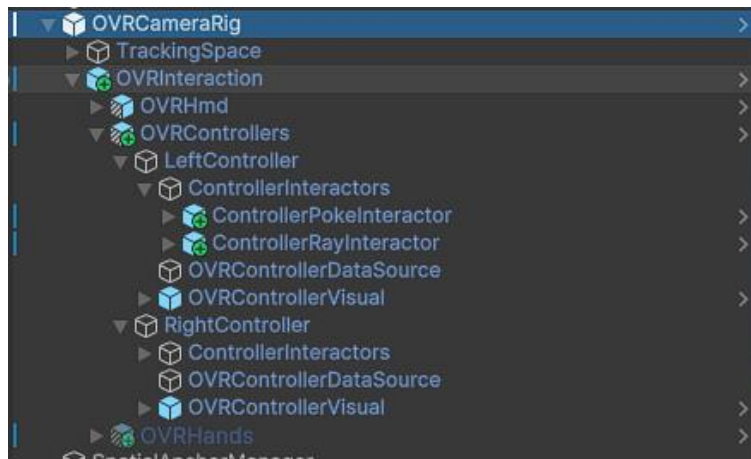


Abbildung 15: OVRCameraRig mit integrierten OVRController und RayInteractor

Das Synchronisieren des 3D Modells geschieht über ein Panel, in welchem der Versuchsleiter die Position und Rotation des 3D Modells anhand seiner aktuellen Position und der Kopffrotation des Kopfes bestimmen kann. Dadurch ist es möglich sich auf den Nullpunkt des 3D Modells im reellen Raum zu platzieren. Durch einen Toggle Button ist er außerdem in der Lage zwischen dem Passthrough Modus und dem 3D Modell (VR) hin und her zu wechseln und dadurch die Genauigkeit der Platzierung abgleichen zu können.

Um mit Panels über die OVRRayInteractors interagieren zu können wird eine PointableCanvasModule Komponente in der Unity Szene benötigt. Zusätzlich benötigt das Canvas eine PointableCanvas, PointableCanvasUnityEventWrapper und RayInteractable Komponente. (*Unity Canvas Integration | Oculus Developers, o. J.*)

Nachdem das Synchronisieren hinreichend genau geschehen ist, soll der Testperson die Meta Quest 3 aufgesetzt werden. Sie sieht ein Panel vor sich, in dem eine kurze Erklärung des nun folgenden Hörtests beschrieben ist. Sollte sie bisher keine der vier Applikationen durchgeführt haben, wird das Warm-up empfohlen, um besser mit der Mechanik der Quest Controller und dem Hörtest vertraut zu werden. Danach folgt der eigentliche Hörtest mit 21 Klangquellen. Am Ende des Tests wird der:die User:in gebeten die Daten über einen Speichern Button anonym zu hinterlegen. Anschließend soll er das VR-Headset an die Versuchsleitung übergeben. Je nachdem welche Anwendungen der:die User:in bereits durchgeführt hat, wird die nächste Applikation von der Versuchsleitung geladen und der Testperson übergeben.

3.3.6 Flussdiagramm

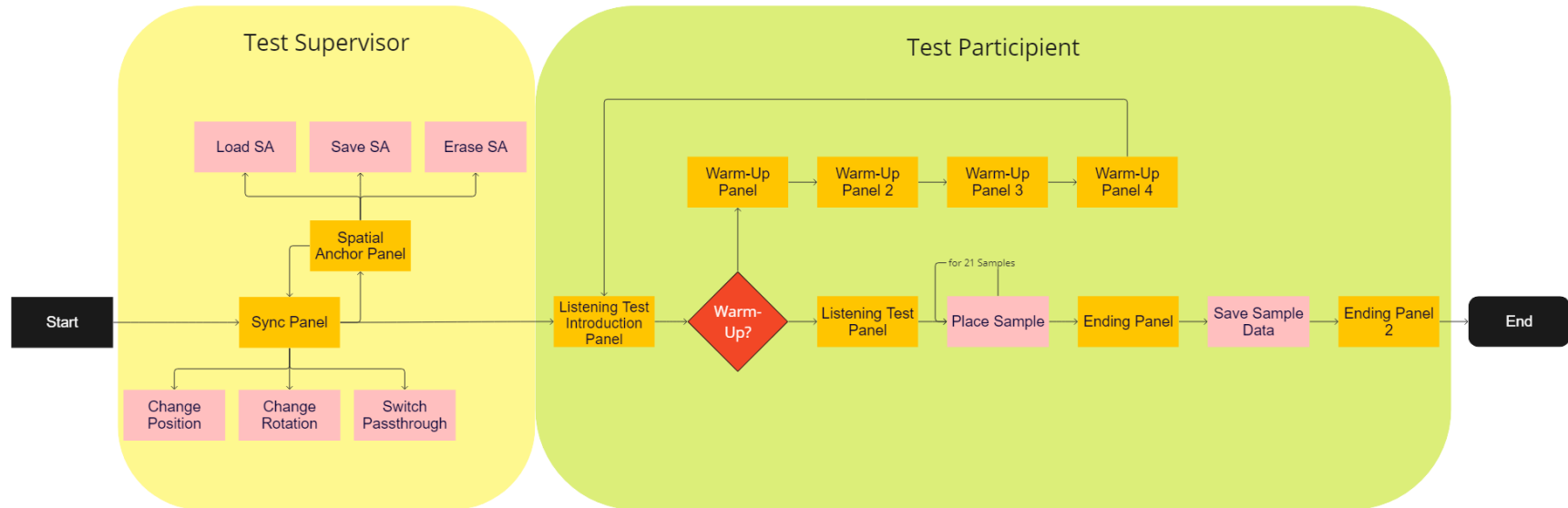


Abbildung 16: Flussdiagramm der Hörtest Applikation

In Abbildung 16 ist der Ablauf der Applikation vom Start bis zum Ende abgebildet. Gelbe Elemente zeigen UI-Ansichten an den:die User:in. Diese können entweder die Versuchsleitung (Test Supervisor) oder die Testperson (Test Participant) sein. Rosa gekennzeichnet sind Aktionen, Rot stellt eine notwendige Entscheidung dar.

### 3.3.7 Implementierung

Im Ordner Scripts wurden alle zur Funktionalität der Applikation notwendigen C# Skripte abgespeichert. Die wichtigsten und ihre Funktion soll nachfolgend dargestellt werden:

#### 3.3.7.1 RoomSyncPanel

Die Klasse `RoomSyncPanel` ist für Funktionalitäten des Sync Panels und des Spatial Anchor Panels verantwortlich. Insbesondere bietet es Funktionen an, auf welche die Buttons des Sync Panels durch das `OnClick` Event zugreifen können. Für die Buttons „Rotieren“ und „Neu zentrieren“ werden Funktionen zum Positionieren des 3D Modells angeboten. Dadurch soll der virtuelle Raum mit dem echten Raum örtlich synchronisiert werden. Die `SwitchPassthrough()` Funktion blendet das 3D Model des Raumes (`GameObject RemoCapLab Sync`) ein bzw. aus. `ShowSpatialAnchors()` wechselt die Ansicht des Panels zum Spatial Anchor Panel. Das Laden, Speichern und Löschen der Spatial Anchors wird ebenfalls über diese Klasse zur Verfügung gestellt.

#### 3.3.7.2 SpatialAnchorManager

In der `SpatialAnchorManager` Klasse werden Spatial Anchors verwaltet. Es dient zum Laden von bereits gespeicherten Spatial Anchors, speichern und überschreiben neuer Spatial Anchors, sowie das Löschen bereits vorhandener Spatial Anchors lokal auf der Meta Quest 3. Dafür benötigt die Komponente einen für das Programm eindeutigen Uuid Key. Dieser dient zur wiederfinden des Spatial Anchors über die Laufzeit hinweg. Ein Spatial Anchor Prefab, welches die OVR Spatial Anchor Komponente beinhaltet, sowie den Raum dessen Position es zu laden und speichern gibt. (*Spatial Anchors Overview: Unity | Oculus Developers*, o. J.)

#### 3.3.7.3 ListeningTestPanel

Die `ListeningTestPanel` Komponente steuert die Ansicht des Hörtest Panels. Dieses wird der Testperson angezeigt.

## 3 Testen der Spatializer

---

### 3.3.7.4 *PanelController*

Der `PanelController` erlaubt das Aktivieren und Deaktivieren der zwei Panels `ListeningTestPanel` und `RoomSyncPanel`. Während des Hörtests wird das `ListeningTestPanel` deaktiviert und am Ende wieder aktiviert.

### 3.3.7.5 *WarmUpController*

Der `WarmUpController` ist für die Erkennung der Usereingaben während der Warm-Up Phase verantwortlich. Es besitzt Unity Events, welche zur Steuerung des Hörtest Panels dienen.

### 3.3.7.6 *SimpleGazeCuror*

Dieses Skript projiziert mithilfe eines Raycasts einen Cursor am Boden, mit welchem die Position der virtuellen Klangquelle verortet werden soll. Dazu existiert ein Model des Cursors als Prefab. Damit der Raycast den Boden als potenziellen Hit erkennen kann, muss dem Boden des unsichtbaren Motion Capturing Lab Prefab ein Mesh Collider hinzugefügt werden. Der Raycast wird vom rechten Controller aus durchgeführt. Ressourcen dazu wurden von Immersive Limit zur Verfügung gestellt. (*Simple Gaze Cursor*, o. J.)

### 3.3.7.7 *PositionCircle*

`Position Circle` zeichnet mithilfe eines `LineRenderers` einen Kreis mit vorgegebenen Segmenten und Radius. Er wird als Prefab `PositionCircle` abgespeichert und ist Kind des `MotionCapturingLab` `GameObject`.

### 3.3.7.8 *ListeningTest*

Das `ListeningTest` Skript implementiert die Logik für den eigentlichen Hörtest. Die Methode `StartListeningTest()` erzeugt unter anderem eine Instanz der `ListeningTestData` Klasse. Diese beinhaltet wiederum eine Liste von `ListeningTestDataPair`, welches zu jedem Hörsample den Namen der `AudioSource` über das `SampleIdentifier` Skript, die Position der `AudioSource` sowie die Position des platzierten Markers hinterlegt. Außerdem wird die benötigte Zeit zum Orten der Klangquelle gespeichert. Die `ListeningTestData` Klasse kann mithilfe des `JsonSerializers` als `.json` Datei serialisiert und auf dem Gerät über die Methode `WriteTestResults()` in der `ListeningTest` Komponente geschrieben werden.

Neben dem zufälligen Anordnen der `AudioSources` für den Test, werden außerdem die Usereingaben und das Erzeugen des `ListeningTestMarker` hier abgehandelt. Dafür wird ein Prefab benötigt, welches ein in die Höhe

### 3 Testen der Spatializer

---

transformierter Cursor Prefab ist. Es wird mithilfe des `SimpleGazeCursors` am Boden instanziiert und speichert somit die Position des verorteten Markers.

Der `ListeningTestName` soll pro Applikation eindeutig sein. Er erzeugt im Dateisystem der Quest 3 einen eigenen Ordner pro Applikation. Die `.json` Datei mit den zugehörigen Testdaten wird dann in diesem Ordner abgespeichert und kann nach dem Test über USB-Kabel ausgelesen werden.

Über die Methoden `StartWarmup()` und `EndWarmup()` kann vom Hörtest Panel über die entsprechenden Buttons das Warm Up gestartet und beendet werden. Es beinhaltet dieselbe Logik wie der eigentliche Hörtest, jedoch werden dabei keine Daten hinterlegt.

#### 3.3.7.9 *RotateToCamera*

Das `RotateToCamera` Skript erlaubt es, einer `AudioSource` stets auf die `MainCamera` gerichtet zu sein. Dies versichert einigen Spatializern durch die Option einer Richtcharakteristik (`Directivity Pattern`) die korrekte Rotation zum `AudioListener`.

## 3.4 Auralisierung

### 3.4.1 Kunstkopfaufnahmen vor Ort

Um den Partizipanten eine virtuelle Hörumgebung präsentieren zu können, die der realen Hörumgebung des Motion Capturing Labors ähnelt, wird eine Referenzaufnahme mithilfe eines Kunstkopfes für die Entwicklung produziert. Diese Referenzaufnahme des Raumes kann dann zum Abgleich beim Einstellen der Parameter der einzelnen Audio Spatializer verwendet werden.

Zu Beginn wurden Kunstkopfaufnahmen für Musikproduktionen eingesetzt. Seit über 40 Jahren werden sie jedoch auch intensiv beispielsweise in der Automobilindustrie genutzt, um den Klang im Fahrzeuginnenraum zu optimieren. Nur ein binaurales Aufnahmesystem wie ein künstlicher Kopf ermöglicht es den Zuhörern, mit kalibrierten und ausgeglichenen Kopfhörern sehr realistische Höreindrücke wahrzunehmen, die dem originalen Klangfeld ähneln. (Genuit, 2018, S. 1)

Diese Kunstkopfaufnahmen dienen als Referenz um die Nachhallparameter eines Audio Spatializer so gut wie möglich der realen Umgebung im Motion Capturing Lab reproduzieren zu können, wenn dieser die entsprechenden Optionen und Parameter zur Verfügung stellt.

Der Raum befindet sich an der FH St. Pölten und besitzt eine Tiefe von etwa 17 Metern, eine Höhe von etwa 3 Meter und eine Weite von etwa 8 Meter. Dies führt zu einem Volumen von insgesamt circa 408 m<sup>3</sup>.

Bei der Aufnahme wird der Kunstkopf in der Mitte des Raumes platziert, exakt dort, wo sich auch die Partizipant:innen während des Hörtests befinden. Insgesamt wurden 8 Positionen mit unterschiedlichen Entfernungen zum Kunstkopf gewählt. Auf diesen Positionen wurde pro Aufnahme ein Lautsprecher platziert. So konnten insgesamt 8 binaurale Aufnahmen produziert werden, die als Referenz dienen.

### 3 Testen der Spatializer

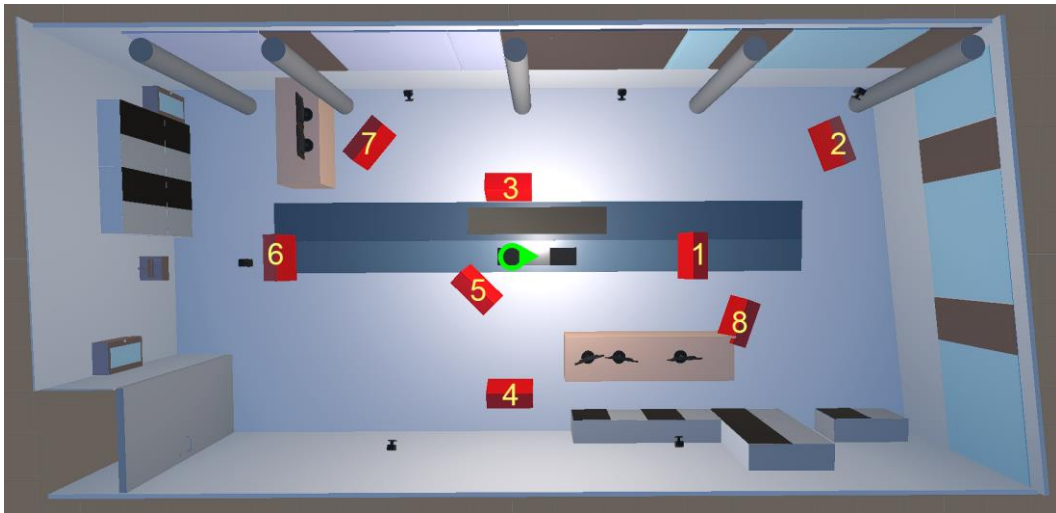


Abbildung 17: Binauralaufnahmen für 8 Monitorpositionen im virtuellen MoCapLab

In Abbildung 17 sind alle 8 Positionen angeführt. Der grün markierte Kreis in der Mitte zeigt die Position sowie Ausrichtung des Kunstkopfes. Der Pfeil ist entlang der Nase ausgerichtet. Diese Richtung ist „vorne“. Insgesamt wurde für jede Position zwei Kunstkopfaufnahmen produziert. Eine Aufnahme, in welcher der Kunstkopf nach vorne blickt und eine, in welcher der Kopf Richtung Klangquelle, also dem Monitor, geneigt ist. Der Monitor befindet sich am Boden und wurde Richtung Kunstkopf gekippt. Eine Neigung des Kopfes findet nur anhand einer Änderung des Azimuth statt.



Abbildung 18: Kunstkopfaufnahme Position 3 mit Blickrichtung vorne.

Für die Kunstkopfaufnahmen wurde der Kunstkopf Davinci Head MK2 genutzt.<sup>48</sup>

#### 3.4.2 Raumakustische Messung

Beim digitalen Nachbau des Klangs der virtuellen Räume mithilfe der Audio Spatializer ist auch der Nachhall eine wichtige Komponente. Die Nachhallzeit lässt sich mithilfe des Parameters RT60 (= Reverberation Time) objektivieren. Die RT60 kann mithilfe eines Messmikrophons und einer entsprechenden Messsoftware berechnet werden.

Dieser Wert wurde im Anschluss auch digital aus den virtuellen Räumen gemessen. Wenn der Audio Spatializer dies über entsprechende Parameter zulässt, kann somit die Nachhallzeit soweit an den RT60 Wert des realen Raumes adjustiert werden, dass sie sich realer und virtueller Raum hinsichtlich der Nachhallzeit nur leicht unterscheiden.



Abbildung 19: Setup für die raumakustische Messung mit der Software REW.

Das Messmikrophon wurde auf 1,8m Höhe in der Mitte des Raumes positioniert. Der Lautsprecher in 4m Entfernung auf das Mikrophon gerichtet. Die Messungen mithilfe der Software REW zeigen folgende RT60 Ergebnisse:

Zwischen 60Hz und 250Hz klingt der Raum mit 550-600ms nach. Die Nachhallzeit steigt dann zwischen 250Hz und 1kHz kontinuierlich an, bis sie schließlich

---

<sup>48</sup> <https://www.davincihead.com/product/davincihead/>

### 3 Testen der Spatializer

zwischen 1kHz und 3kHz bis zu 900-950ms erreichen kann. Bis 10kHz sinkt RT60 wieder auf unter 650ms.

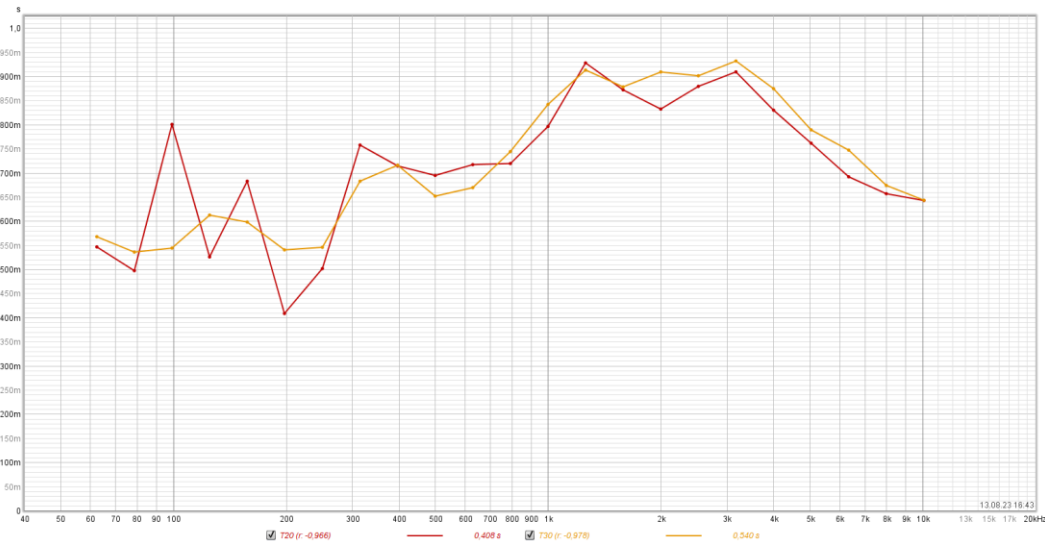


Abbildung 20: RT60 mit T20 in rot und T30 in orange.

Für diese Messung wurde das Messmikrofon Beyerdynamic MM1 genutzt.<sup>49</sup> Als Lautsprecher wurde die Full-Range PA Aktivbox RCF ART-310-A genutzt.<sup>50</sup>

#### 3.4.3 Raummessungen durch Audio Spatializer

Nachdem die Messung für den realen Raum erhoben wurde, kann ein Vergleich mit den digitalen Messungen angefertigt werden. Dazu wurde der virtuelle Raum in Unity genutzt, in dem eine Klangquelle an derselben Position platziert wurde, an dem der Lautsprecher im realen Raum positioniert war. Das virtuelle Mikrofon oder `AudioListener` wurde an die Stelle des echten Mikrofons fix platziert und blickt in Richtung der Klangquelle (Azimuth = 0 Grad zur Klangquelle). Anschließend wurde der Sweep von der Messsoftware an Unity gesendet, wo er als Ausgabe für die virtuelle Klangquelle diente. Das virtuelle Mikrofon wurde als Empfänger genutzt, um das System des digitalen Raumes in Echtzeit zu messen. Daraufhin wurde das RT-60-Diagramm generiert und mit dem des realen Raums verglichen. Durch diese Messungen ist ein objektiver Vergleich der Nachhallzeiten zwischen den virtuellen Räumen, aber auch dem echten Raum möglich. Sollte es möglich sein durch Anpassungen der Parameter der Spatializer eine Annäherung des RT-

<sup>49</sup> <https://global.beyerdynamic.com/mm-1.html>

<sup>50</sup> <https://www.rcf.it/de/products/product-detail/art-310-a-mkiii>

### 3 Testen der Spatializer

---

60 Diagramms zum realen Raum zu erzeugen, so wurde diese ebenfalls dokumentiert.

Für die Raummessungen über den Audio Spatializer wurden 4 Unity Szenen angelegt. Die SpatializingSoundTemplate Szene konnte kopiert werden, jedoch müssen manche Prefabs die für den Hörtest benötigt wurden angepasst oder gänzlich gelöscht werden. Die Applikation zur Messung wurde im Unity Editor ausgeführt.

Um den Sweep aus der Messsoftware REW in Unity auszugeben, wurde ein Skript für Unity geschrieben. Das Skript `MicInput` zeigt alle verfügbaren Mikrophone des Systems an und kann über die Wahl der Device Array ID und der Eingabe der Sampling Rate das Eingangssignal am gewählten Mikrophon als Ausgabe über eine `AudioSource` Komponente wiedergeben. Der `AudioListener` in der Szene wiederum gibt das Standardausgabegerät in Windows wieder. Diese Kanäle werden in REW als Eingangssignal eingespeist. Um das Routing zu verwirklichen, wurde auf die Software VoiceMeeter Banana<sup>51</sup> von VB-Audio zurückgegriffen.

---

<sup>51</sup> <https://vb-audio.com/VoiceMeeter/banana.htm>

### 3 Testen der Spatializer

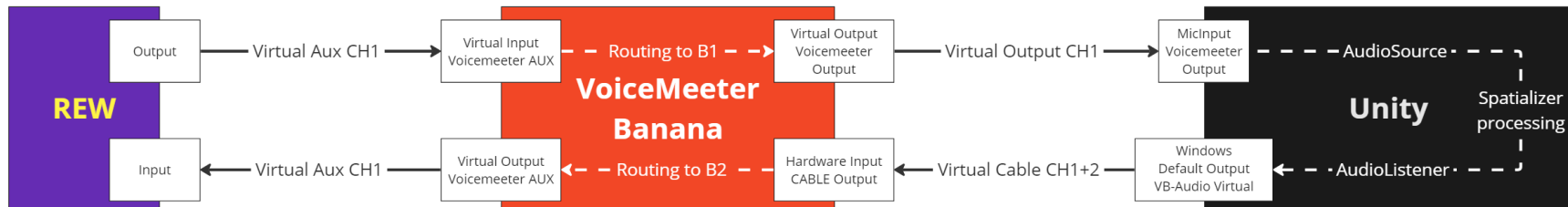


Abbildung 21: Routing des Audiosignals von REW nach Unity und wieder zurück mithilfe von VoiceMeeter Banana.

In Abbildung 21 ist zu sehen, wie das Audiosignal (Sweep) von REW über die virtuelle Routingsoftware VoiceMeeter Banana nach Unity und das durch die Audio Spatializer verarbeitete Signal wieder zurück in die Messsoftware gelangt. In REW unter den Einstellungen wurde dafür der Audiotreiber ASIO und unter ASIO Device das virtuelle VoiceMeeter AUX Virtual ASIO Device gewählt. Für den Output wurde der Kanal VM-VAIO-1 selektiert. Über diesen Kanal wurde der Sweep ausgespielt. Als Input wiederum wurde ebenfalls VM-VAIO-1 ausgewählt. Dieser Kanal empfing das prozessierte Signal und leitete es zur Messung weiter. Siehe hierzu Abbildung 22.

VoiceMeeter Banana empfing nun das Audiosignal aus REW über die Virtual Inputs unter VoiceMeeter AUX. Um dieses über den virtuellen Output VoiceMeeter Output weiterzuleiten, musste das Signal nach B1 geschickt werden.

### 3 Testen der Spatializer

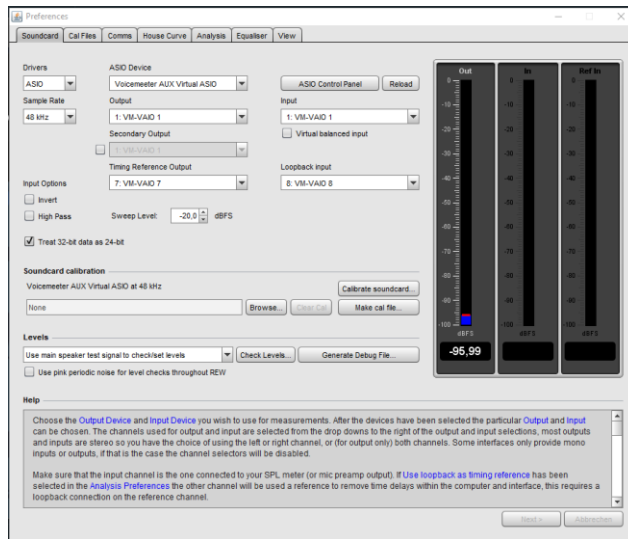


Abbildung 22: ASIO Routing in REW

In Unity nahm das `MicInput` Skript nun das Signal entgegen, indem der entsprechende Output als Mikrophon eingespeist wurde. Über die `AudioSource` wurde das Signal nun in der Unity Szene ausgegeben und vom verwendeten Audio Spatializer prozessiert.

Der `AudioListener` gab das verarbeitete Signal nun durch das in Windows selektierten Standard-Audioausgabegerät wieder aus. Daher wurde hier das VB-Audio Virtual Cable ausgewählt, um das Signal in VoiceMeeter Banana abgreifen zu können. Hier musste es nun noch über B2 auf den Virtual Output VoiceMeeter AUX geroutet werden, um am Ende der Kette wieder in REW verarbeitet werden zu können. Die Konfiguration des VoiceMeeter Banana ist in Abbildung 23 zu sehen.



Abbildung 23: VoiceMeeter Banana Konfiguration

### 3.4.4 Meta Spatializer

Der MetaXR Spatializer wurde in der Version 59.0 verwendet. Er ist zur Integration in Unity als Spatializer Plugin direkt verfügbar. Zusätzlich kann der Spatializer auch über die Audio Middleware FMOD verwendet werden.

Um den MetaXR Spatializer nun in Unity zu integrieren, muss das Unity Plugin<sup>52</sup> in Unity installiert werden. Danach muss das Spatializer Plugin unter dem Menüpunkt Edit > Project Settings und der Kategorie Audio ausgewählt werden. Weiters wurde „Best latency“ gewählt, um die Audiolatenz zu reduzieren. Das Ambisonics Decoder Plugin ist für die Applikation irrelevant. (*Meta XR Audio SDK Plugin for Unity - Requirements and Setup: Unity | Oculus Developers*, o. J.)

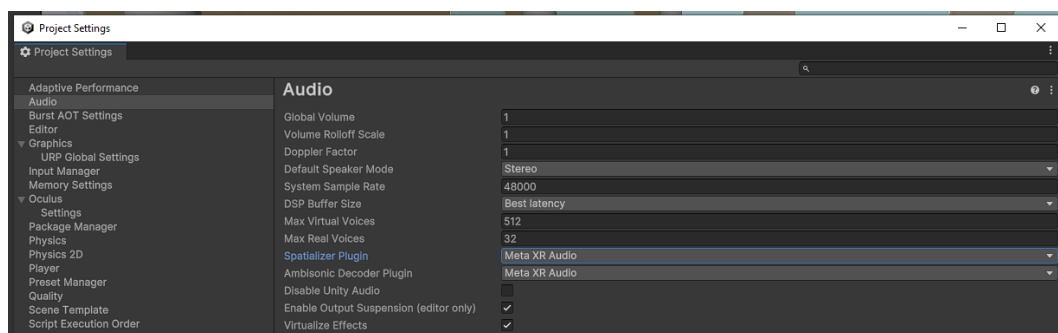


Abbildung 24: Meta XR Audio Spatializer Plugin Auswahl in den Unity Project Settings

Nun muss die `AudioSource` um die `MetaXRAudioSource` Komponente erweitert werden. Um mit dem Meta XR Spatializer zu funktionieren, muss das der „Spatialize“ Parameter aktiviert und der Parameter „Spatial Blend“ der `AudioSource` auf 1 gesetzt. Alle nativen Parameter der `AudioSource` können weiterverwendet werden.

Die `MetaXRAudioSource` Komponente erweitert die Gestaltungsmöglichkeiten. Zusätzlich wurde die `MetaXRAudioSourceExperimentalFeatures` Komponente hinzugefügt. Es lässt noch mehr Spielraum in der Gestaltung der `AudioSource`.

Die Parameter der `MetaXRAudioSource` blieben in ihren Standardwerten. `Enable Spatialization` erlaubt das Aktivieren und Deaktivieren der Spatialisierung dieser `AudioSource`. Der `Gain Boost Db` Parameter verstärkt die Lautstärke über den Unity Standardwert. `Enable Acoustics` ermöglicht das An- und Abschalten der

---

<sup>52</sup> <https://assetstore.unity.com/packages/tools/integration/meta-xr-audio-sdk-264557>

### 3 Testen der Spatializer

---

Raum Akustik für diese `AudioSource`. Mit `Reverb Send Db` kann der Raumhall extra verstärkt oder gedämpft werden, was zu einem stärker verhalten bzw. trockeneren Klang führt.

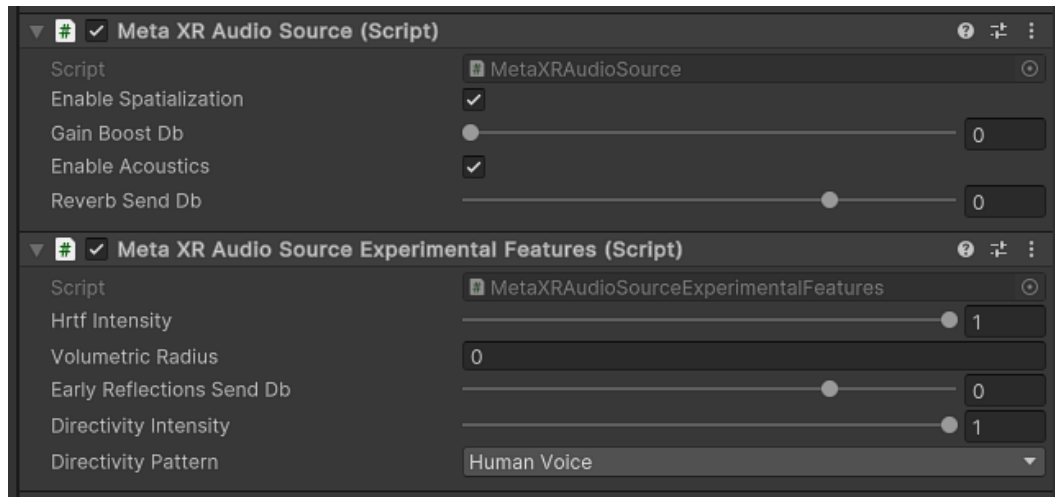


Abbildung 25: Parameter der `MetaXRAudioSource` und `MetaXRAudioSourceExperimentalFeatures`.

Wird dir `Hrtf intensity` auf 0 gesetzt, so werden vereinfachte HRTFs zur Faltung des Signals angewendet. Meta beschreibt ein „a stereo pan (with ITD still applied)“. Da dies jedoch ebenfalls zur Reduktion der Lokalisierbarkeit führt, wurde hier der Wert 1 gewählt. Der `Volumetric Radius` steuert den Eindruck der Größe des Klangobjekts. Da es sich bei dem Test um Punktschallquellen handelt, blieb dieser Parameter ebenfalls unberührt auf 0. `Early Reflections Send Db` steuert den Grad der Erstreflexionen die zugemischt werden. Der `Directivity Pattern` Parameter wurde auf `Human Voice` gestellt. Dadurch werden Filter so auf die `AudioSource` angewendet, dass eine gerichtete Klangquelle entsteht. Dieser Effekt kann mit der `Directivity Intensity` gesteuert werden und blieb somit auf 1. (*Apply Spatialization in Unity: Unity | Oculus Developers, o. J.*)

Um die Raumakustik über bestimmte Materialien noch besser realisieren zu können verfügt der MetaXR Spatializer über den Meta XR Audio Reflection Effekt. Dafür musste ein eigener Master Mixer für alle MetaXR Spatializer Szenen angelegt werden. Der Effekt wurde dann auf den Master Audio Mixer Group Controller angewendet. An allen `AudioSource`n der Szenen wurde dieser Master als Output selektiert.

In den Parametern dieses Effekts ist es möglich, die Erstreflexionen sowie den Nachhall ein und auszuschalten. Außerdem kann der gesamte Hallanteil in dB angepasst werden. Der Parameter `Voice Limiting` steuert die Anzahl an

### 3 Testen der Spatializer

---

spatialisierbaren `AudioSources`, die diesen Mixer als Output selektiert haben. Für die Applikation blieben diese Werte in Ausgangsstellung.

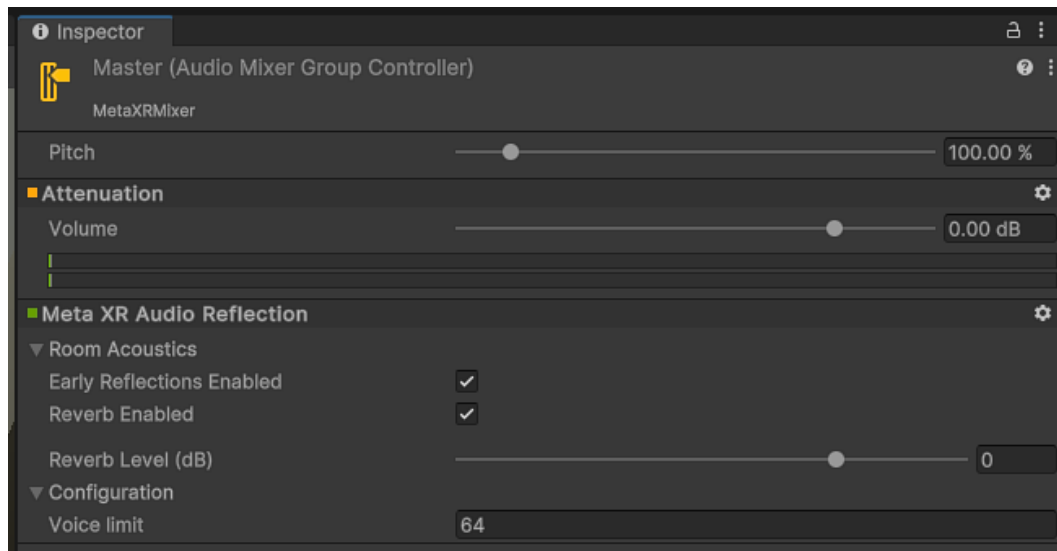


Abbildung 26: Einstellungen des Meta XR Audio Reflection Effekts

Um den Erstreflexionen und Nachhall im virtuellen Raum besser modellieren zu können bietet Meta die `MetaXRAudioRoomAcousticProperties` Komponente an. Diese muss einmalig in der Szene auf dem `RoomAcoustics` GameObject platziert werden. Alle `MetaXRAudioSources` der Szene, welche `Enable Acoustics` aktiviert haben, wurden daraufhin mithilfe dieser Komponente spatialisiert.

Für die Dimensionen des Raumes wurden die Abmessungen des echten Raumes verwendet. Mithilfe des `MetaXRRoomAcoustic` Gizmos kann die Position des erstellten akustischen Quaders über den `Motion Capturing Lab` gelegt werden. Dies geschah durch die Verschiebung des `RoomAcoustics` GameObject. Weiters konnten die Materialien gewählt werden, die im Raum ausgestattet sind. Da Meta keine Möglichkeit zur Einstellung der Absorptionsfaktoren der Materialien anbietet, ist man an die definierte Liste gebunden. Da die linke und vordere Wand (Blickrichtung aus dem Zentrum: vorne) zu großen Teilen aus Fenstern besteht wurde hier `Glass` als Material gewählt. Für die Decke `Acoustic Tile`, und die restlichen Flächen `Gypsum Board`. Dies wurde mit der offiziellen Dokumentation der Absorptionsgrade Metas abgeglichen. Der `Clutter Factor` Parameter erlaubt es die Fülle des Raumes durch andere Gegenstände oder Personen zu modellieren. Da im Raum ebenfalls Tische, Schränke und Säulen sowie andere Objekte vorhanden sind, wurde dieser Wert auf 0.5 belassen. (*Apply Room Acoustics in Unity: Unity | Oculus Developers, o. J.*)

### 3 Testen der Spatializer

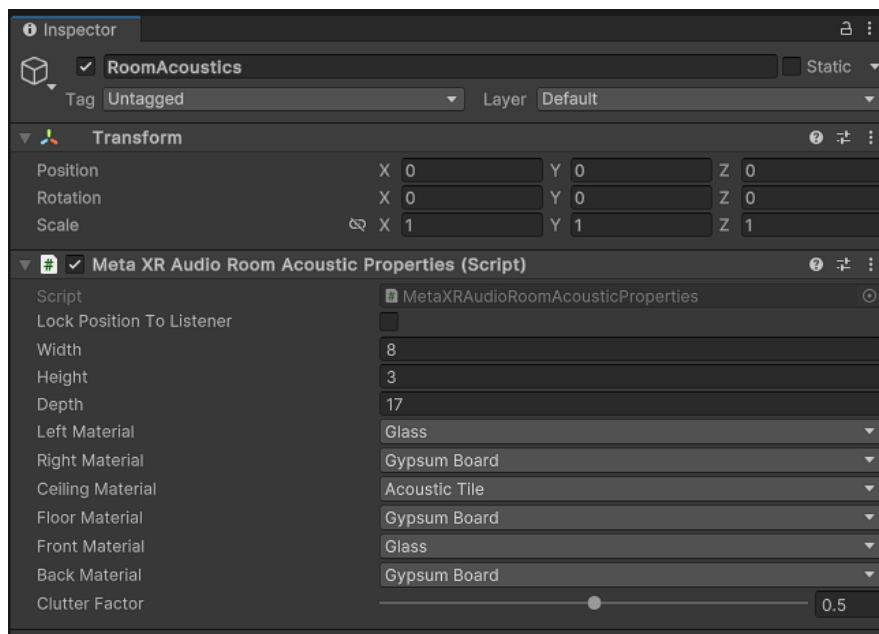


Abbildung 27: Parameter der `MetaXRAudioRoomAcousticProperties` Komponente auf dem `RoomAcoustics` GameObject.

In der Szene `MetaXRMeasurement` konnte nach diesem Setup nun die Messung mit REW erfolgen. Die Applikation wurde im Unity Editor gestartet und das virtuelle Mikrophon über die `MicInput` Komponente aktiviert. Nach dem Klick auf Start in REW wurde der Sweep durch den Spatializer auralisiert und kann in REW gemessen werden.

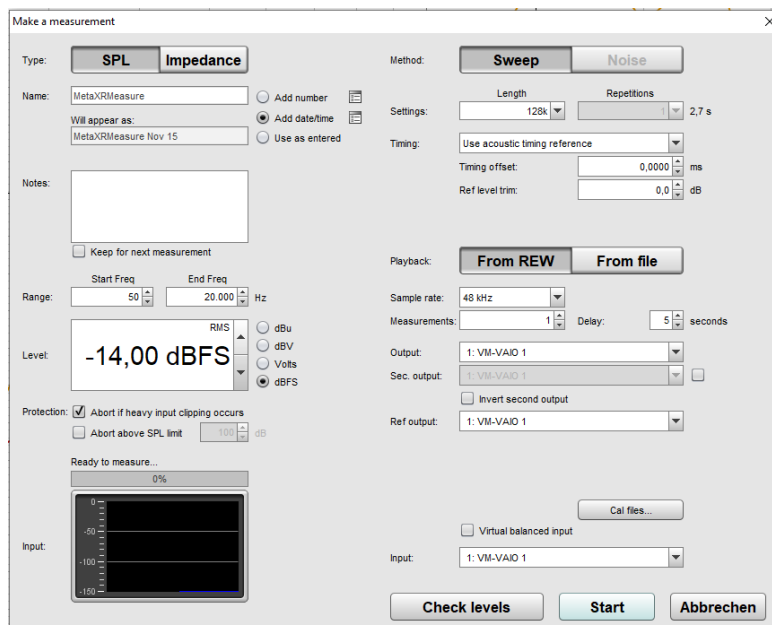


Abbildung 28: Messeinstellungen in REW für den MetaXR Spatializer

### 3 Testen der Spatializer

---



Abbildung 29: RT60 des MetaXR Spatializers im virtuellen Motion Capturing Lab.

Die Messung zeigt eine ähnliche virtuelle Hallentwicklung im Vergleich zum echten Raum besonders in den Mitten zwischen 400Hz und 3kHz.

#### 3.4.5 DearVR Spatializer

Der DearVR Spatializer für Unity in der Version 1.6 wurde für diese Masterarbeit von DearReality zur Verfügung gestellt. Es liegt als .unitypackage vor und kann so in Unity importiert werden.

Zunächst wurde eine neue Szene DearVRMeasurement erzeugt und die DearVR Engine in den Project Settings als Spatializer Plugin gewählt. DearVR benötigt den DearVRManager sowie eine DearVRSource in der Unity Szene, um eine AudioSource zu spatialisieren. Wie in der mitgelieferten Dokumentation beschrieben wird nun die AudioSource wie folgt erweitert:

Der Spatialize Parameter muss aktiviert und Spatial Blend auf 0 bzw. 2D eingestellt sein. Alle anderen Parameter bleiben bzw. haben keinen Effekt sobald DearVRSource als Komponente vorhanden ist. (Dear Reality, 2020, S. 24–25) Der Output der AudioSource muss ebenfalls auf eine neue DearVR Mixer Group gesetzt werden.

Nach hinzufügen der DearVRSource Komponente an der AudioSource, wurde der Internal Reverb deaktiviert. Alle DearVRSources in der Szene sollen über Reverb Sends an das DearVR Reverb Plugin gesendet werden. Dadurch wird die Berechnungslast wesentlich verringert. (Dear Reality, 2020, S. 18)

### 3 Testen der Spatializer

---

Die Levels Parameter blieben unberührt. Bass Boost, Unity Distance Graph (zum Deaktivieren der internen Distanzberechnung), Distance und Phi Angle Correction blieben ebenfalls auf den Standardwerten.

Der Auralization Parameter hingegen wurde aktiviert, um die Berechnung der Erstreflexionen in Echtzeit anzustoßen. Dies erfordert die Room Geometry for Auralization Einstellungen am DearVRManager. Der DearVRManager verfügt über einen Algorithmus zur Raumanalyse. Dieser erkennt automatisch den Abstand zu den Wänden des Raumes inkl. Decke und Boden. Dafür jedoch benötigt der Raum eine Begrenzung durch einen Unity Collider. Zusätzlich musste also ein Mesh aus den 6 Seiten des Raumes erzeugt werden. Nun ist es möglich über die Option Debug Room Analyzer (Gizmos), sowie den Werten in Manual Room Geometry das Ergebnis der Raumanalyse zu überprüfen. Diese lieferten den Abstand zwischen AudioListener und den Wänden bzw. der Decke/Boden des Motion Capturing Lab.

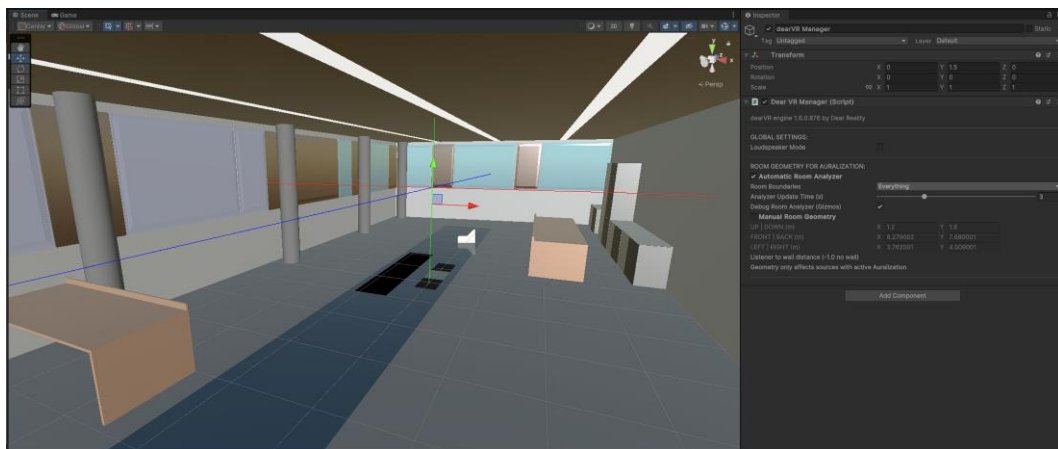


Abbildung 30: Automatische Raumanalyse mit dem DearVRManager.

Die Occlusion und Obstruction Parameter sind für die Applikation ohne Relevanz, da es keine Objekte zwischen AudioListener und AudioSource gibt die zu einer Beugung bzw. zum Durchdringen eines Objekts einer Klangquelle führen. Der DearVR Play Performance Mode blieb ebenfalls unberührt.

### 3 Testen der Spatializer

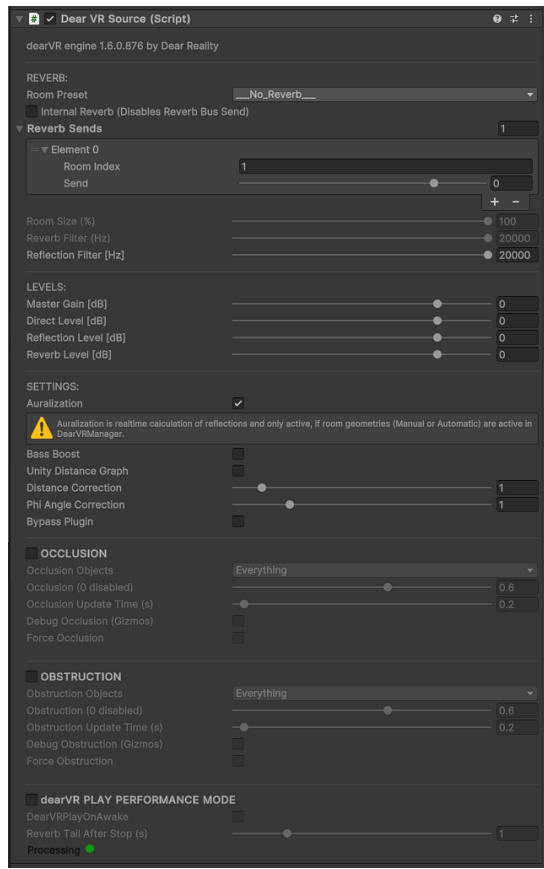


Abbildung 31: DearVRSource Parameter

Abschließend wurde im DearVR Reverb Plugin das Room Preset gewählt und die Reverb ID 1 gewählt. Diese wurde zuvor in der DearVRSource vergeben. Das Plugin wurde auf einen eigene Audio Mixer Group Controller namens „Reverb“ gelegt. Anschließend konnte in REW die Messung durchgeführt werden.

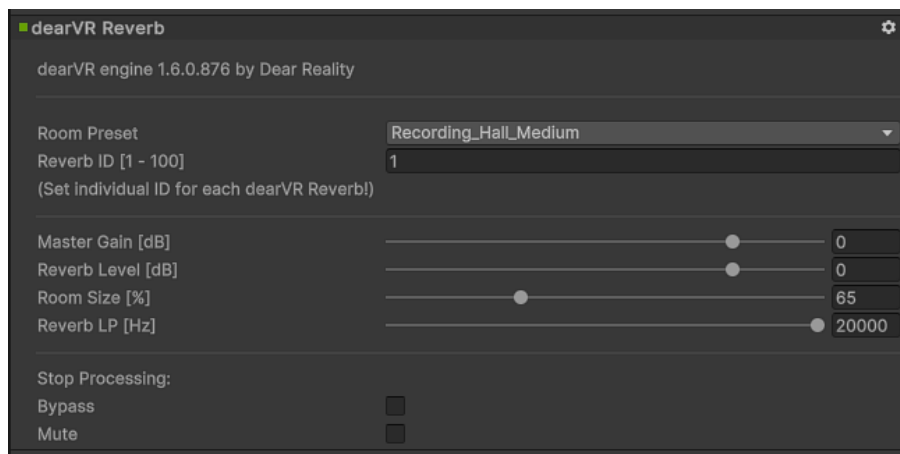


Abbildung 32: Das DearVR Reverb Plugin

### 3 Testen der Spatializer

Bei den Messungen stellte sich heraus, dass das verwendete Reverb Preset maßgeblich die RT60 Messungen beeinflusst. Eine Messung mit dem Preset Room\_Large führte zu folgender RT60:



Abbildung 33: RT60 des Reverbpreset „Room\_Large“ mit dem DearVR Spatializer.

Eine Annäherung an den echten Raum konnte mit dem Preset Recording\_Hall\_Medium und einer Roomsize von 65% erreicht werden:

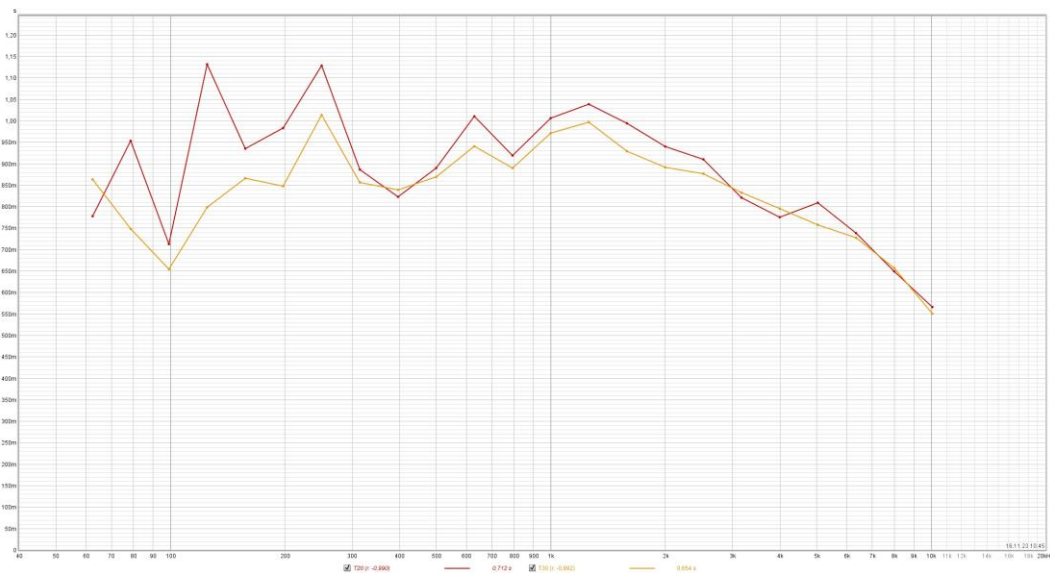


Abbildung 34: RT60 mit Reverbpreset „Recording\_Hall\_Medium“ mit dem DearVR Spatializer.

### 3.4.6 Steam Audio

Der Steam Audio Spatializer wurde in der Version 4.4.1 verwendet. Er ist zur Integration in Unity als Spatializer Plugin direkt verfügbar. Zusätzlich kann der Spatializer auch als Unreal Engine Plugin, über die Audio Middleware FMOD oder direkt mithilfe der C API verwendet werden.

Zunächst wurde die Unity Szene SteamAudioMeasurement angelegt. Der Spatializer wurde über das .unitypackage importiert. Der Ordner FMOD wurde dabei ausgenommen. Anschließend wurde der Steam Audio Spatializer in den Project Settings als Spatializer Plugin gewählt. (*Getting Started — Steam Audio Unity Integration documentation*, o. J.) An der AudioSource musste der Spatialize Parameter aktiviert werden. Der Spatial Blend Parameter wurde auf 1 (3D) gestellt. Alle anderen Parameter der AudioSource blieben unberührt.

Der AudioSource wurde die SteamAudioSource Komponente hinzugefügt. Distance Attenuation und Air Absorption wurde aktiviert, um den Lautstärkeabfall und Filterung über die Distanz vom Spatializer berechnen zu lassen. Directivity wurde ebenfalls aktiviert, um den Lautsprecher mit Nieren-Richtcharakter und nicht als Punktschallquelle darstellen zu können. Reflections wurden aktiviert. Mit dem Reflection Type Realtime werden während der Laufzeit Raycasts ausgesendet um so, auf Kosten der CPU, Reflexionen berechnen zu können. Apply HRTF To Reflections resultiert in eine bessere Lokalisierbarkeit, weshalb auch dieser Parameter aktiviert wurde. Pathing und Occlusion sind für die Applikation nicht relevant. (*Steam Audio Source — Steam Audio Unity Integration documentation*, o. J.)

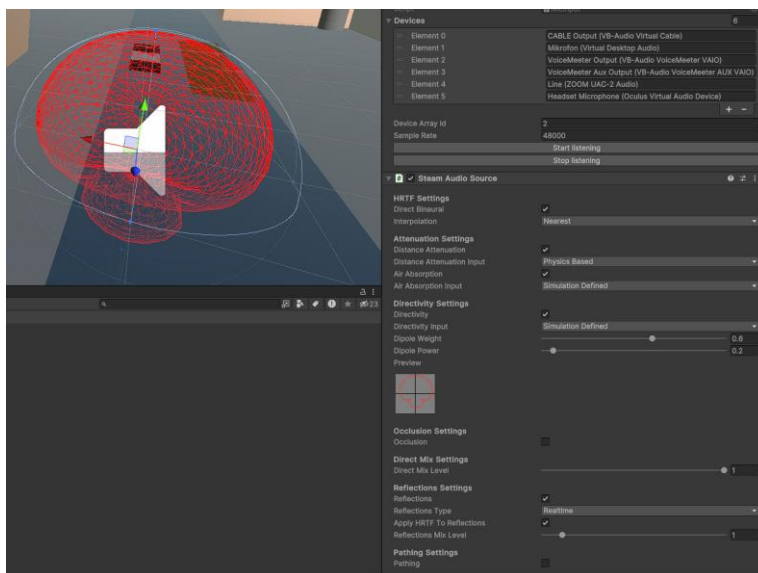


Abbildung 35: SteamAudioSource Parameter

### 3 Testen der Spatializer

---

Mithilfe der `SteamAudioGeometry` Komponente ist es möglich die Reflexionen durch den Spatializer berechnen zulassen. Dafür wurde eine Prefab Variante `SteamAudioRemoCapLab` aus dem Prefab `RemoCapLab Invisible` erzeugt. In diesem wurden nun alle Gegenstände inkl. Wände, Decke und Boden mit dem passenden Material versehen. In dem Asset Ordner `Plugins > SteamAudio > Ressources > Materials` sind alle vordefinierten Materialien und ihr Absorptionsverhalten definiert. Für die Wände und Decke wurde das Material `Default` gewählt, für Fenster das Material `Glass`, die Säulen und Boden das Material `Concrete`, die Tische und Schränke `Wood`, das Waschbecken `Ceramic`. (*Steam Audio Geometry — Steam Audio Unity Integration documentation, o. J.*)

Da sich die Position des gesamten virtuellen Motion Capturing Labor Prefab während der Laufzeit durch die neue Positionierung über die Synchronisation ändern kann, musste die `SteamAudioDynamicObject` Komponente dem hierarchisch am höchsten liegenden `GameObject` hinzugefügt werden. Die Export `Dynamic Object` Funktion durchsucht das `GameObject` nach `SteamAudioGeometry` Komponenten und erzeugt ein Asset für den Spatializer. Das Asset wurde im Ordner Prefab abgelegt. (*Steam Audio Dynamic Object — Steam Audio Unity Integration documentation, o. J.*)

Um den “physics-based” Hall berechnen zu lassen musste eine neue Mixer Gruppe angelegt werden. Diesem wurden eine Gruppe `Reverb` angelegt, um das `Steam Audio Reverb` Plugin darin hinzuzufügen. Der `Apply HRTF` Parameter wurde für eine bessere Spatialisierung aktiviert. Zudem musste dem `AudioListener` ein `SteamAudioListener` hinzugefügt werden und `Apply Reverb` aktiviert sowie der `Reverb Type` auf `Realtime` gestellt werden. (*User’s Guide — Steam Audio Unity Integration documentation, o. J.*)

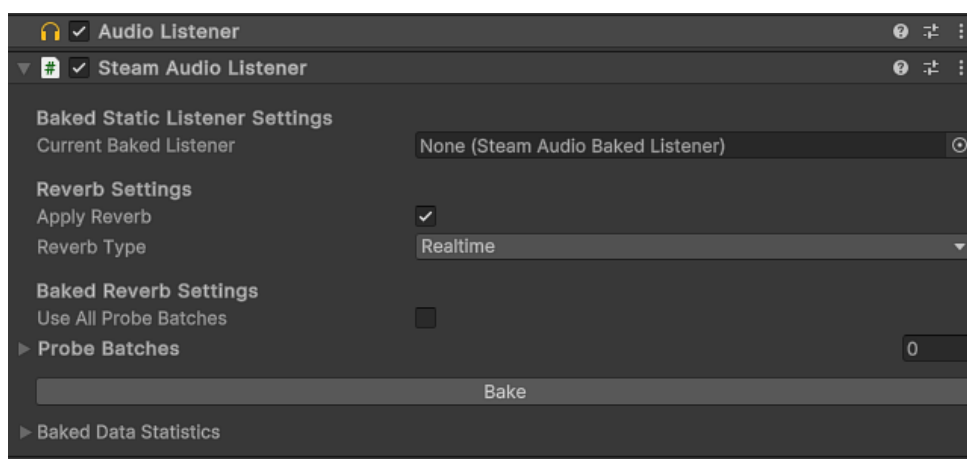


Abbildung 36: `SteamAudioListener` zur Aktivierung des Reverbs mit dem `Steam Audio Spatializer`.

### 3 Testen der Spatializer

---

Die Messungen mit REW zeigen folgende RT60:

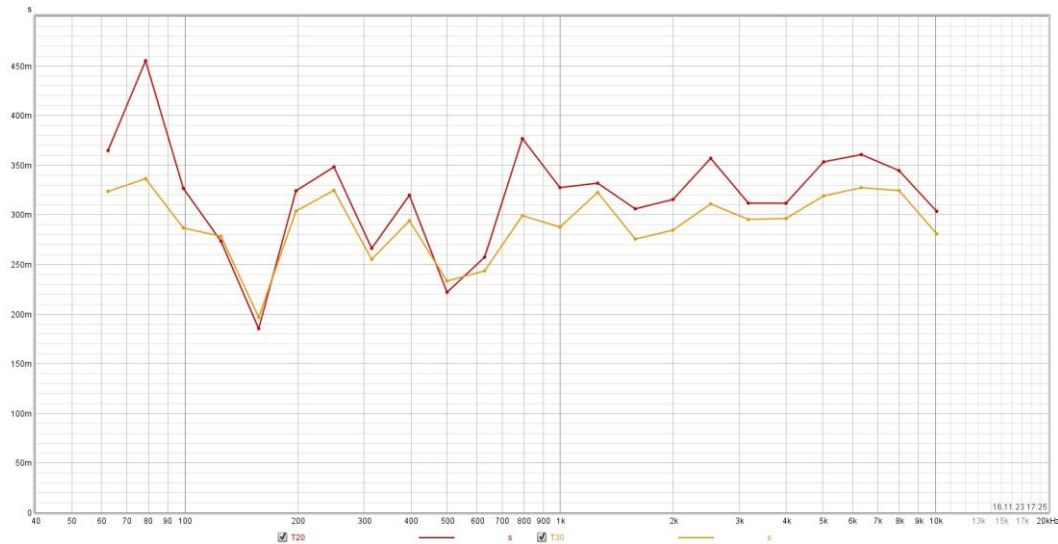


Abbildung 37: RT60 mit dem Steam Audio Spatializer

Es zeigt einen hallärmeren Raum. Steam Audio lässt keine Beeinflussung des Reverbs zu, da er in Echtzeit prozessiert wird. Lediglich die Ratio des Direktschalls zum Nachhall kann über die Lautstärke der Mixer Gruppen beeinflusst werden. Dies führt jedoch zu keiner Änderung an der RT60.

#### 3.4.7 Microsoft Acoustics

Das Microsoft Acoustics Spatializer Plugin liegt in der Version 2022.1 vor.<sup>53</sup> Es kann neben Unity auch in der Unreal Engine oder als Wwise Plugin verwendet werden.

Zunächst wurde das .unitypackage in Unity importiert. Eine neue Szene MSAcousticMeasurement wurde angelegt. In den Project Settings musste Microsoft Acoustics als Spatializer Plugin gewählt werden. Daraufhin wurde eine neue Mixer Gruppe angelegt und das Project Acoustics Mixer Plugin darauf angewendet.

In der AudioSource wurde Spatialize aktiviert und der Spatial Blend auf 3D gesetzt. Nun musste noch die AcousticsAdjust Komponente der AudioSource hinzugefügt werden. (kevinasg, 2022b)

---

<sup>53</sup> <https://www.microsoft.com/en-us/download/details.aspx?id=57346>

### 3 Testen der Spatializer

---

Der Workflow in Microsoft Acoustics wird auf in drei Schritten beschrieben:

Vorab-Berechnung (Pre-bake): Zunächst wird jede Geometrie ausgewählt, die akustisch relevant ist. Danach wird ihr ein Material zugewiesen. Zusätzlich wird eine "navigation area" ausgewählt, auf der sich der `AudioListener` in der Szene bewegen kann. Debug-Anzeigen helfen dabei zu zeigen, welche Daten vor der Durchführung der Berechnung im Berechnungsprozess (Baking) einbezogen werden.

Berechnung (Bake): Die akustische Simulation erfordert intensive Rechenleistung. Die von der Vorab-Berechnung erzeugten Dateien werden an Azure gesendet, um den Simulationsprozess zu beschleunigen. Nach Abschluss wird die resultierende akustische Information als Unity Asset in das Projekt integriert.

Laufzeit: Während der Laufzeit kann im Editor die Akustik pro Klangquelle kontrolliert und verändert werden. (kevinasg, 2023)

Das „Baking“ in der Szene wurde über das Acoustics Window (Window > Acoustics) in Unity vorberechnet. Dafür wurde zunächst ein neue Prefab Variante aus dem RemoCapLab Invisible Prefab erstellt. Im Reiter Objects wurden nun alle relevanten Geometrien ausgewählt. Im Hierarchy Window von Unity wurde das `MSAcousticRemoCapLab` ausgewählt. Durch einen Klick auf Acoustics Geometry wurden allen `GameObjects` eine `AcousticsGeometry` Komponente hinzugefügt. Auf den Boden und Gang+Matte im Prefab wurde Acoustics Navigation angewendet. Für die Wände, Säulen, Boden, Fenster, Tische und Schränke im Raum wurden eigene unsichtbare Materialien angelegt. Diese wurden den Gegenständen des `MSAcousticRemoCapLab` entsprechend zugewiesen. Nun konnten die Materialien im Acoustics Window > Materials entsprechenden Absorptionsgraden zugewiesen werden.

### 3 Testen der Spatializer

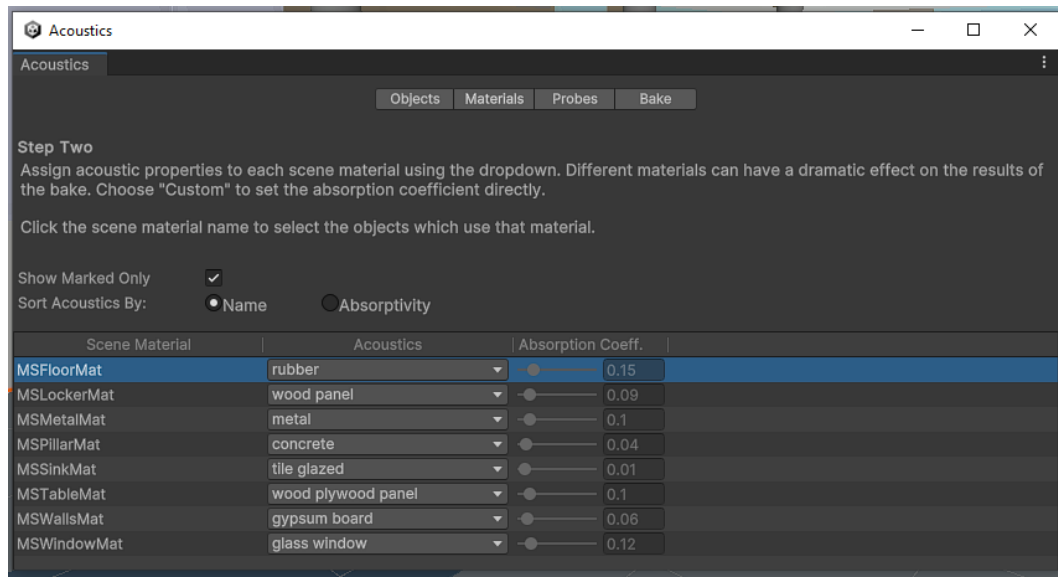


Abbildung 38: Acoustics Material Tab mit zugewiesenen Absorptionsgraden pro Material in der Szene.

Danach konnte die Vorberechnung der „probe points“ durchgeführt werden. Die Simulation Resolution Fine wurde gewählt, um die Frequenz der durchgeführten Simulationen zu verdoppeln. Dies erhöht jedoch auch die Zeit der Berechnung (Bake). Mithilfe der Gizmos AcousticsProbes und AcousticsVoxels ist nun die Vorberechnung überprüfbar. Die berechneten Daten wurden in den Assets unter AcousticsData abgespeichert.

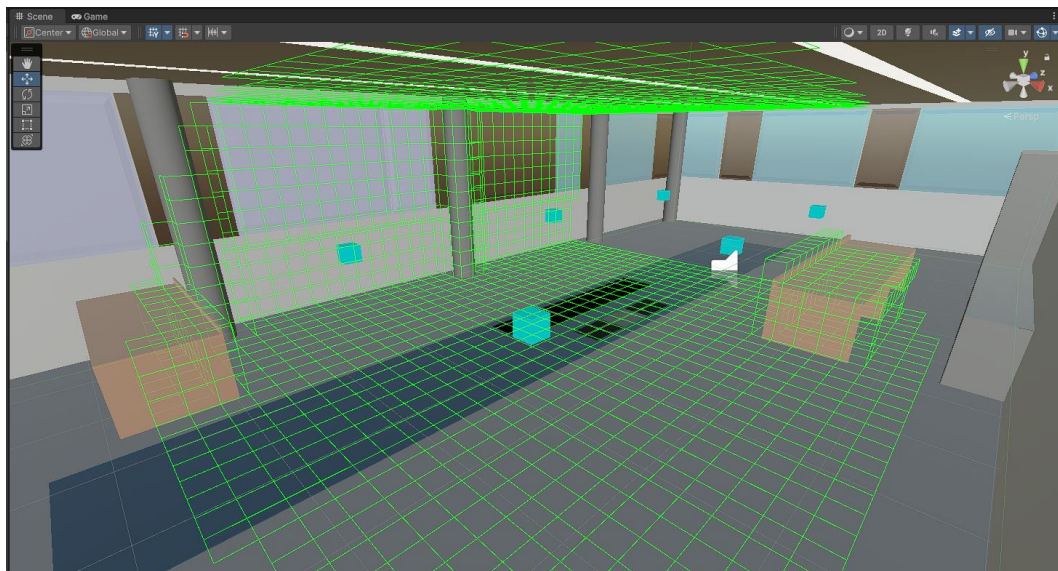


Abbildung 39: Berechnete Probe Points (blaue Würfel) und Voxels (grüne Linien) innerhalb der Unity Szene.

### 3 Testen der Spatializer

Diese können nun im letzten Schritt, dem Baking, in die Azure Cloud hochgeladen und die Berechnung durchgeführt. Dazu muss ein Azure Konto angelegt und entsprechend vorkonfiguriert werden.<sup>54</sup> Alternativ steht auch eine Offlineberechnung über einen Docker Container zur Verfügung.<sup>55</sup> Für das Projekt wurde die Szene offline berechnet. Das MicrosoftAcoustics Prefab wurde nun der Szene hinzugefügt und die berechnete Ace Datei mit der Endung .bytes dem AcousticsManager übergeben. (kevinasg, 2022a)

Nun kann der Raum mit REW getestet werden. Die gemessenen RT60 Werte liegen, bis auf den Bassbereich zwischen 60Hz-100Hz, in einem vergleichbaren Bereich zum realen Raum.

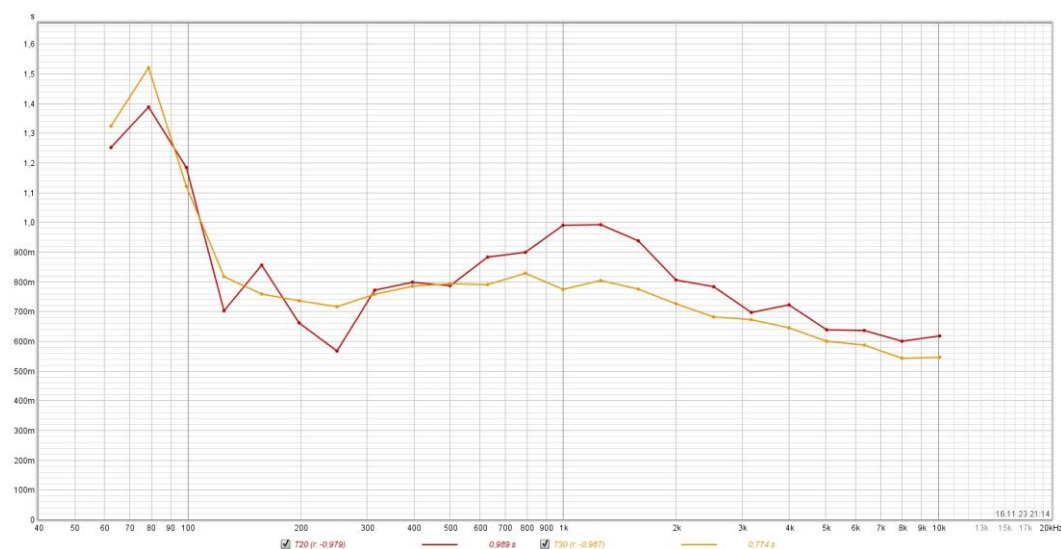


Abbildung 40: RT60 des virtuellen Motion Capturing Labs mit dem Microsoft Acoustics Spatializer.

#### 3.4.8 Build

Für die vier Hörtest-Applikationen wurden nun vier Szenen aus der Vorlage SpatializingSoundTemplate angelegt. Jede der Szenen integrierte einen Audio Spatializer wie in den Kapiteln oben beschrieben. Dafür wurden weitere Prefab Varianten der Prefabs AudioSource und SoundSources erstellt.

Nachdem jede der vier Hörtest-Szenen im Unity Editor über Quest Link auf der Meta Quest 3 getestet wurden, kann der Build für jede Szene einzeln angestoßen werden. Da es nicht möglich ist, das Spatializer Plugin während der Laufzeit zu

<sup>54</sup> <https://learn.microsoft.com/en-us/gaming/acoustics/create-azure-account>

<sup>55</sup> <https://learn.microsoft.com/en-us/gaming/acoustics/unity-baking-bake#Local-bake>

ändern, handelt es sich also um vier Applikationen, die auf der Meta Quest 3 vorinstalliert sind. Dieses Testgerät konnte nun für den Hörtest verwendet werden. Die Applikationen wurden in den Projektordner Builds hinterlegt.

## 3.5 Online-Fragebogen

Zusätzlich zu den in den Applikationen erhobenen Daten, sollten die Testperson mithilfe einer Online-Fragebogens durch das Testprozedere geleitet werden. Der pseudonymisierte Fragebogen erhebt demographische Daten. Indem jeder Testperson eine eindeutige Teilnehmeridentifikation zugewiesen wurde, konnten die Daten der Hörtestapplikationen und des Fragebogens miteinander verknüpft werden. Der Online-Fragebogen wurde mit dem Online-Tool Google Forms erstellt und der Testperson vor Ort auf einem PC über einen Browser angezeigt. Zur Erstellung eines Datensatzes ist keine Anmeldung von Nöten.

### 3.5.1 Aufbau

Der Online-Fragebogen besteht aus insgesamt 11 Seiten, wobei die erste Seite der Einführung dient. Der gezeigte Text soll die Testperson auf das nun kommende vorbereiten und Dank zur Teilnahme aussprechen.

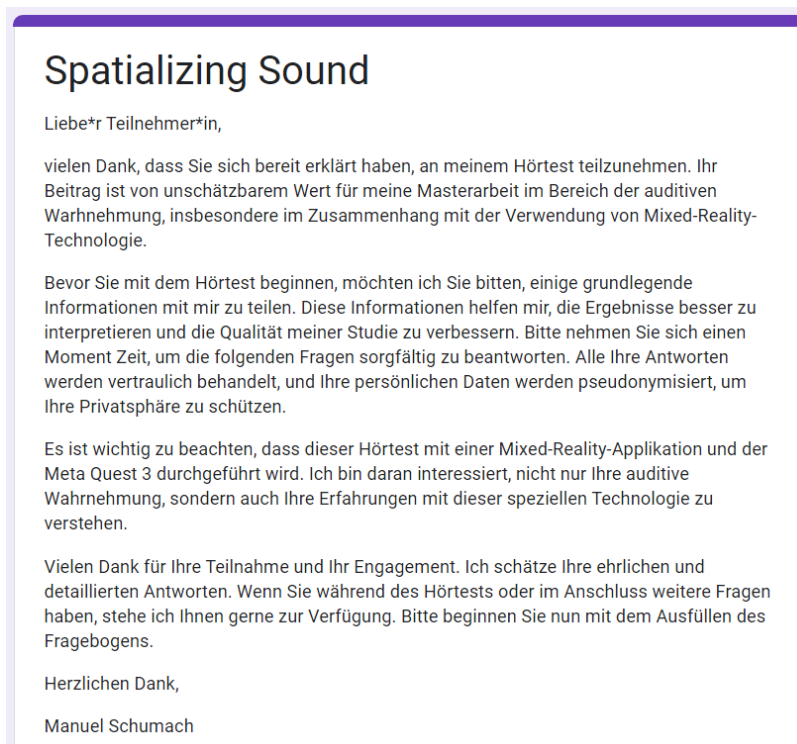


Abbildung 41: Einführung des Online-Fragebogen

### 3 Testen der Spatializer

---

Es folgt eine Abfrage einiger demographischen Daten zur Person, sowie die Eingabe der von der Versuchsleitung vergebenen Teilnehmeridentifikation. Neben Fragen zu Alter, Geschlecht und Beruf/Ausbildung wird auch die Erfahrung mit Virtual bzw. Mixed Reality Anwendungen abgefragt. Sollte die Testperson außerdem unter bekannten Hörschwierigkeiten leiden, so führt der Datensatz zu einem Ausschluss.

Es folgt eine kurze Einweisung auf den ersten Hörtest mit der ersten Mixed Reality Applikation. Die Testperson wird daraufhin von der Versuchsleitung angewiesen zur Mitte des Raumes zu kommen und die Meta Quest 3 sowie die Kopfhörer aufzusetzen. Es folgt der erste Durchlauf der Applikation, sowie der einmalige Warm-Up, um die Interaktion mit der Applikation zu verstehen.

Wurde der Hörtest abgeschlossen und die Daten auf der Meta Quest 3 gespeichert, wird sie mit Hilfe der Versuchsleitung wieder abgesetzt und die Testperson kann mit der Beantwortung des Fragebogens fortsetzen. Es folgen einige Fragen zu diesem ersten Hörtest, welche sich allerdings von den folgenden 3 Hörtests nicht unterscheiden.

The image shows a questionnaire titled "Hörtest 1 Fragen" with a purple header. Below the header, there is a white box with the instruction: "Beantworten Sie nun folgende Fragen zum eben durchgeführten ersten Hörtest." The questionnaire consists of five sections, each with a question and a 5-point Likert scale:

- Section 1: "Wie genau konnten Sie die Stimme im Raum lokalisieren? \*" with options from "Sehr ungenau" to "Genau".
- Section 2: "Inwiefern empfanden Sie die Wahrnehmung der Stimme als natürlich im Vergleich \* zur realen Stimme?" with options from "Überhaupt nicht natürlich" to "Sehr natürlich".
- Section 3: "Wie würden Sie den Schwierigkeitsgrad bewerten, die Stimme im Raum zu lokalisieren? \*" with options from "Sehr schwierig" to "Sehr einfach".
- Section 4: "Ist Ihnen während der Durchführung des Tests etwas Besonderes aufgefallen? (Zu laut/leise, Störgeräusche,...)" with a text input field labeled "Your answer".

Abbildung 42: Fragebogen zu der Hörtestapplikation

Dies wiederholt sich nun für die restlichen 3 Applikationen bis schlussendlich die abschließenden Fragen zum Hörtest gestellt werden. Sie fragen die Testpersonen nach etwaigen Herausforderungen während des Tests, allgemeinen Eindrücken und Verbesserungspotenzial.

## 3.6 Hörtest

Der erste Hörtest fand am 25.11.2023 an der FH St.Pölten im Motion Capturing Labor statt. Zwei Personen wurden im Abstand von 45 Minuten ab 11:00 gebeten, an der FH zu erscheinen. Vorbereitend wurden die Spatial Anchor der Testapplikationen für den Raum gespeichert. Dies gewährleistet eine konstante Positionierung des virtuellen Raums über alle Testteilnehmer hinweg.

Die Reihenfolge der getesteten Applikationen pro Person ist:

- MS Acoustics
- Meta Spatializer
- Steam Spatializer
- DearVR Spatializer

Die Testperson begann mit der Eingabe der Teilnehmeridentifikation im Online-Fragebogen. Demographische Daten wurden zusätzlich erhoben. Die Testperson wurde gebeten mit der ersten Applikation den Hörtest zu beginnen. Diese wurde zuvor von der Testleitung gestartet und auf korrekte Positionierung überprüft.

Zu Beginn des ersten Tests mit dem MS Acoustics Spatializer wurden dem:der Testteilnehmer:in grundsätzliche Interaktionsmöglichkeiten mit der Meta Quest 3 anhand des Quest Menus beigebracht. Um nicht nachvollziehbare Daten während des Hörtests zu vermeiden, wurde die Testperson gebeten in das VR-Menü der Meta Quest 3 zu wechseln. Die Passthrough Funktion war nun deaktiviert. Die Testteilnehmer:innen wurden gebeten auf die Testleitung zu zeigen. Diese hat sich in 3 Positionen um die Testperson im Raum bewegt und begann zu sprechen. Zeigte die Testperson auf die Testleitung, so konnte man von einer rationalen Lokalisierfähigkeit ausgehen.

Nun wurde die erste Applikation von der Testleitung gestartet und nach der Kontrolle zur Positionierung des virtuellen Raumes wurde das Headset an die Testperson übergeben. Zusätzlich wurde ihr die Kopfhörer für den Test aufgesetzt. Da nicht alle der Testteilnehmer:innen mit der Steuerung der Controller und der

### 3 Testen der Spatializer

---

Testapplikation vertraut waren, wurde zunächst das Warmup in die Applikation ausgeführt.

Nachdem die Daten zur ersten Applikation gesammelt wurden, übergab der:die Testteilnehmer:in das Headset inklusive Kopfhörer an die Testleitung. Es folgte die Befragung zum ersten Hörtest im Online-Fragebogen. Danach wiederholte sich der Hörtest für die restlichen 3 Applikationen. Von hier an konnte auf das Warmup verzichtet werden.



*Abbildung 43: Hörtest mit vorheriger Lokalisierungsprüfung*

Die ermittelten Daten wurden lokal auf der Meta Quest 3 gespeichert. Nachdem alle vier Applikationen getestet und die Beantwortung des Fragebogens abgeschlossen wurde, können die Daten der:die Testteilnehmer:in mithilfe der Teilnehmeridentifikation, der Uhrzeit und der Reihenfolge der Applikationen korrekt zugewiesen werden.

Es folgt eine Auflistung aller Hörtesttermine. Zu den entsprechenden Uhrzeiten begannen die Tests für die teilnehmenden Testpersonen mit der TeilnehmerID. Anhand dieser Liste ist eine spätere Zuweisung der Testdaten zur Testperson möglich.

Datum	Uhrzeit	TeilnehmerID
25.11.2023	11:00	1
	11:45	2
03.12.2023	12:30	4

### 3 Testen der Spatializer

---

	13:15	5
	14:00	7
	14:45	6
	15:45	8
	16:30	9
	17:15	10
	18:00	11
09.12.2023	11:30	18
	12:15	12
	13:00	13
	13:45	14
	14:30	16
	15:15	15
	16:00	17

*Tabelle 1: Zeitpunkte und zugehörige TeilnehmerIDs der Hörtests*

Weiters wurde die Applikation nach dem ersten Hörtestdatum geringfügig angepasst. Die Synchronisation des Raumes durch die Spatial Anchors wird nun automatisch beim Aufsetzen der Meta Quest 3 gestartet. Bis dahin musste dies manuell durch Drücken des Universal Menu-Buttons erfolgen. Außerdem wurde der virtuelle Kreis am Boden des Raumes, um eine Anzeige erweitert, die nun die Sicht „nach vorne“ bestätigt. Dadurch konnte der:die Testteilnehmer:in nach dem Aufsetzen des Headsets gefragt werden, ob die Positionierung des virtuellen Raumes nach dem Synchronisieren korrekt ist. Daraufhin konnte der Hörtest in einem korrekt positionierten und rotierten Raum durchgeführt werden.



*Abbildung 44: Testteilnehmer bei der Ausführung des Hörtest mit einer der Hörtest Applikationen*

## 4 Evaluation

Um die vorhandenen Daten nun für ein Statistikprogramm (SPSS) vorzubereiten, wurde in Unity ein weiteres Skript geschrieben. Darin wurden alle von den Testpersonen generierten Daten gesammelt und in eine für SPSS verwertbare .csv Datei transformiert. Im nachfolgenden werden die zur Beantwortung der Fragestellungen dieser Masterarbeit notwendigen Variablen definiert und aus dem vorhandenen Datensätzen extrahiert.

Für beide Fragestellungen dieser Masterarbeit sind die Werte der Distanzdifferenz und der Winkelabweichung von größter Bedeutung. Sie werden für die Forschungsfrage nach der Präzision einer Klangquelle und dem Einreihen der Spatializer nach diesem Kriterium benötigt, aber auch für die Frage nach der Eignung, indem der Hörtest innerhalb der Toleranzbereiche Werte liefert. (Siehe hierzu Kapitel 2.1.6 Toleranzbereiche)

Für die Varianzanalyse der abhängigen Variablen „Distanzdifferenz“ und „Azimuthdifferenz“ werden die Positionen jedes gesetzten Markers innerhalb eines Spatializers (statistische Gruppe oder Kategorie) aus den vorhandenen Datensätzen benötigt. Diese sollen wiederum mit den korrekten Positionen (Ursprung) der virtuellen Klangquelle verglichen werden.

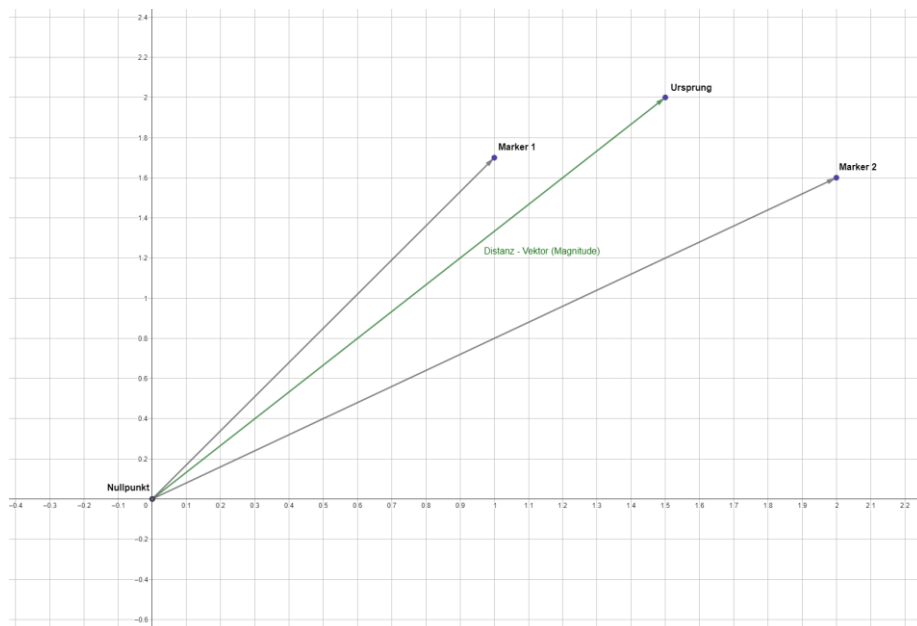


Abbildung 45: Möglicher Nullpunkt, Ursprung und zwei Marker in einem Koordinatensystem

## 4 Evaluation

---

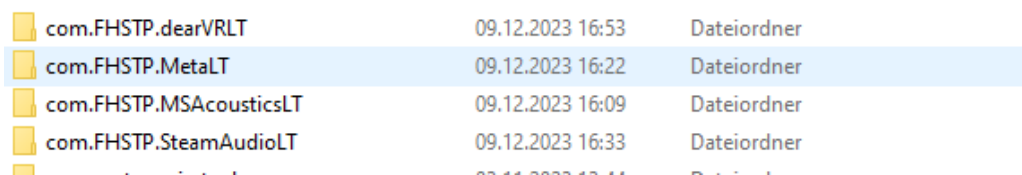
Um die Distanzabweichung zu berechnen, muss der Vektor des Ursprungs, von dem des Markers subtrahiert werden. Daraus ergibt sich eine Länge, welche positiv ist, wenn die Testperson die Klangquelle zu weit entfernt verortet hat. Ist die Länge negativ, wurde die Klangquelle zu nah wahrgenommen.

Für die Winkelabweichung muss das Skalarprodukt der beiden Vektoren, Ursprung und Marker, gebildet werden. Es ergibt den Winkel zwischen den Vektoren. Ist der Winkel positiv so wurde die Klangquelle weiter rechts verortet. Ist der Winkel negativ, so wurde die Klangquelle links verortet.

In Unity wurden nun diese Daten so transformiert, dass sie für die Statistiksoftware SPSS verwertbar sind. Dies geschah durch eine automatisierte Umwandlung in Unity in eine .csv Datei. Daraufhin kann in SPSS eine einfaktorielle Varianzanalyse berechnet und entsprechende Grafiken (Plots) erzeugt werden.

### 4.1.1 Datenaufbereitung in Unity

Alle .json Dateien wurden nach dem Hörtest auf der Meta Quest 3 in vier Ordner der Applikationen abgelegt. Diese vier Ordner finden sich im Internal Shared Storage unter Android/Data.



com.FHSTP.dearVRLT	09.12.2023 16:53	Dateiordner
com.FHSTP.MetaLT	09.12.2023 16:22	Dateiordner
com.FHSTP.MSAcousticsLT	09.12.2023 16:09	Dateiordner
com.FHSTP.SteamAudioLT	09.12.2023 16:33	Dateiordner

Abbildung 46: Ordner zu den abgelegten .json Dateien des Hörtests

Der Pfad zu den .json Dateien für die DearVR Applikation lag nun zum Beispiel unter „Android\data\com.FHSTP.dearVRLT\files\dearVRLT“.

Alle Dateien der vier Applikationen wurden in einen lokalen Ordner mit folgender Hierarchie abgelegt:

```
1_MSAcoustic
2_MetaLT
3_SteamAudioLT
4_dearVRLT
output
SpssData
```

## 4 Evaluation

---

Zusätzlich wurde allen .json Dateien händisch ein Prefix angefügt, welche die Teilnehmeridentifikation darstellt. Die korrekte Zuweisung wird durch den Zeitpunkt der abgegebenen Fragebögen bestimmt.

Name	Änderungsdatum	Typ	Größe
1_MSAcoustic_2023_11_25_11_22_33.json	25.11.2023 11:26	JSON-Quelldatei	5 KB
2_MSAcoustic_2023_11_25_12_01_54.json	25.11.2023 12:07	JSON-Quelldatei	5 KB
4_MSAcoustic_2023_12_03_12_42_53.json	03.12.2023 12:46	JSON-Quelldatei	5 KB
5_MSAcoustic_2023_12_03_13_36_37.json	03.12.2023 13:40	JSON-Quelldatei	5 KB
7_MSAcoustic_2023_12_03_15_19_08.json	03.12.2023 15:22	JSON-Quelldatei	5 KB
6_MSAcoustic_2023_12_03_15_42_51.json	03.12.2023 15:46	JSON-Quelldatei	5 KB
8_MSAcoustic_2023_12_03_16_25_31.json	03.12.2023 16:28	JSON-Quelldatei	5 KB
9_MSAcoustic_2023_12_03_16_58_17.json	03.12.2023 17:02	JSON-Quelldatei	5 KB
10_MSAcoustic_2023_12_03_17_39_21.json	03.12.2023 17:44	JSON-Quelldatei	5 KB
11_MSAcoustic_2023_12_03_18_28_01.json	03.12.2023 18:31	JSON-Quelldatei	5 KB
18_MSAcoustic_2023_12_09_12_03_23.json	09.12.2023 12:09	JSON-Quelldatei	5 KB
12_MSAcoustic_2023_12_09_12_50_46.json	09.12.2023 12:54	JSON-Quelldatei	5 KB
13_MSAcoustic_2023_12_09_13_24_57.json	09.12.2023 13:29	JSON-Quelldatei	5 KB
14_MSAcoustic_2023_12_09_14_06_27.json	09.12.2023 14:13	JSON-Quelldatei	5 KB
16_MSAcoustic_2023_12_09_15_01_20.json	09.12.2023 15:05	JSON-Quelldatei	5 KB
15_MSAcoustic_2023_12_09_15_30_07.json	09.12.2023 15:33	JSON-Quelldatei	5 KB
17_MSAcoustic_2023_12_09_16_14_58.json	09.12.2023 16:21	JSON-Quelldatei	5 KB

Abbildung 47: Hörtestergebnisse mit Teilnehmeridentifikation für die MSAcoustic Applikation

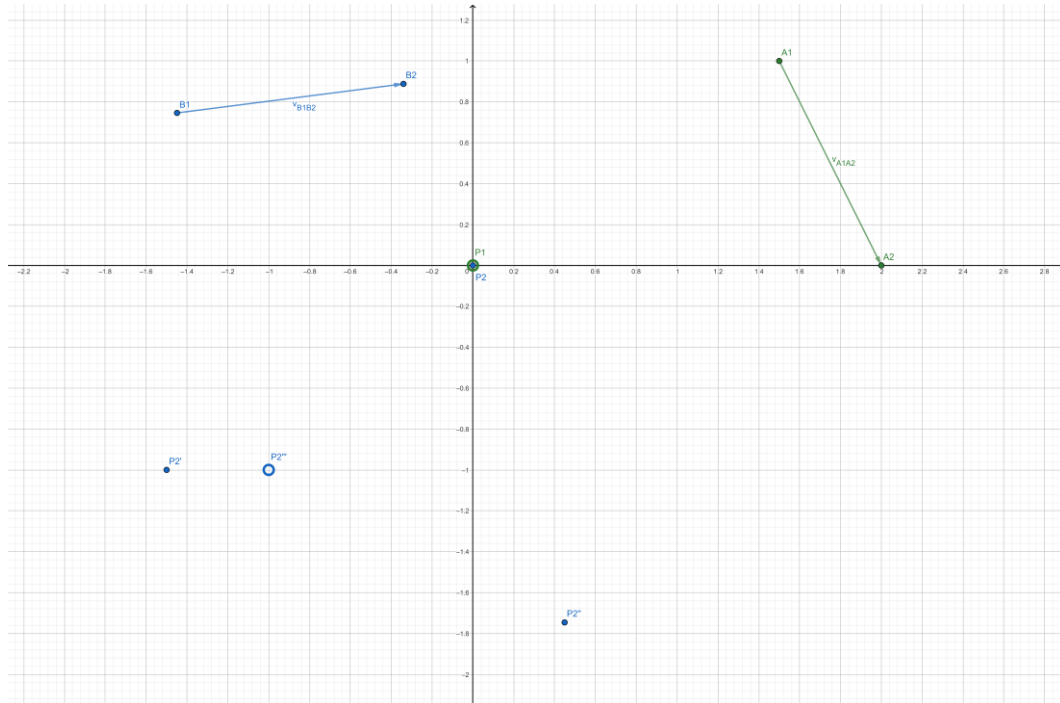
Die Daten liegen als .json serialisiertes `ListeningTestData` Klasse vor. (Siehe 3.3.7.8 `ListeningTest`) Dieses beschreibt pro Klangquelle einen Punkt im `Vektor3` Datentyp als platzierten Marker und sein Ursprungspendant. `Vektor3` beschreibt einen Punkt im dreidimensionalen Raum durch die mit dem `float` Datentyp gespeicherten Komponenten X, Y und Z. Unity bemisst mit Y die Höhe einer „Transform“ Komponente. Die Höhenwerte sind für die Fragestellungen jedoch nicht von Interesse. Pro Sample (Klangquelle), Testperson und Spatializer liegen also  $21 \cdot 17 \cdot 4 = 1428$  Datensätze an platzierten Markern (`Vektor3`) und ihren Ursprüngen (`Vektor3`) vor.

Durch das Verwenden der Spatial Anchor sind die während des Tests erzeugten Datensätze jedoch nicht auf dem Nullpunkt der im Unity Editor verwendeten Ursprünge ( $X=0, Z=0$ ). Beim Aufsetzen des Meta Quest 3 Headsets ist der Nullpunkt der Szene nicht der Nullpunkt des 3D Objekts. Der virtuelle Raum wird an die Position des erkannten Spatial Anchor gesetzt. Daraus resultiert, dass die gespeicherten Marker und zugehörigen Ursprünge an anderer Stelle platziert und auch rotiert werden als mit dem „echten“ Nullpunkt der Unity Szene. Da jedoch die

## 4 Evaluation

---

Ursprünge ihre Relation zueinander behalten, ist eine Berechnung des neuen Nullpunkts, aber auch eine Transformation aller Punkte zu dem echten Nullpunkt  $P(0,0)$  möglich.



*Abbildung 48: Veranschaulichung der Berechnung des Punktes  $P2''$  durch Transformation des Punktes  $P2$  in GeoGebra.*

In Abbildung 48 sind die Schritte der Transformation in einem kartesischen zweidimensionalen Koordinatensystem beispielhaft veranschaulicht. Die Punkte  $A1$  und  $A2$  bilden dabei zwei Ursprünge in einer Unity Szene ab.  $B1$  und  $B2$  sind Ursprünge, die aus den Daten der Hörtests ermittelt wurden. In Unity sind die Punkte  $A1$  und  $A2$  Klangquellen innerhalb des virtuellen Raumes. Da der Raum bei der Neuplatzierung durch den Spatial Anchor innerhalb des Koordinatensystems einer Translation und einer Rotation unterliegt, können die neuen Punkte  $B1$  und  $B2$  irgendwo im Raum auf der X und Z-Achse liegen. Ihre Relation zueinander bleibt aber erhalten. Das bedeutet, dass alle Punkte gleichzeitig verschoben und um deren Nullpunkt rotiert werden.

## 4 Evaluation

---

Zur Vereinfachung wurde für dieses Beispiel ein Lösungsansatz in GeoGebra<sup>56</sup> beschrieben. Die Punkte A1 und A2 sowie deren Nullpunkt P1 werden folgendermaßen definiert:

```
A1=(1.5,1)
A2=(2,0)
P1=(0,0)
```

Die Punkte B1 und B2 werden nun um 51.5 Grad rotiert und dann um den Wert 1 subtrahiert. Dies entspricht einer beispielhaften zufälligen Translation und Rotation, die durch das Platzieren des virtuellen Raumes auf den Spatial Anchor vollführt wird.

```
angle=51.5
B2=Rotate(A2,angle)-1
B1=Rotate(A1,angle)-1
```

Die Herausforderung besteht nun darin, ohne diese Transformationsinformationen den Nullpunkt P2“(-1,-1) zu finden. Dafür wird der neue Nullpunkt P2 in drei Schritten transformiert:

```
P2=(0,0)
P2'=P2-(A1-P1)
```

Zunächst wird der Punkt P2' um den Vektor A1-P1 subtrahiert. Daraufhin muss mithilfe der Vektoren A2-A1 und B2-B1 der Winkel zwischen den Vektoren kalkuliert werden. Dies geschieht über das Skalarprodukt. In GeoGebra liefert die Funktion `Angle(Vektor1, Vektor2)` den gewünschten Winkel in Grad.

```
v_{B1B2}=Vector(B1,B2)
v_{A1A2}=Vector(A1,A2)
theta=Angle(v_{A1A2},v_{B1B2})
```

Jetzt wird der Punkt P2' um den berechneten Winkel theta rotiert. GeoGebra liefert diesen mit der Funktion `Rotate(Punkt, Winkel)`.

```
P2''=Rotate(P2',theta)
```

Nun muss der Punkt P2“ um den verschobenen A1-P1 Vektorpendant B1-P1 addiert werden.

```
P2'''=P2''+B1-P1
```

Das Ergebnis der Transformationen ist der Punkt P2“(-1,-1), welcher dem Nullpunkt der transformierten Punkte B1 und B2 entspricht.

Mit dieser Methodik kann der Nullpunkt der transformierten Ursprünge in Unity gefunden werden, und anschließend durch eine Rücktransformation alle transformierten Ursprünge wieder auf die in der Unity Szene gesetzten Punkt

---

<sup>56</sup> <https://www.geogebra.org/>

## 4 Evaluation

---

vereinheitlicht werden. So können alle benötigten Variablen über den echten gemeinsamen Nullpunkt der Szene berechnet werden.

In Unity wurde dafür eine Kopie der Szene „SpatializingSoundTemplate“ angelegt. Diese wurde in EvaluationScene umbenannt. Darin sind nun alle Ursprünge und deren Positionen zu finden. Alle GameObjects innerhalb der Szene wurden deaktiviert. Von Interesse ist lediglich das MotionCapturingLab GameObject, welches die Klangquellen im GameObject SoundSources als Prefab beinhaltet. Der Szene wurde ein GameObject TestCalculation hinzugefügt. Das Skript `TestDeserialization` konnte nun die oben beschriebenen Transformationen durchführen und danach alle Variablen in .csv Dateien kulminiert abspeichern. Dafür benötigt die Komponente die Origin Points, also das SoundSources GameObject, zur Berechnung der Transformationen. Außerdem fordert es die Rohdaten der .json Dateien sowie die Ergebnisse des Fragebogens in Form einer Pfadangabe. Weiters wurde der Ausgabepfad für die SPSS-Dateien angegeben. Beim Starten der Szene wurde die Datentransformation automatisch angestoßen.

Zunächst werden pro Ordner alle ListeningTestData Instanzen aus den .json Dateien deserialisiert. Pro Spatializer werden nun die Testdaten in ein neues Datenset überführt. Die `SPSSPlottingDataSet` Klasse beschreibt neben dem `UserIdentifier` und der `SampleID` die Punkte X und Y der Ursprünge und deren platzierte Marker, nach der Transformation auf den Nullpunkt der Unity Szene (0,0). Außerdem wird die berechnete Distanzdifferenz und Azimuthdifferenz zwischen dem Ursprung und Marker sowie die benötigte Zeit zum Platzieren des Markers gespeichert. Mit diesem Datenset kann wiederum der Durchschnitt pro Testperson berechnet werden. Diese werden mit der Klasse `SPSSPlottingMeanDataSet` definiert.

Das Skript erzeugt nun in der `Start()` Methode eine Liste an `SPSSPlottingData`. Alle deserialisierten `ListeningTestData` Instanzen durchlaufen die Methode `GenerateSPSSPlottingDataSet`. Zunächst wird dem Datenset wie oben beschrieben der neue Nullpunkt mithilfe der Methode `GetCenterPointFromTestData` gefunden.

Um in Unity eine Translation und Rotation durchführen zu können, muss ein GameObject in der Szene erzeugt werden. Dieses GameObject hat nun eine `Transform` Komponente, welche auf dem Nullpunkt der Szene liegt. Aus den `OriginPoints` werden die Punkte AS01 und AS02 zur Vektor Berechnung und Translation extrahiert. Diese Punkte AS01 und AS02 werden ebenfalls aus dem Testdatensatz extrahiert.

## 4 Evaluation

---

```
// get all data
var centerOld = Vector3.zero;
var centerNew = new GameObject().transform;
var as01Old =
OriginPoints.GetComponentInChildren<SampleIdentifier>().FirstOrDefault(i => i.Name == "AS01");
var as01New = testData.Results.FirstOrDefault(p => p.Name == "AS01").Origin;
```

Nun werden die Vektoren erzeugt sowie deren einschließender Winkel mithilfe der `Vector3.SignedAngle(Vector3 from, Vector3 to, Vector3 axis)`<sup>57</sup>

Methode berechnet.

```
// set AS01 as new center
centerNew.position = centerOld - as01Old.transform.position;

// rotate
var normalizedVOld = GetNormalizedOriginVector("AS01", "AS02");
var normalizedVNew = GetNormalizedMarkerVector(testData, "AS01", "AS02");
var degrees = Vector2.SignedAngle(normalizedVNew, normalizedVOld);
centerNew.RotateAround(centerOld, Vector3.up, degrees);
```

Mit der Methode `Transform.RotateAround(Vector3 point, Vector3 axis, float angle)`<sup>58</sup> kann das `GameObject` rotiert werden. Translationen werden über die Subtraktion oder Addition der `Vector3` Klassen durchgeführt.

```
// add new AS01 position
centerNew.position = centerNew.position + as01New;
var center = new Vector2(centerNew.position.x, centerNew.position.z);
Destroy(centerNew.gameObject);
return center;
```

Mit dem neuen Nullpunkt kann nun jeder Ursprung und Marker aus dem `ListeningTestData` auf den echten Nullpunkt überführt werden. Dies geschieht in der Methode `GenerateSPSSPlottingDataSet` für jeden Datenpunkt wie folgt:

```
originToMove.position = dataPoint.Origin - centerNew;
markerToMove.position = dataPoint.Marker - centerNew;
originToMove.RotateAround(centerOld, Vector3.up, degrees);
markerToMove.RotateAround(centerOld, Vector3.up, degrees);
originToMove.position = originToMove.position + centerOld;
markerToMove.position = markerToMove.position + centerOld;
```

---

<sup>57</sup> <https://docs.unity3d.com/ScriptReference/Vector3.SignedAngle.html>

<sup>58</sup> <https://docs.unity3d.com/ScriptReference/Transform.RotateAround.html>

## 4 Evaluation

---

Nun werden die benötigten Werte Distanz- und Azimuthdifferenz berechnet. Alle Datensätze in Form der `SPSSPlottingDataSet` Klasse werden nun in einer Liste gesammelt und der `GenerateSPSSPlottingDataMean` Methode übergeben. Diese berechnet nun die Mittelwerte der Variablen Distanzdifferenz, Azimuthdifferenz und die X und Y Koordinaten der Marker aller Teilnehmer pro Sample und Spatializer. Pro Eintrag wird außerdem noch ein zusätzlicher Eintrag erzeugt, welcher für das Abspeichern der Ursprungspositionen verantwortlich ist. Dieser Eintrag ist vor allem zur graphischen Veranschaulichung da, da er sich durch die vorangegangene Transformation nicht von den anderen Ursprüngen seines Samples unterscheidet. Mithilfe des `CSVFile` Nuget Packages konnte die transformierte Liste nun als „`spssPlottingDataMeanFile.csv`“ gespeichert und in SPSS geöffnet werden.

Weiters sind für die Evaluation die Antworten des Online-Fragebogens von Interesse. Die Klasse `SPSSDataSet` speichert die im Fragebogen selbst eingeschätzten Werte zur Lokalisationsgenauigkeit (`Localization_Accuracy`), Natürlichkeit (`Voice_Naturalness`) und Lokalisierungsschwierigkeit (`Location_Difficulty`) pro Teilnehmeridentifikation und Spatializer. Zusätzlich wird der Durchschnitt der Distanzdifferenz und Azimuthdifferenz, aber auch der Durchschnitt bei weiter entfernten Klangquellen (> 5 Meter) und der nahen Quellen (<= 5m) berechnet und pro Sample aufsummiert. Diese Werte sind nicht Teil einer Forschungsfrage oder Hypothese, können aber zur Aufschlüsselung und graphischen Darstellung der Unterschiede der Audio Spatializer beitragen. Die Berechnung der `SPSSDataSets` geschieht in der `CalculateDataSetForUser` Methode, die Transformation des `QuestionnaireDataSet` in das `SPSSDataSet` geschieht in der Methode `AppendQuestionairData`.

Die `QuestionnaireDataSet` Klasse ist die Abbildung der von Google Forms heruntergeladenen `.csv` Datei. Eine Zeile in der `.csv` Datei entspricht einem Datensatz einer Testperson mit einer Teilnehmeridentifikation. Diese Zeile wird in der Methode `GetQuestionairDataSet()` in eine Instanz von `QuestionnaireDataSet` überführt. Sie speichert also die Teilnehmeridentifikation und die `Localization_Accuracy`, `Voice_Naturalness` und `Location_Difficulty` pro Spatializer.

Auch dieses Datenset wurde nun gesammelt in die Datei „`spssDataFile.csv`“ gespeichert, um sie mit SPSS auszulesen.

## 4.2 Ergebnisse

### 4.2.1 Forschungsfrage 1

**Wie präzise lassen sich virtuelle Soundquellen mit unterschiedlichen Audio Spatializern in einer Mixed Reality Umgebung verorten?**

Hypothese:

Aufgrund der klanglichen Unterschiede in Bezug auf die Auralisation (wie Erstreflexionen oder Nachhall) einer virtuellen Klangquelle durch einen Audio Spatializer unterscheidet sich die Lokalisierung derselben Klangquelle, unabhängig ihrer spektralen und temporalen Eigenschaften, im Vergleich zu den anderen Spatializern.

Nullhypothese:

Es findet sich kein signifikanter Unterschied in der Lokalisierbarkeit von virtuellen Klangquellen zwischen den vier getesteten Audio Spatializern.

Die in Unity aufbereiteten Daten liefern die durchschnittliche X und Z Positionen der gesetzten Marker. Diese können mithilfe eines Streudiagramms zu jeder virtuellen Klangquelle pro Audio Spatializer dargestellt werden. Zusätzlich wird der Ursprung zu jedem Sample angezeigt. Daraus kann eine erste Einschätzung der Lokalisierbarkeit der Quellen gewonnen werden. Darauf folgen die Ergebnisse zu den Variablen Distanzdifferenz und Azimuthdifferenz.

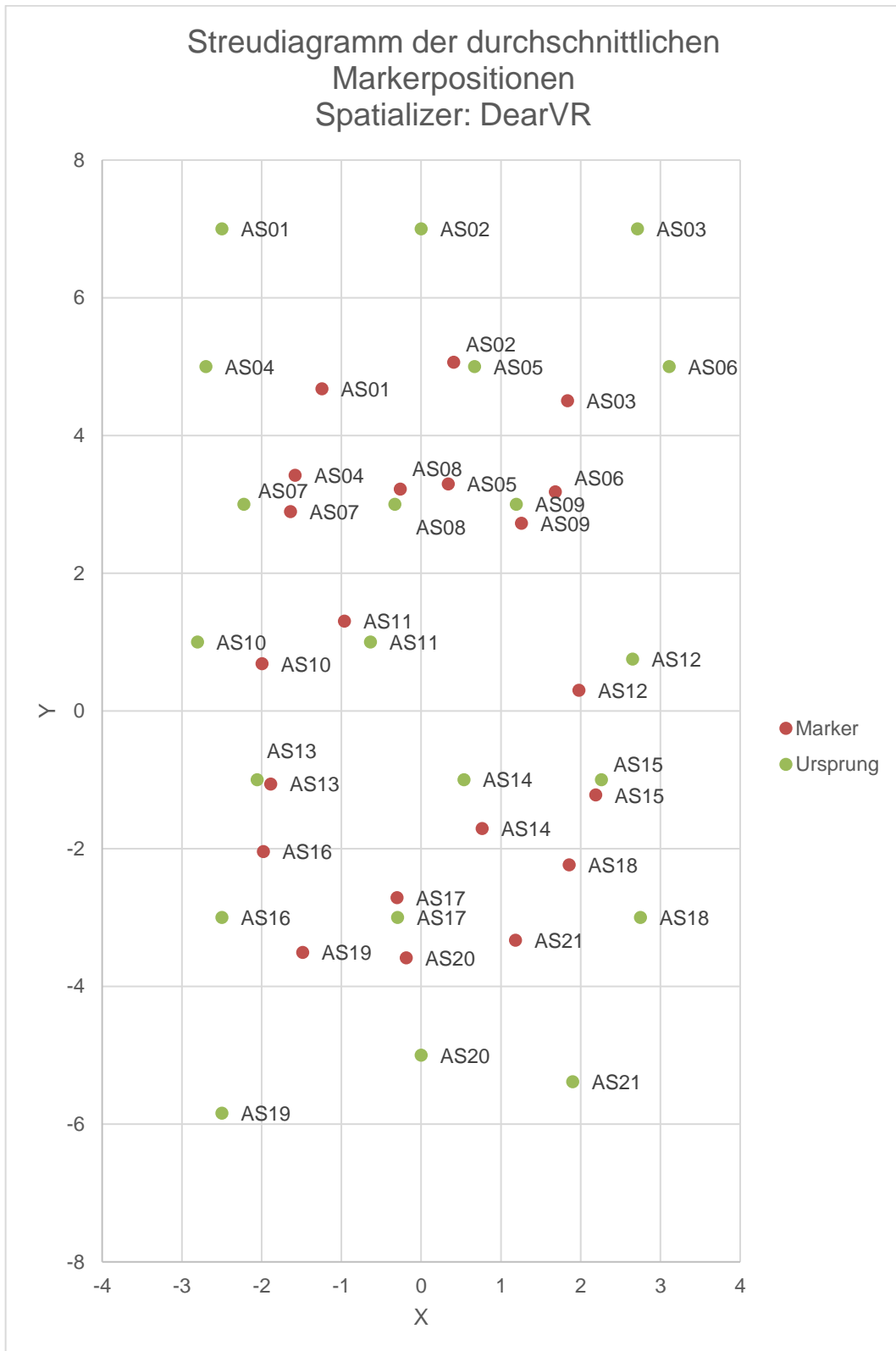


Abbildung 49: Streudiagramm der durchschnittlichen Markerpositionen des DearVR Spatializers

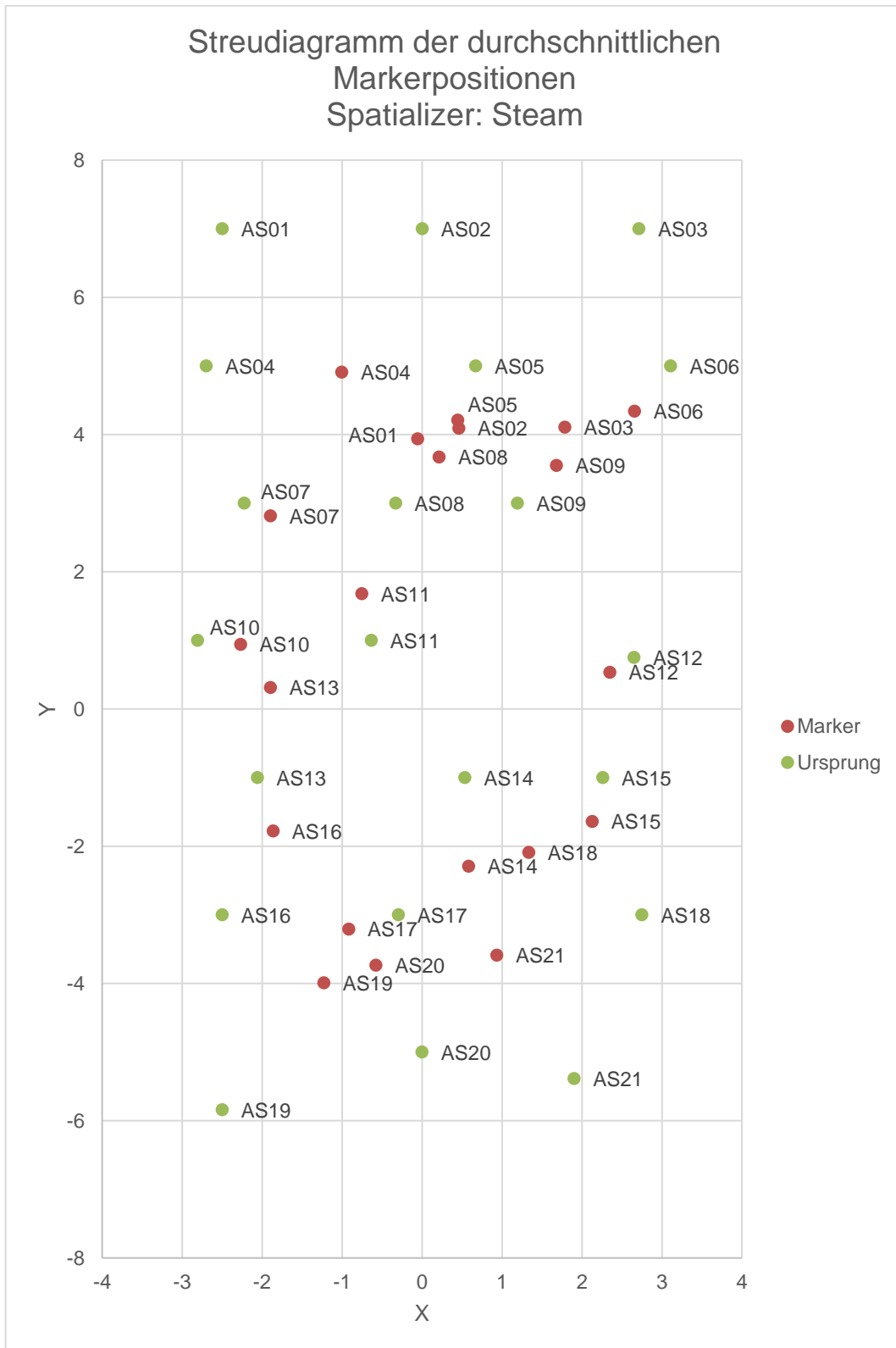


Abbildung 50: Streudiagramm der durchschnittlichen Markerpositionen des Steam Spatializers

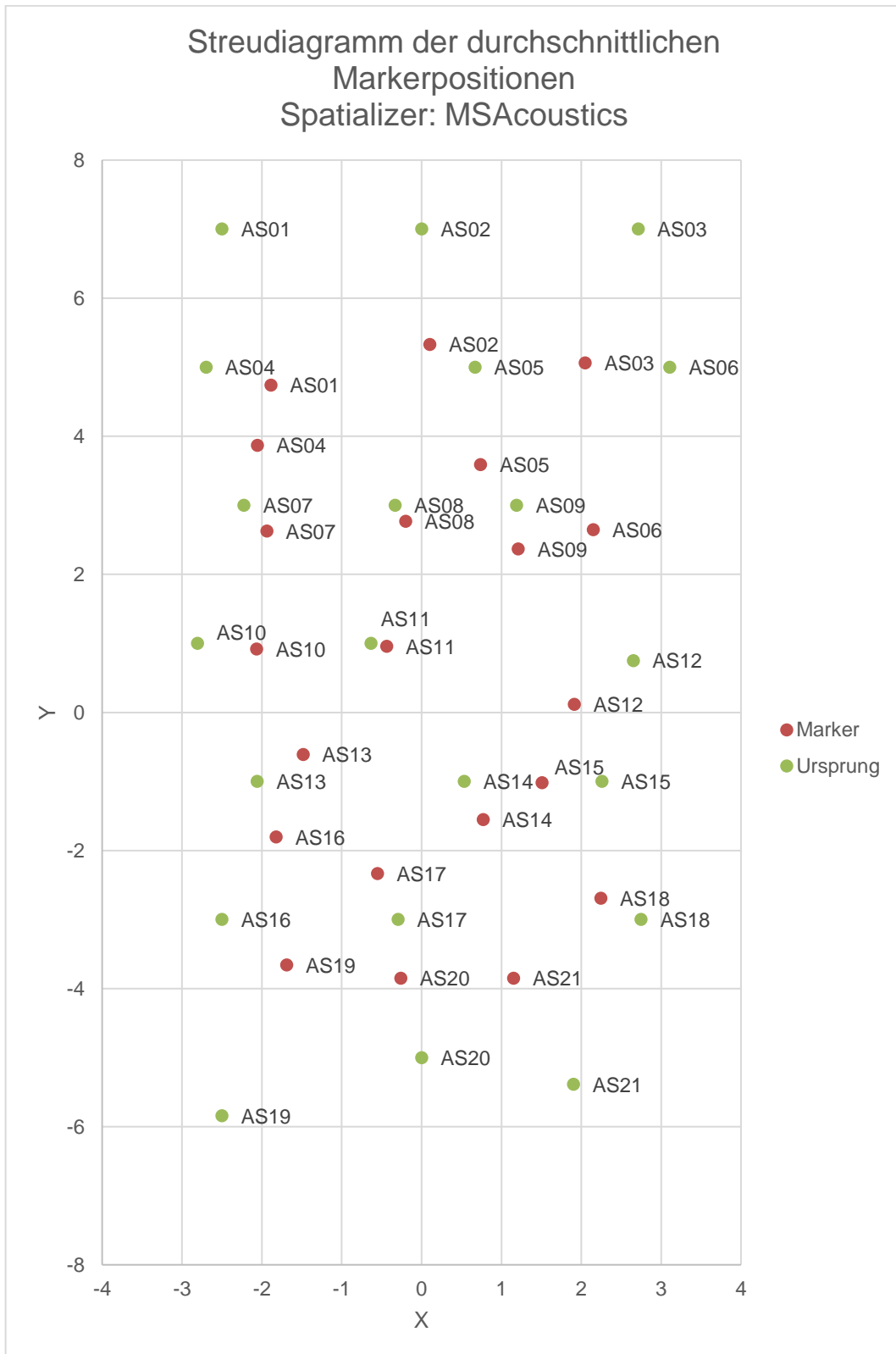


Abbildung 51: Streudiagramm der durchschnittlichen Markerpositionen des MSAcoustics Spatializers

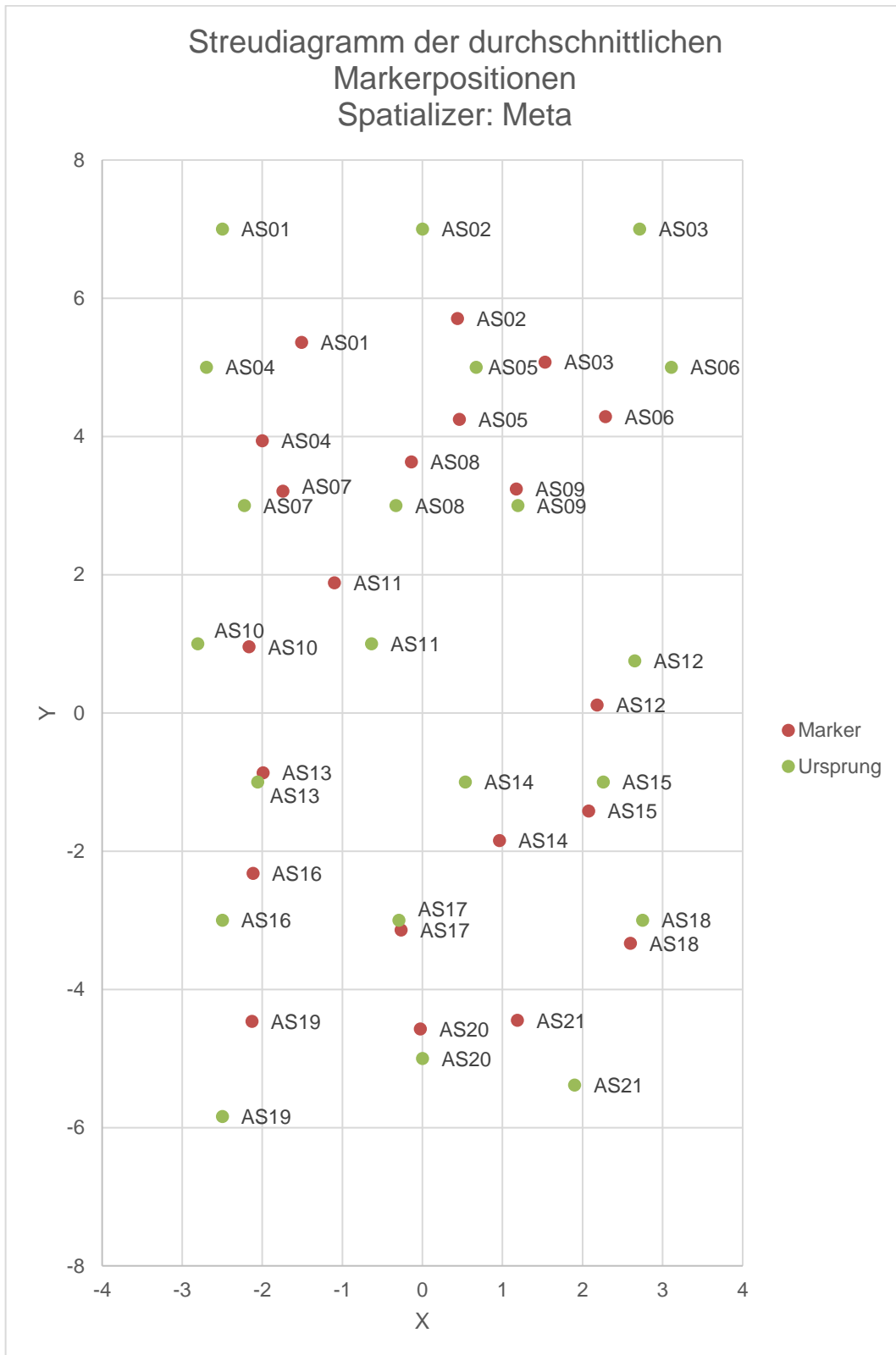


Abbildung 52: Streudiagramm der durchschnittlichen Markerpositionen des Meta Spatializers

### 4.2.1.1 Distanzdifferenz

Ein Faktor zur Beschreibung der Lokalisierbarkeit von Klangquellen ist die Distanzerkennung (siehe 2.1.3 Distanzerkennung). Die in Unity aufbereiteten Daten liefern die durchschnittliche Distanzdifferenzen zwischen Marker und Ursprung zu jeder virtuellen Klangquelle pro Audio Spatializer. Diese berechnet sich aus der Länge des Vektors des Markers und die des Ursprungs. Ist eine Klangquelle zu kurz verortet worden, so hat die Distanzdifferenz ein negatives Vorzeichen. In SPSS werden die Daten graphisch aufbereitet:

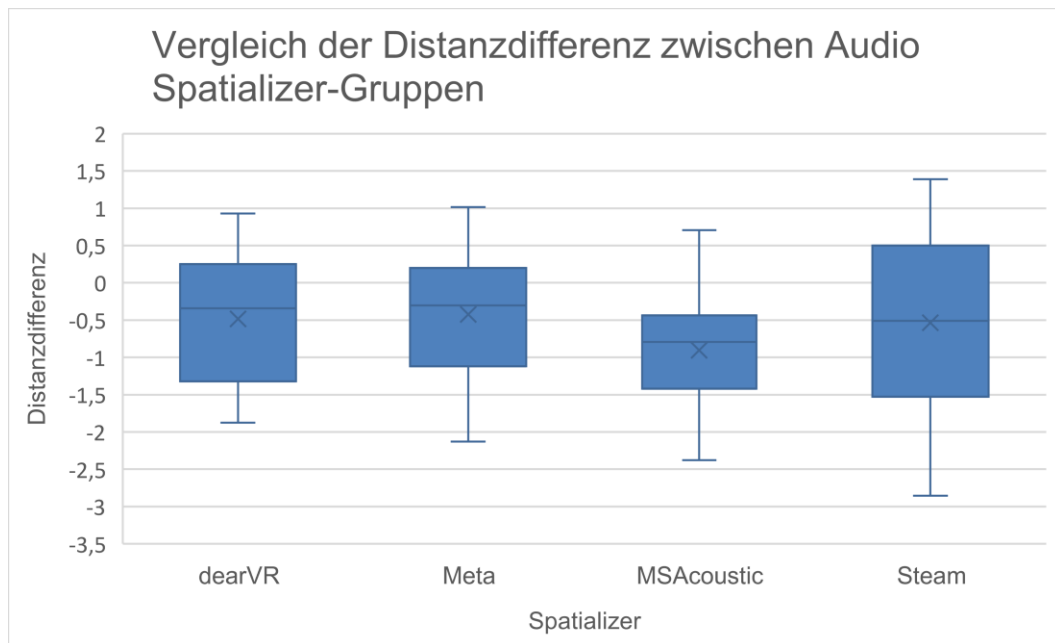


Abbildung 53: Boxplot - Vergleich der Distanzdifferenz zwischen Audio Spatializer

In dieser Zusammenfassung werden die deskriptiven Statistiken der Distanzdifferenzen für die verschiedenen Audio Spatializer präsentiert. Die Daten wurden von insgesamt 17 Testpersonen generiert, wobei jede Person für alle vier Spatializer jeweils 21 Werte produzierte.

#### DearVR:

Die durchschnittliche Distanzdifferenz für DearVR beträgt -0,49m, basierend auf den 21 von den Testpersonen generierten Werten. Die Spannweite der Werte erstreckt sich von -1,88m bis 0,93m, mit einer Standardabweichung von 0,87m.

#### Steam:

Für Steam ergibt sich eine durchschnittliche Distanzdifferenz von -0,53m, basierend auf den 21 Werten der Testpersonen. Die Werte schwanken zwischen -2,86m und 1,39m, und die Standardabweichung beträgt 1,22m.

## 4 Evaluation

---

### MSAcoustics:

Die durchschnittliche Distanzdifferenz bei MSAcoustics beträgt  $-0,90\text{m}$ , basierend auf den 21 Werten der Testpersonen. Die Spannweite der Werte liegt zwischen  $-2,38\text{m}$  und  $0,71\text{m}$ , mit einer Standardabweichung von  $0,73\text{m}$ .

### Meta:

Für Meta beträgt die durchschnittliche Distanzdifferenz  $-0,42\text{m}$ , basierend auf den 21 Werten der Testpersonen. Die Werte erstrecken sich von  $-2,13\text{m}$  bis  $1,02\text{m}$ , und die Standardabweichung beträgt  $0,86\text{m}$ .

Bemerkenswert ist, dass mit allen Spatializer im Mittel zu kurz geschätzt wurde. Die könnte auf einen von den Spatializern zu klein auralisierten Raum hinweisen.

### 4.2.1.2 Azimuttdifferenz

Ein weiterer Faktor zur Beschreibung der Lokalisierbarkeit von Klangquellen ist die Lokalisierung in horizontaler Ebene (siehe 2.1.1 Lokalisierung in horizontaler Ebene). Die in Unity aufbereiteten Daten liefern die durchschnittliche Azimuttdifferenz zwischen Marker und Ursprung zu jeder virtuellen Klangquelle pro Audio Spatializer. Diese berechnet sich aus dem Winkel zwischen den Vektoren des Markers und des Ursprungs. Ist eine Klangquelle zu weit links verortet worden, so hat die Azimuttdifferenz ein negatives Vorzeichen. In SPSS werden die Daten graphisch aufbereitet:

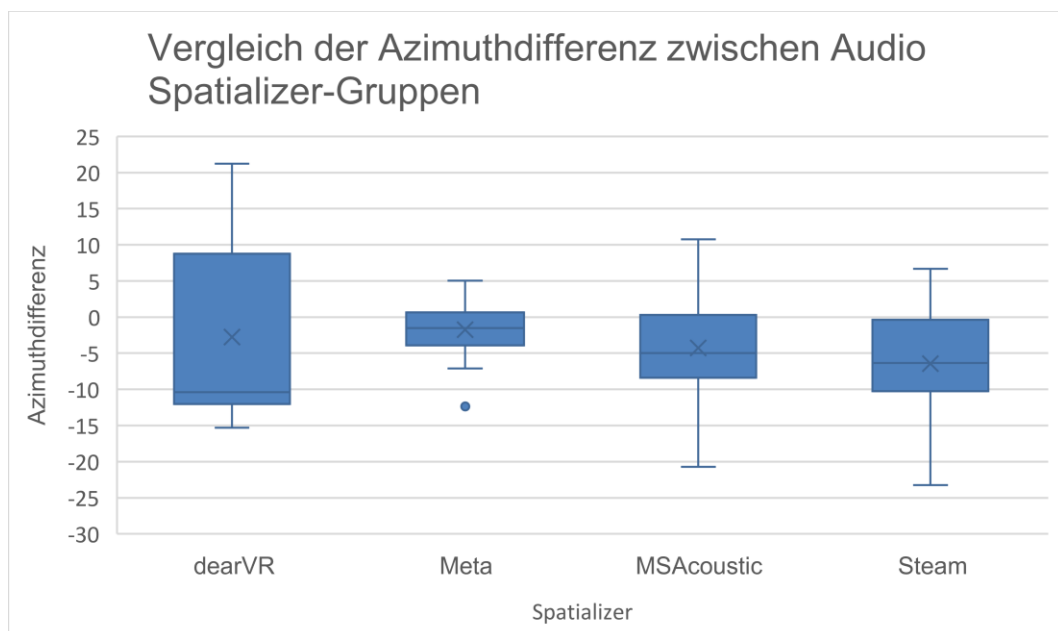


Abbildung 54: Boxplot - Vergleich der Azimuttdifferenz zwischen Audio Spatializer

## 4 Evaluation

---

In dieser Zusammenfassung werden die deskriptiven Statistiken der Azimuthdifferenzen für die verschiedenen Audio Spatializer präsentiert. Die Daten wurden von insgesamt 17 Testpersonen generiert, wobei jede Person für alle Gruppen jeweils 21 Werte produzierte.

### DearVR:

Die durchschnittliche Azimuthdifferenz für DearVR beträgt etwa  $-2,76^\circ$ , basierend auf den 21 von den Testpersonen generierten Werten. Die Spannweite der Werte erstreckt sich von  $-15,29^\circ$  bis  $21,23^\circ$ , mit einer Standardabweichung von  $11,44^\circ$ .

### Steam:

Für Steam beträgt die durchschnittliche Azimuthdifferenz etwa  $-6,43^\circ$ , basierend auf den 21 Werten der Testpersonen. Die Werte erstrecken sich von  $-23,24^\circ$  bis  $6,67^\circ$ , und die Standardabweichung beträgt  $7,64^\circ$ .

### MSAcoustics:

Die durchschnittliche Azimuthdifferenz bei MSAcoustics beträgt etwa  $-4,25^\circ$ , basierend auf den 21 Werten der Testpersonen. Die Spannweite der Werte liegt zwischen  $-20,70^\circ$  und  $10,77^\circ$ , mit einer Standardabweichung von  $6,46^\circ$ .

### Meta:

Für Meta ergibt sich eine durchschnittliche Azimuthdifferenz von etwa  $-1,74^\circ$ , basierend auf den 21 Werten der Testpersonen. Die Werte schwanken zwischen  $-12,36^\circ$  und  $5,05^\circ$ , und die Standardabweichung beträgt  $3,93^\circ$ .

Interessanterweise wurden in mit allen Spatializern im Mittel die Marker zu weit links platziert, was zu den negativen Werten führt.

### 4.2.2 Forschungsfrage 2

**Sind aktuell am Markt verfügbare Audio Spatializer zur Lokalisierung von virtuellen Klangquellen ohne weitere unterstützender Hard/Software wie personalisierte HRTFs (= head-related transfer functions) in einer Mixed Reality Umgebung geeignet?**

Hypothese:

Manche Spatializer externalisieren virtuelle Klangquellen besser als andere. Dadurch ist eine im Vergleich bessere Lokalisierung gewährleistet und beeinflusst die Wahl des Audio Spatializers.

Nullhypothese:

Die Spatializer unterscheiden sich in ihrer Externalisierung nicht untereinander. Es lässt sich keine Wahl zur besten Lokalisierung treffen.

#### 4.2.2.1 Distanzdifferenz

Demnach zeigt die Analyse der Distanzdifferenzen, dass der Meta Spatializer die präziseste Distanzlokalisierung aufweist, charakterisiert durch die geringste Standardabweichung von 0,86m und die niedrigste durchschnittliche Distanzdifferenz von -0,42m. Aufgrund dieser Ergebnisse lässt sich schlussfolgern, dass die räumliche Verortung von Klangquellen mit dem Meta Spatializer im Vergleich zu den anderen Spatializern tendenziell genauer und konsistenter erfolgt.

Darauf folgt der DearVR Spatializer, der trotz einer leicht höheren durchschnittlichen Distanzdifferenz von -0,49m eine geringere Standardabweichung von 0,87m aufweist. Diese Feststellung deutet darauf hin, dass, obwohl die durchschnittliche Distanzdifferenz etwas höher ist, eine ähnlich konsistente Distanzlokalisierung mit dem DearVR Spatializer gewährleistet ist, wie mit dem Meta Spatializer.

Die weiteren Spatializer, MSAcoustics und Steam, zeigen eine absteigende Präzision in der Distanzlokalisierung, wobei MSAcoustics eine höhere durchschnittliche Distanzdifferenz von -0,90m und eine Standardabweichung von 0,73m aufweist, während Steam eine durchschnittliche Distanzdifferenz von -0,53m und eine höhere Standardabweichung von 1,22m verzeichnet.

Die im Kapitel 2.1.6 Toleranzbereiche beschriebenen Werte zur Distanzdifferenz zeigen, dass sich die erhobenen Werte des DearVR, Meta und Steam Spatializer innerhalb des Toleranzbereichs von  $\pm 0,55\text{m}$  befinden.

### 4.2.2.2 Azimuttdifferenz

Die Untersuchung der Azimuttdifferenzen zeigt, dass der Meta Spatializer durch die geringste Standardabweichung von  $3,93^\circ$  und die niedrigste durchschnittliche Azimuttdifferenz von  $-1,74^\circ$  die präziseste Lokalisierbarkeit in horizontaler Ebene aufweist. Die Ergebnisse des Hörtests lassen darauf schließen, dass der Meta Spatializer bei der Lokalisierung von Klangquellen im Vergleich zu den anderen Spatializern tendenziell eine höhere Genauigkeit und Konsistenz aufweist.

Wird lediglich die durchschnittlichen Azimuttdifferenz von  $-2,76^\circ$  betrachtet, ist der DearVR Spatializer nur um etwa  $1^\circ$  unpräziser als der Meta Spatializer. Jedoch zeigt sich hier die höchste Standardabweichung von  $11,44^\circ$ . Dies deutet darauf hin, dass, die Lokalisierung auf horizontaler Ebene mit diesem Spatializer inkonsistenter ist, als mit den anderen.

Die Spatializer MSAcoustics und Steam zeigen eine schwächere Lokalisierungspräzision mit ähnlichen Werten. MSAcoustics weist eine höhere durchschnittliche Azimuttdifferenz von  $-4,25^\circ$  und eine Standardabweichung von  $6,46^\circ$  auf, während Steam eine durchschnittliche Azimuttdifferenz von  $-6,43^\circ$  und eine höhere Standardabweichung von  $7,64^\circ$  verzeichnet.

Der beschriebene Toleranzbereich umfasst  $\pm 3,6^\circ$ . Der Meta Spatializer und der DearVR Spatializer erfüllen dieses Kriterium.

### 4.2.2.3 Übersicht

Spatializer	Distanz-differenz	Toleranzbereich (+-0,55m)	Azimuth-differenz	Toleranzbereich (+-3,6°)
Meta	-0,42m	Ja	$-1,74^\circ$	Ja
DearVR	-0,49m	Ja	$-2,76^\circ$	Ja
Steam	-0,53m	Ja	$-6,43^\circ$	Nein
MSAcoustics	-0,90m	Nein	$-4,25^\circ$	Nein

*Tabelle 2: Übersicht der Mittelwerte der Distanzdifferenzen und Azimuttdifferenzen und Einordnung in Toleranzbereiche*

### 4.2.3 Varianzanalyse

In diesem Experiment wurden vier verschiedene Audio-Spatializer (Gruppen) auf ihre Auswirkungen auf die wahrgenommenen Distanz- und Azimuttdifferenzen bei 21 virtuellen Klangquellen getestet. Die abhängigen Variablen waren "Distanzdifferenz" und "Azimuttdifferenz", gemessen in einem Stichprobenumfang von 17 Personen.

### **Distanzdifferenz:**

- Zwischen den Gruppen: Quadratsumme = 2,933, Freiheitsgrade (df) = 3, Mittel der Quadrate = 0,978, F-Wert = 1,111, p-Wert = 0,350 (nicht signifikant).
- Innerhalb der Gruppen: Quadratsumme = 70,391, df = 80, Mittel der Quadrate = 0,880.
- Gesamt: Quadratsumme = 73,324, df = 83.

Die einfaktorielle ANOVA für die Distanzdifferenz zeigt keine signifikanten Unterschiede zwischen den getesteten Audio Spatializern ( $F(3, 80) = 1,111$ ,  $p = 0,350$ ).

### **Azimuthdifferenz:**

- Zwischen den Gruppen: Quadratsumme = 261,199, df = 3, Mittel der Quadrate = 87,066, F-Wert = 1,414, p-Wert = 0,245 (nicht signifikant).
- Innerhalb der Gruppen: Quadratsumme = 4926,899, df = 80, Mittel der Quadrate = 61,586.
- Gesamt: Quadratsumme = 5188,098, df = 83.

Auch für die Azimuthdifferenz konnte keine signifikante Varianz zwischen den Audio Spatializern nachgewiesen werden ( $F(3, 80) = 1,414$ ,  $p = 0,245$ ).

**Die Hypothesen der Forschungsfrage 1 und 2 müssen daher verworfen werden.**

## 4.3 Weitere Ergebnisse

Neben den durch die Forschungsfragen erforderlichen durchschnittlichen Werte zu Distanz- und Azimuthdifferenz pro virtueller Klangquelle wurden noch weitere Daten erhoben die im nachfolgenden beleuchtet werden sollen.

### 4.3.1 Online-Fragebogen

Die durch die in Unity Transformierten Datensätze bilden die Ergebnisse des Fragebogens in Form von drei Variablen ab: Lokalisationsgenauigkeit (Localization\_Accuracy), Natürlichkeit der Stimme (Voice\_Naturalness) und Lokalisationsschwierigkeit (Location\_Difficulty). Die Testteilnehmer:innen wurden gebeten, nach jedem Hörtest zur aktuellen Applikation die drei Variablen zwischen 1 und 5 zu bewerten.

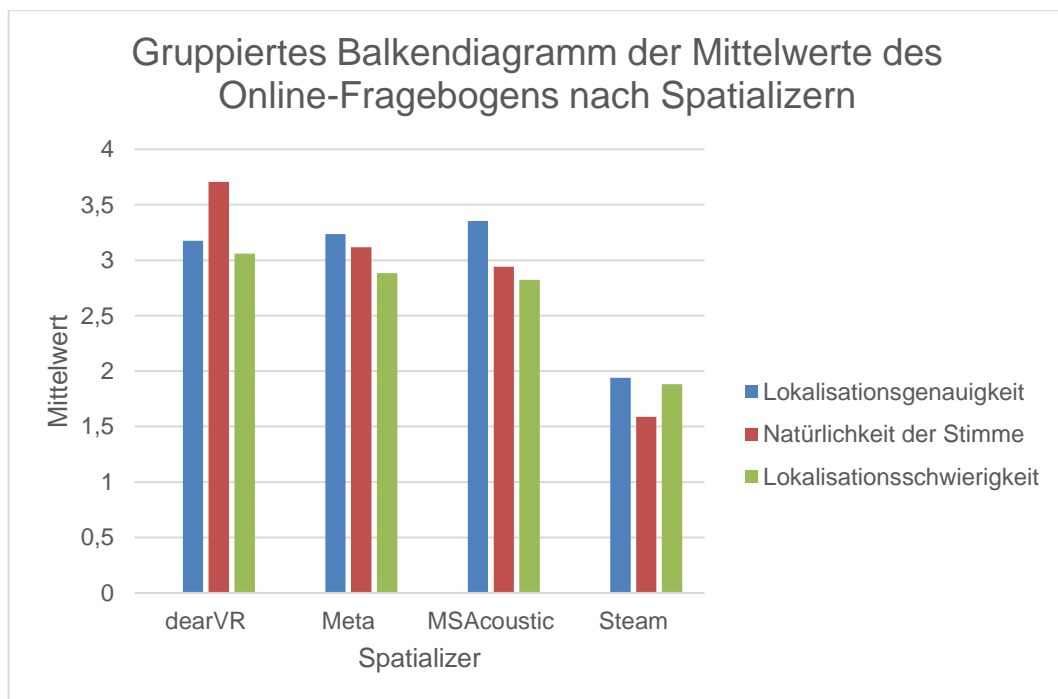


Abbildung 55: Gruppirtes Balkendiagramm der Mittelwerte des Online-Fragebogens nach Spatializern

Die Lokalisationsgenauigkeit beschreibt dabei die von der Testperson nach einem Hörtest mit einer Mixed-Reality-Applikation selbst eingeschätzte Leistung beim Lokalisieren der virtuellen Klangquellen. Je höher dieser Wert ist, umso eher schätzten die Testpersonen die korrekten Positionen ein. Da es für die Testteilnehmer:innen nicht möglich ist während des Hörtests einzusehen, ob sie

## 4 Evaluation

---

eine Hörquelle gut oder schlecht platziert haben, ist dieser Parameter eine reine Selbsteinschätzung.

Im Gegensatz dazu wird mit der Lokalisationsschwierigkeit die Schwierigkeit abgefragt, mit der eine Klangquelle verortet werden kann. Je höher der Wert, umso leichter fiel es der Testperson, die Klangquelle zu verorten.

Die Natürlichkeit der Stimme beschreibt den Eindruck der Testpersonen auf die Auralisierung. Wurde die Stimme als natürlich wahrgenommen wird diese mit einem hohen Wert beschrieben.

Spatializer	Lokalisationsgenauigkeit		Natürlichkeit der Stimme		Lokalisations-schwierigkeit	
	M	SD	M	SD	M	SD
DearVR	3,18	0,951	3,71	0,772	3,06	0,827
Meta	3,24	0,752	3,12	0,781	2,88	0,781
MSAcoustics	3,35	0,862	2,94	1,029	2,82	0,809
Steam	1,94	0,827	1,59	0,618	1,88	0,857

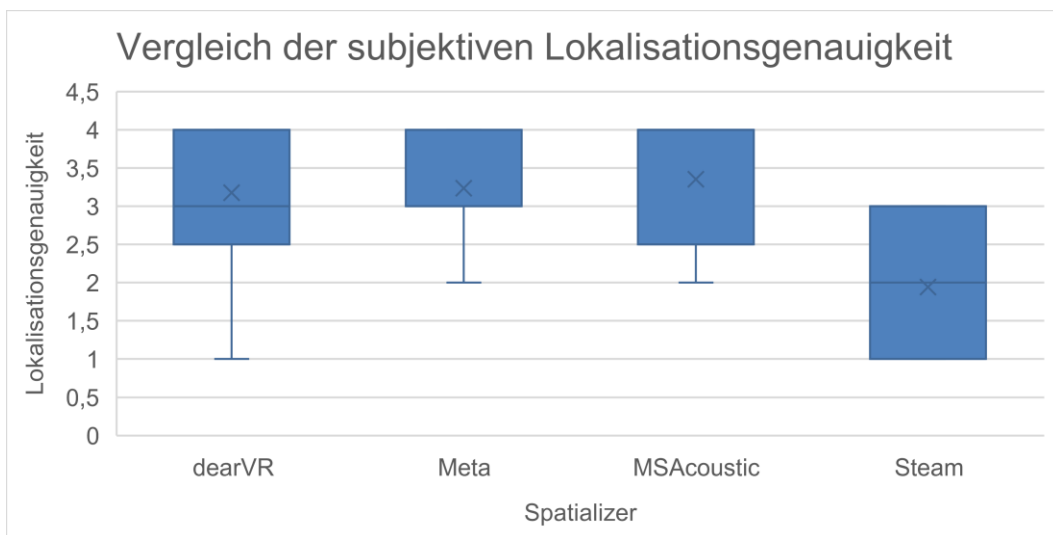


Abbildung 56: Boxplot – Vergleich der Lokalisationsgenauigkeit. Ein hoher Wert bedeutet eine hohe Lokalisationsgenauigkeit (Selbsteinschätzung).

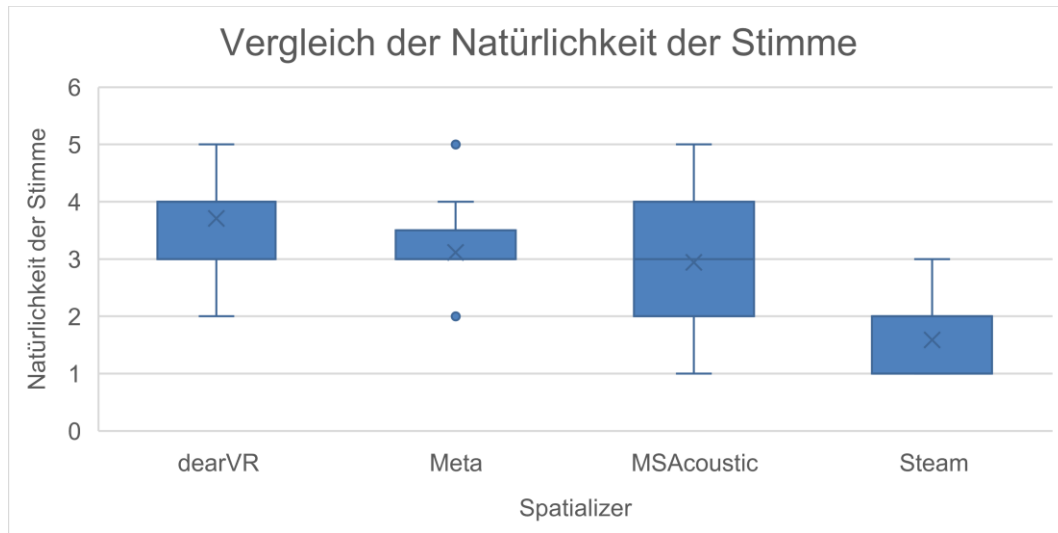


Abbildung 57: Boxplot – Vergleich der Natürlichkeit der Stimme. Ein hoher Wert bedeutet eine hohe Natürlichkeit.

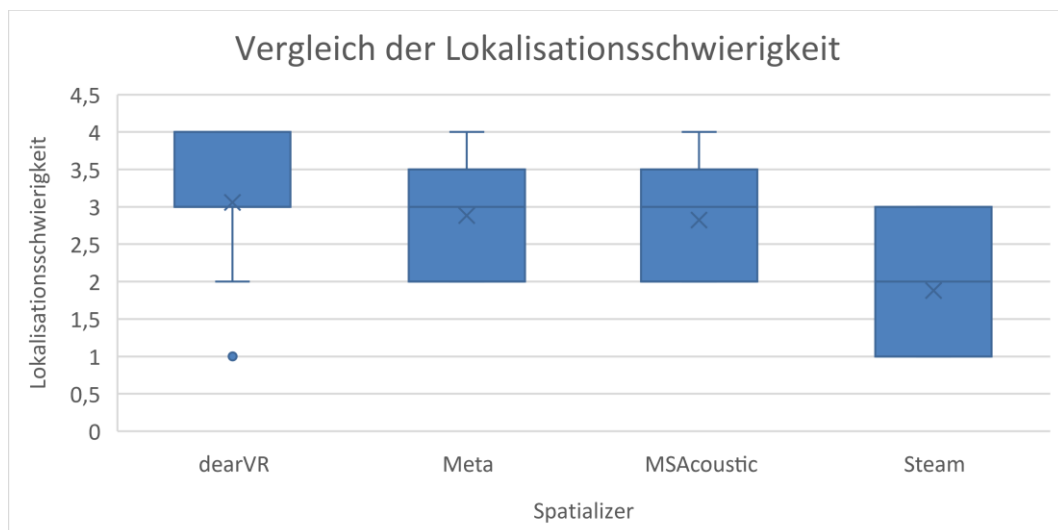


Abbildung 58: Boxplot – Vergleich der Lokalisationsschwierigkeit. Ein hoher Wert bedeutet eine einfache Lokalisierbarkeit.

Zusätzlich zu den Bewertungen konnten die Testpersonen zu jedem Spatializer einen Text verfassen. Sie wurden gebeten, Besonderheiten bei der Durchführung zu beschreiben. Im nachfolgenden werden diese Texte sinngemäß zusammengefasst.

### 4.3.1.1 DearVR

Die Rückmeldungen der Testpersonen zum DearVR Spatializer lassen folgende wesentliche Punkte erkennen:

## 4 Evaluation

---

### Klangcharakteristiken und Phasenstörungen:

- Eigenartige Phasenstörungen werden beim Drehen und Lokalisieren in der Nähe wahrgenommen.
- Die Nähe ist leichter zu orten als weit entfernte Klangquellen. Auf Distanz wirkt die Stimme eher wie aus Lautsprechern kommend und füllt den Raum.

### Natürlichkeit der Stimme:

- Die Stimme wird als natürlicher wahrgenommen als in vorherigen Tests.
- In manchen Situationen wird ein leichtes Rauschen als unangenehm empfunden.

### Hall und Raumwahrnehmung:

- Ähnlich dem ersten Test (MSAcoustics), jedoch mit weniger Hall. Die Klangquelle lässt sich besser lokalisieren, passt jedoch nicht perfekt zum Raum.

### Hintergrundgeräusche und Fiepen:

- Hintergrundgeräusche sind nicht störend, jedoch wird ein hohes Fiepen wahrgenommen.

### Tiefenwahrnehmung und Hall:

- Die Tiefenwahrnehmung gestaltet sich als schwierig, während die Links/Rechts-Wahrnehmung (Azimuth) als gut beschrieben wird.
- Der Hall wird als recht groß wahrgenommen.

### Bester Hall und Qualität des Audiosamples:

- Es wird angenommen, dass dies der beste Hall in Bezug auf die Lokalisierung der Stimme im Raum ist.
- Die Klarheit der Stimme wird als besonders gut wahrgenommen.

Insgesamt deuten die Rückmeldungen darauf hin, dass der DearVR Spatializer positiv in Bezug auf die Natürlichkeit der Stimme und die Lokalisierung im Vergleich zu vorherigen Tests bewertet wird. Während des Tests konnte bei zwei Testteilnehmern ein Fiepen wahrgenommen werden, welches beim Abspielen der Klangquellen wahrnehmbar war.

### 4.3.1.2 *Meta*

Die Rückmeldungen der Testpersonen zum Meta Spatializer können wie folgt zusammengefasst werden:

Lautstärke und Distanzeinschätzung:

- Die Stimme wird als zu leise wahrgenommen, was die genaue Lokalisierung erschwert.
- Die meisten Klangquellen fühlen sich weit entfernt an, und die genaue Distanz ist schwer zu bestimmen.
- Einige Teilnehmer hatten Schwierigkeiten, die Distanz wegen der Lautstärke zu schätzen.
- Die Distanzeinschätzung war generell schwieriger als beim ersten Test (MSAcoustics), insbesondere bei maximaler Distanz.

Richtungslokalisierung und Raumwahrnehmung:

- Die Richtungslokalisierung wird als leicht empfunden, während die Distanzeinschätzung Schwierigkeiten bereitet.
- Die Rückmeldung deutet darauf hin, dass der virtuelle Raum als kleiner oder enger empfunden wird als der echte Raum.

Hintergrundgeräusche und Hall:

- Hintergrundgeräusche wurden als etwas störender empfunden, insbesondere weil die Versuchsstimme leiser war.
- Der Hall wird als natürlicher empfunden.

Sonstige Beobachtungen:

- Die Stimme wird als angenehm wahrnehmbar beschrieben.

Insgesamt deuten die Rückmeldungen darauf hin, dass die Lautstärke, Distanzeinschätzung und Richtungslokalisierung herausfordernde Aspekte beim Test mit dem Meta Spatializer waren. Es gibt jedoch auch positive Aspekte wie die als natürlicher empfundene Raumakustik und angenehme Stimme.

### 4.3.1.3 *MSAcoustics*

Reverb und Raumcharakteristik:

- Die Reverb-Simulation wird als nicht der Realität des Raumes entsprechend wahrgenommen.

## 4 Evaluation

---

- Die Entfernung ist aufgrund der „Audioeffekte“ in der Simulation schwierig zu lokalisieren, insbesondere aufgrund der Abweichung von der realen Hallcharakteristik.

Richtungslokalisierung und Kopfbewegungen:

- Die Lokalisation der Richtung, aus der die Stimme kommt, ist einfach, besonders wenn der Kopf gedreht wird.
- Es wird angemerkt, dass die Stimme wie aus einem Lautsprecher klingt.

Entfernungslokalisierung:

- Die Einschätzung der Entfernungen, insbesondere weit und nah, gestaltet sich als herausfordernd.
- Die maximale Distanz der Stimme ist am schwierigsten einzuschätzen, besonders aufgrund der Testumgebung.

Raumhall und Hintergrundgeräusche:

- Der Hall wird als unnatürlich empfunden, insbesondere für den Raum.
- Hintergrundgeräusche der sprechenden Personen werden wahrgenommen, jedoch nicht als allzu störend.

Schwierigkeiten und Unsicherheiten:

- Einige Testpersonen hatten Schwierigkeiten, die Distanz der Stimme konkret einzuschätzen.
- Die Raumtiefe der Stimme wird als etwas schwer zu beurteilen wahrgenommen.
- Es gibt Unsicherheiten bei der Einschätzung der Entfernung der Stimme.

Insgesamt zeigen die Rückmeldungen eine Vielfalt an Empfindungen hinsichtlich der Lokalisationsgenauigkeit, vor allem in Bezug auf Entfernungen und Raumcharakteristiken. Es scheint, dass die Richtungslokalisierung positiv bewertet wird im Vergleich zur Entfernungslokalisierung und der Wahrnehmung des Raumes.

### 4.3.1.4 Steam

Die Rückmeldungen der Testpersonen zum Steam Spatializer lassen folgende wesentliche Punkte erkennen:

Raumhall und Lokalisationsdefizite:

- Die Stimme wird als sehr hallig wahrgenommen.
- Die Position und Entfernung können bei einigen Klangquellen nicht bestimmt werden.
- Es wird angemerkt, dass die Stimme eher "überall und nirgends" und zentral im Kopf wahrgenommen wird.

Akustik:

- Die Akustik wird mit einem Stahlcontainer oder Raumschiff verglichen und als nicht realistisch empfunden.

Herausforderungen bei der Richtungserkennung:

- Die Richtungserkennung wird als schwieriger empfunden, besonders zu Beginn des Tests.
- Links/Rechts-Wahrnehmung (Azimuth) gestaltet sich als schwieriger im Vergleich zu vorherigen Tests.

Verzerrungen und Klangcharakteristiken:

- Die Stimme wird als "blechig", "metallen", „boxy“ und verzerrt beschrieben.

Orientierungsprobleme und Gewöhnung:

- Zu Beginn wird eine "komplette Orientierungslosigkeit" erlebt, jedoch fällt die Lokalisierung mit der Zeit leichter.
- Die Stimme wird als dumpfer wahrgenommen als zuvor, was die Lokalisierung erschwert.

Insgesamt zeigen die Rückmeldungen eine Vielzahl von Herausforderungen, vor allem im Hinblick auf den Hall, den Klangcharakter und die Schwierigkeiten bei der Richtungserkennung. Es scheint, dass die Akustik des Steam Spatializers als weniger realistisch empfunden wird im Vergleich zu den anderen Spatializern.

### **4.3.2 Lokalisierbarkeit auf Distanz**

Einige Testpersonen erwähnen im Online-Fragebogen, dass es einfacher sei eine Klangquelle die nah liegt zu verorten. Daher wird bei der Datentransformation in Unity zusätzlich die durchschnittliche Distanz- und Azimuthdifferenz in Abhängigkeit zur Entfernung des Ursprungs in fern ( $>5\text{m}$ ) und nah ( $\leq 5\text{m}$ ) berechnet.

4.3.2.1 Distanzdifferenz

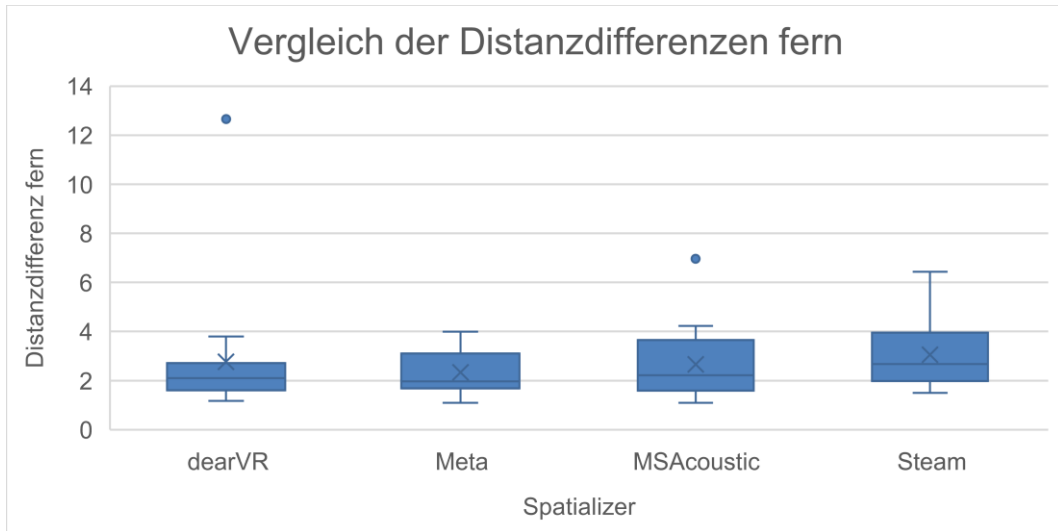


Abbildung 59: Boxplot – Vergleich der Distanzdifferenzen fern

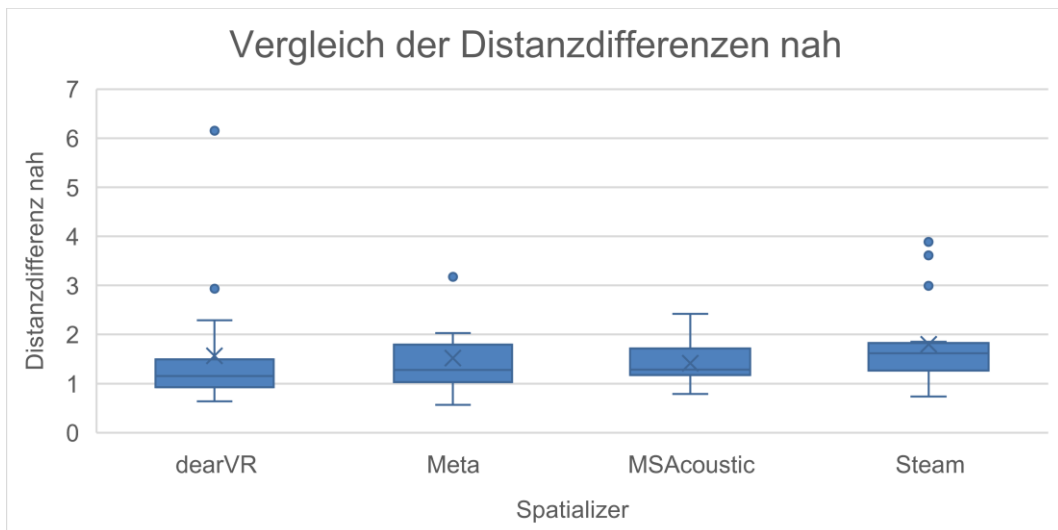


Abbildung 60: Boxplot – Vergleich der Distanzdifferenzen nah

Spatializer	Nahe Klangquellen		Ferne Klangquellen	
	M	SD	M	SD
DearVR	1,56m	1,31m	2,76m	2,64m
Meta	1,52m	0,74m	2,33m	0,91m
MSAcoustics	1,41m	0,41m	2,66m	1,49m
Steam	1,80m	0,88m	3,05m	1,36m

Tabelle 3: Mittelwerte (M) und Standardabweichung (SD) der nahen und fernen Distanzdifferenzen pro Audio Spatializer

## 4 Evaluation

In allen Spatializern tendieren die Distanzwerte bei fernen Klangquellen dazu, höher zu sein als bei nahen Klangquellen. DearVR zeigt insgesamt die geringste Distanz bei nahen Klangquellen. Steam hat den höchsten Mittelwert für Distanzen bei fernen Klangquellen. Die Standardabweichungen variieren, wobei Steam die höchsten Abweichungen sowohl für nahe als auch ferne Klangquellen aufweist.

### 4.3.2.2 Azimuttdifferenz

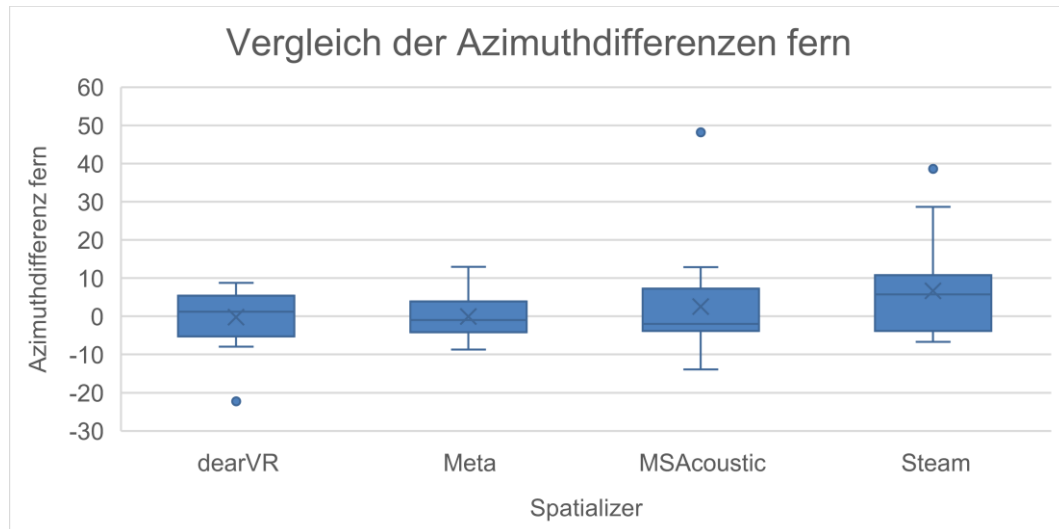


Abbildung 61: Boxplot – Vergleich der Azimuttdifferenzen fern

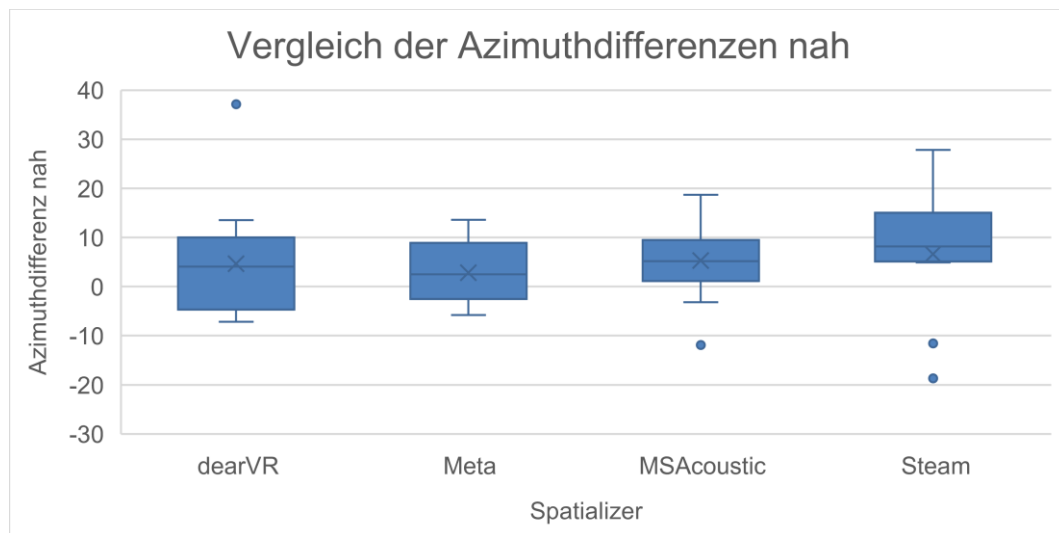


Abbildung 62: Boxplot - Vergleich der Azimuttdifferenzen nah

## 4 Evaluation

---

Spatializer	Nahe Klangquellen		Ferne Klangquellen	
	M	SD	M	SD
DearVR	4,66°	10,90°	-0,19°	7,74°
Meta	2,80°	6,49°	-0,02°	6,13°
MSAcoustics	5,33°	7,21°	2,57°	13,71°
Steam	6,64°	12,62°	6,61°	12,26°

*Tabelle 4: Mittelwerte (M) und Standardabweichung (SD) der nahen und fernen Azimuthdifferenzen pro Audio Spatializer*

Die Azimuthdifferenzen bei fernen Klangquellen neigen dazu, näher an 0 zu liegen. Der Steam Spatializer zeigt den höchsten Mittelwert für die Azimuthdifferenzen bei fernen Klangquellen, während DearVR den niedrigsten aufweist. Bei nahen Klangquellen variieren die Azimuthdifferenzen stärker, insbesondere bei MSAcoustics. Die Standardabweichungen sind bei allen Spatializern höher für nahe Klangquellen im Vergleich zu fernen Klangquellen.

## 5 Fazit

In dieser Arbeit wurde die Lokalisierbarkeit virtueller Klangquellen in einer Mixed Reality-Umgebung mithilfe verschiedener Audio Spatializer untersucht. Die vorgestellten Audio Spatializer bieten unterschiedliche Features und Arbeitsweisen, die es den Entwickler:innen und Sound Designer:innen ermöglichen, die gewünschte Klangkulisse in ihren Mixed Reality-Anwendungen umzusetzen. Es ist wichtig, die individuellen Anforderungen und Präferenzen bei der Auswahl eines Audio Spatializers zu berücksichtigen. Durch einen Hörtest wurden die Präzision und die Eignung der Spatializer für die Lokalisierung von Klangquellen bewertet.

Die Ergebnisse zeigen, dass die Auswahl des Spatializers keinen signifikanten Einfluss auf die Lokalisierbarkeit hat. Dennoch sind einige Spatializer, wie Meta und DearVR, tendenziell präziser in der Lokalisierung als im Vergleich mit Microsoft Acoustics und Steam Audio. Die Ergebnisse legen nahe, dass zwei der aktuell verfügbaren Audio Spatializer, Meta und DearVR, ohne zusätzliche Hard- oder Software für die Lokalisierung virtueller Klangquellen in Mixed Reality-Anwendungen geeignet sind, da sie die in dieser Arbeit definierten Toleranzbereiche für die Distanzdifferenz und Azimuthdifferenz einhalten. Diese Erkenntnisse können Entwickler:innen helfen, den passenden Audio Spatializer für ihre Anwendungen auszuwählen und so realistische und immersive Mixed Reality-Erlebnisse zu schaffen.

Dieses Ergebnis spiegelt auch die Meinung der Testpersonen wider, welche diese über den Online-Fragebogen kundtaten. Es zeigt sich eine sehr starke Tendenz zu den Spatializern Meta und DearVR, wobei vor allem der Steam Spatializer aufgrund der verzerrten Wiedergabe der Klangquellen nicht nur durch die subjektive Bewertung schlecht abschnitt.

Bei dem durchgeführten Hörtest konnte dank des Feedbacks einiger Testteilnehmer:innen rechtzeitig wichtige Anpassungen am Hörtest vorgenommen werden. Dies bezieht sich vor allem auf die Implementierung der Mixed-Reality Applikationen. Durch die automatisierte Erkennung des Spatial Anchor im Raum nach dem erneuten Aufsetzen der Meta Quest 3 konnte so die Datenvalidität gewährleistet werden. Um komplizierte Neuberechnung der Positionierungen der einzelnen Marker und Ursprünge zu vermeiden, sollte zu einem Datenset stets auch der Nullpunkt des virtuellen Raumes in der Unity Szene gespeichert werden.

Kritisch aufgefallen und von den Testpersonen häufig erwähnt ist auch die ungleiche Lautstärke zwischen den Testapplikationen. Diese sollte bei neuerlicher Durchführung angeglichen sein.

Die vier Spatializer haben unterschiedliche Auralisierungsmethoden. Der DearVR Spatializer beispielsweise nutzt die automatische Raumanalyse und ein von dem:der Sound Designer:in gewähltes Preset zur Erzeugung des Raumhalls. Der Steam und auch der MSAcoustics Spatializer hingegen berechnen diesen vollautomatisch. Lediglich durch die Anpassungen der Materialien im Raum kann eine klangliche Veränderung des Raumhalls angepasst werden. Interessanterweise sind genau diese zwei Audio Spatializer am schlechtesten bemessen und von den Testpersonen bewertet worden. Dies könnte darauf schließen lassen, dass die automatisierten Auralisierungsmethoden noch nicht stark genug, sind um einen:eine Sound Designer:in zu ersetzen.

Abschließend lässt sich sagen, dass die vorliegende Arbeit einen Beitrag zum Verständnis und zur Auswahl von Audio Spatializern für die Lokalisierung von virtuellen Klangquellen in Mixed Reality-Anwendungen leistet. Es gibt jedoch weiterhin Raum für zukünftige Forschung und Verbesserungen der Methodik, um die Entwicklung von noch realistischeren und immersiveren Mixed Reality-Erlebnissen zu unterstützen.

## 5.1 Ausblick

Eine mögliche Erweiterung der bisherigen Hörtests besteht darin, die Lokalisierung in der vertikalen Hörebene zu erfassen und die y-Achse in die Analyse einzubeziehen. Dies ermöglicht eine umfassendere Untersuchung der Lokalisierungsfähigkeiten der Audio Spatializer in allen drei Raumdimensionen. Durch die Erweiterung der Datenerhebung um die vertikale Position können genauere Rückschlüsse auf die Präzision und Zuverlässigkeit der Lokalisierung gezogen werden.

Zusätzlich sollte auf eine gute Abhörmöglichkeit Wert gelegt werden. Die verwendeten Kopfhörer sollten geschlossen sein, um die akustischen Eindrücke der Audio Spatializer nicht mit jenen der realen Welt zu vermischen. Dadurch könnte eine noch bessere Leistung in der Lokalisierung gewährleistet werden.

Ein weiterer vielversprechender Forschungsbereich liegt im Sound Design und den Möglichkeiten der Audio Spatializer in Bezug auf die Gestaltung von Klängen und die Schaffung immersiver Erfahrungen. Hierbei können verschiedene

## 5 Fazit

---

Parameter und Techniken untersucht werden, um die Wirkung und Realitätsnähe der virtuellen Klangquellen weiter zu optimieren. Eine detaillierte Analyse der Auswirkungen von Klangparametern wie Frequenzspektrum, Klangfarbe und Raumreflexionen auf die Lokalisierbarkeit kann dazu beitragen, das Sound Design in Mixed Reality-Anwendungen zu verfeinern.

Darüber hinaus bietet die Untersuchung von Augmented Reality-Anwendungen, bei denen die reale Umgebung automatisch von einem Mixed Reality-Headset erkannt wird, ein interessantes Forschungsgebiet. Hierbei steht die Frage im Mittelpunkt, ob die Audio Spatializer mit dieser stetig wechselnden Umgebung umgehen können und weiterhin eine präzise Lokalisierung ermöglichen. Die Analyse der Leistungsfähigkeit der Spatializer in solchen dynamischen Umgebungen kann wichtige Erkenntnisse liefern und zur Verbesserung der Mixed Reality-Erfahrungen beitragen.

Weiters ist die Frage nach der automatischen Auralisierungsmethode beim Hörtest von Interesse geworden. Eine Hypothese könnte sein, dass ein vollautomatisch generierter Raumhall schlechter abschneidet als ein in den Parametern des Spatializers an den echten Raum angepasster Raumhall. Der Hörtest dieser Masterarbeit lässt darauf schließen. Eine weitere These wäre noch, ob Audio Spatializer in Zukunft stark genug sind, um Sound Designer für diese Aufgabe der Erzeugung eines realitätsnahen Raumhalls abzulösen.

Zusammenfassend bieten diese vorgeschlagenen Forschungsrichtungen vielversprechende Möglichkeiten, um die Lokalisierung virtueller Klangquellen in Mixed Reality weiterzuentwickeln und zu optimieren. Die Untersuchung der Lokalisierung in der vertikalen Hörebene, die Verfeinerung des Sound Designs sowie die Analyse der Leistungsfähigkeit in dynamischen Umgebungen können dazu beitragen, die Immersion und Realitätsnähe von Mixed Reality-Erlebnissen zu steigern.

# Literaturverzeichnis

*3D Audio Spatialization | Oculus Developers.* (o. J.). Abgerufen 9. Januar 2024, von <https://developer.oculus.com/resources/audio-intro-spatialization/>

*Apply Room Acoustics in Unity: Unity | Oculus Developers.* (o. J.). Abgerufen 15. November 2023, von <https://developer.oculus.com/documentation/unity/meta-xr-audio-sdk-unity-room-acoustics/>

*Apply Spatialization in Unity: Unity | Oculus Developers.* (o. J.). Abgerufen 15. November 2023, von <https://developer.oculus.com/documentation/unity/meta-xr-audio-sdk-unity-spatialize/>

Begault, D. R. (1994). *3-D sound for virtual reality and multimedia*. AP Professional.

Begault, D. R., Lee, A. S., Wenzel, E. M., & Anderson, M. R. (2000, Februar 1). *Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source*. Audio Engineering Society Convention 108. <https://www.aes.org/e-lib/browse.cfm?elib=9204>

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization* (2. Aufl.). The MIT Press. <https://doi.org/10.7551/mitpress/6391.001.0001>

*Cross-platform mixed reality development on Meta Quest 3.* (o. J.). Unity Blog. Abgerufen 7. November 2023, von <https://blog.unity.com/engine-platform/cross-platform-mixed-reality-development-on-meta-quest-3>

Dear Reality. (2020). *DearVR Unity—User Manual v.1.6*.

Firat, H. B., Maffei, L., & Masullo, M. (2022). 3D sound spatialization with game engines: The virtual acoustics performance of a game engine and a middleware for interactive audio design. *Virtual Reality*, 26(2), 539–558. <https://doi.org/10.1007/s10055-021-00589-0>

*FMOD.* (o. J.). Abgerufen 3. Juni 2023, von <https://www.fmod.com/>

Garner, T. A. (2018). *Echoes of Other Worlds: Sound in Virtual Reality: Past, Present and Future* (1st ed. 2018). Springer International Publishing: Imprint: Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-65708-0>

*General information on SOFA - Sofaconventions.* (o. J.). Abgerufen 9. Juni 2023, von [https://www.sofaconventions.org/mediawiki/index.php/General\\_information\\_on\\_SOFA](https://www.sofaconventions.org/mediawiki/index.php/General_information_on_SOFA)

Genuit, K. (2018). *Standardization of Soundscape: Request of Binaural Recording. Conference Proceedings.*

*Get Started with Passthrough | Oculus Developers.* (o. J.). Abgerufen 13. November 2023, von <https://developer.oculus.com/documentation/unity/unity-passthrough-gs/>

*Getting Started—Steam Audio Unity Integration documentation.* (o. J.). Abgerufen 16. November 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/getting-started.html>

Gildersleeve, B. (2023, Februar 7). *Build Immersive Audio Experiences with Audio SDK.* Meta Developer Blog. <https://developer.oculus.com/presence-platform/>

Görne, T. (2017). *Sounddesign: Klang, Wahrnehmung, Emotion.* Hanser.

*Installing the Universal Render Pipeline into an existing Project | Universal RP | 17.0.1.* (o. J.). Abgerufen 12. November 2023, von <https://docs.unity3d.com/Packages/com.unity.render-pipelines.universal@17.0/manual/InstallURPIntoAProject.html>

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), 1515–1529. <https://doi.org/10.1007/s10639-017-9676-0>

kevinasg. (2022a, Oktober 10). *Project Acoustics Unity Baking—Bake Step—Microsoft Game Dev.* <https://learn.microsoft.com/en-us/gaming/acoustics/unity-baking-bake>

kevinasg. (2022b, Dezember 5). *Project Acoustics Unity Integration and Deployment—Microsoft Game Dev.* <https://learn.microsoft.com/en-us/gaming/acoustics/unity-integration>

kevinasg. (2023, Februar 3). *Project Acoustics Overview—Microsoft Game Dev.* <https://learn.microsoft.com/en-us/gaming/acoustics/what-is-acoustics>

Liaquat, M. U., Munawar, H. S., Rahman, A., Qadir, Z., Kouzani, A. Z., & Mahmud, M. A. P. (2021). Localization of Sound Sources: A Systematic Review. *Energies*, 14(13), 3910. <https://doi.org/10.3390/en14133910>

*Listening Devices | Oculus Developers.* (o. J.). Abgerufen 4. Juni 2023, von <https://developer.oculus.com/resources/audio-intro-devices/>

McGill, M., Brewster, S., McGookin, D., & Wilson, G. (2020). Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376702>

*Meta XR Audio SDK Features: Unity | Oculus Developers*. (o. J.). Abgerufen 4. Juni 2023, von [https://developer.oculus.com/documentation/unity/meta-xr-audio-sdk-features/?intern\\_source=devblog&intern\\_content=build-immersive-audio-experiences-meta-quest-sdk](https://developer.oculus.com/documentation/unity/meta-xr-audio-sdk-features/?intern_source=devblog&intern_content=build-immersive-audio-experiences-meta-quest-sdk)

*Meta XR Audio SDK Plugin for Unity—Requirements and Setup: Unity | Oculus Developers*. (o. J.). Abgerufen 15. November 2023, von <https://developer.oculus.com/documentation/unity/meta-xr-audio-sdk-unity-req-setup/>

Middlebrooks, J. C. (2015). Sound localization. In *Handbook of Clinical Neurology* (Bd. 129, S. 99–116). Elsevier. <https://doi.org/10.1016/B978-0-444-62630-1.00006-8>

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: A class of displays on the reality-virtuality continuum* (H. Das, Hrsg.; S. 282–292). <https://doi.org/10.1117/12.197321>

*Mixed Reality Market Report | Size, Share, Growth & Trends (2022-27)*. (o. J.). Abgerufen 22. Januar 2023, von <https://www.mordorintelligence.com/industry-reports/mixed-reality-market>

Mizutani, W. K. (2017). *VORPAL: A middleware for real-time soundtracks in digital games* [Mestrado em Ciência da Computação, Universidade de São Paulo]. <https://doi.org/10.11606/D.45.2017.tde-18052017-112539>

Neidhardt, A., Schneiderwind, C., & Klein, F. (2022). Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms—Literature Review and Theoretical Framework. *Trends in Hearing*, 26, 23312165221092919. <https://doi.org/10.1177/23312165221092919>

*Oculus-Entwicklerzentrum | Integration der Unity-Spiele-Engine für VR*. (o. J.). Abgerufen 12. November 2023, von <https://developer.oculus.com/unity/>

Project Triton—Immersive sound propagation. (o. J.). *Microsoft Research*. Abgerufen 10. Juni 2023, von <https://www.microsoft.com/en-us/research/project/project-triton/>

Pulkki, V. (2001). *Spatial sound generation and perception by amplitude panning techniques*. Helsinki University of Technology.

Roginska, A., & Geluso, P. (Hrsg.). (2018). *Immersive sound: The art and science of binaural and multi-channel audio*. Routledge, Taylor & Francis Group.

Rossi, S., Viola, I., Toni, L., & Cesar, P. (2021). A New Challenge: Behavioural Analysis Of 6-DOF User When Consuming Immersive Media. *2021 IEEE International Conference on Image Processing (ICIP)*, 3423–3427. <https://doi.org/10.1109/ICIP42928.2021.9506525>

Schütze, S. (2018). *Virtual sound: A practical guide to audio, dialogue and music in VR and AR*. Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc.

*Simple Gaze Cursor*. (o. J.). Immersive Limit. Abgerufen 13. November 2023, von <https://www.immersivelimit.com/simple-gaze-cursor-old>

*Simulating Directivity Using Cone-Shaped Boundaries*. (o. J.). Abgerufen 10. Juni 2023, von [https://www.audiokinetic.com/en/library/edge/?source=Help&id=simulating\\_directivity\\_using\\_cone\\_shaped\\_boundaries](https://www.audiokinetic.com/en/library/edge/?source=Help&id=simulating_directivity_using_cone_shaped_boundaries)

*Spatial Anchors Overview: Unity | Oculus Developers*. (o. J.). Abgerufen 13. November 2023, von <https://developer.oculus.com/documentation/unity/unity-spatial-anchors-overview/>

*Steam Audio Dynamic Object—Steam Audio Unity Integration documentation*. (o. J.). Abgerufen 16. November 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/dynamic-object.html>

*Steam Audio Geometry—Steam Audio Unity Integration documentation*. (o. J.). Abgerufen 16. November 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/geometry.html>

*Steam Audio Source—Steam Audio Unity Integration documentation*. (o. J.). Abgerufen 16. November 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/source.html>

*Steam Audio Unity Integration—Steam Audio Unity Integration documentation*. (o. J.). Abgerufen 9. Juni 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/index.html>

Stirling, P. (o. J.). *Volumetric Sounds*. Abgerufen 4. Juni 2023, von <https://developer.oculus.com/blog/volumetric-sounds>

Technologies, U. (o. J.-a). *Unity - Manual: Audio Spatializer SDK*. Abgerufen 30. März 2023, von <https://docs.unity3d.com/Manual/AudioSpatializerSDK.html>

Technologies, U. (o. J.-b). *Unity - Manual: Prefabs*. Abgerufen 8. November 2023, von <https://docs.unity3d.com/Manual/Prefabs.html>

Toftedahl, M., & Engström, H. (2019). *A Taxonomy of Game Engines and the Tools that Drive the Industry*.

*Unity Canvas Integration | Oculus Developers*. (o. J.). Abgerufen 13. November 2023, von <https://developer.oculus.com/documentation/unity/unity-isdk-canvas-integration/>

*Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine*. (o. J.). Unity. Abgerufen 3. Juni 2023, von <https://unity.com>

*Unreal Engine | Das stärkste Werkzeug für 3D-Echtzeit-Entwicklung*. (o. J.). Unreal Engine. Abgerufen 3. Juni 2023, von <https://www.unrealengine.com/de>

*Use Shared Spatial Anchors: Unity | Oculus Developers*. (o. J.). Abgerufen 12. Januar 2024, von <https://developer.oculus.com/documentation/unity/unity-shared-spatial-anchors/>

*User's Guide—Steam Audio Unity Integration documentation*. (o. J.). Abgerufen 16. November 2023, von <https://valvesoftware.github.io/steam-audio/doc/unity/guide.html>

Villegas, J. (2015). Locating virtual sound sources at arbitrary distances in real-time binaural reproduction. *Virtual Reality*, 19(3–4), 201–212. <https://doi.org/10.1007/s10055-015-0278-0>

Vorländer, M. (2016). *From acoustic simulation to virtual auditory displays*.

Wefers, F. (2015). *Partitioned convolution algorithms for real-time auralization*. Logos Verlag Berlin GmbH.

Weinzierl, S. (Hrsg.). (2008). *Handbuch der Audiotechnik* (Ausgabe in 2 Bänden). Springer.

Werner, S., Klein, F., Neidhardt, A., Sloma, U., Schneiderwind, C., & Brandenburg, K. (2021). Creation of Auditory Augmented Reality Using a Position-Dynamic Binaural Synthesis System—Technical Components, Psychoacoustic Needs, and Perceptual Evaluation. *Applied Sciences*, 11(3), 1150. <https://doi.org/10.3390/app11031150>

*Wwise | Audiokinetic*. (o. J.). Abgerufen 3. Juni 2023, von <https://www.audiokinetic.com/en/products/wwise/>

# Abbildungsverzeichnis

Abbildung 1: Das kopfbezogene Koordinatensystem nach Blauert (Blauert, 1996, S. 11).....	15
Abbildung 2: Blauert beschreibt die Position der einfallenden Klangquelle als „auditory event“ und die wahrgenommene Richtung als „sound event“. Hier die Ortsunschärfe für die horizontale Hörebene. (Blauert, 1996, S. 41) .....	21
Abbildung 3: Ortsunschärfe auf vertikaler Hörebene (Blauert, 1996, S. 44) .....	22
Abbildung 4: Ortsunschärfe bei der Distanzerkennung(Blauert, 1996, S. 47).....	23
Abbildung 5: Ein Querschnitt des Ohres. „1: Semicircular canals. 2: Cochlea. 3: Eardrum-tensioning muscles. 4: Eustachian tube. 5: Cavum conchae. 6: External auditory canal. 7: Eardrum. 8: Hammer.“ (Blauert, 1996, S. 52)....	24
Abbildung 6: Die Blöcke Klanggeneration, Ausbreitung und Wiedergabe der Auralisation nach Vorländer (Vorländer, 2016, S. 3) .....	30
Abbildung 7: Das Konzept des "perceptual matching" von Raumakustik für AAR umfasst: a) die originalgetreue Wiedergabe einer Klangquelle in einem Raum, b) einen physisch perfekt abgestimmten virtuellen Raum, welche in der Praxis jedoch schwer umzusetzen ist, c) einen physisch geringfügig unterschiedlichen, aber wahrnehmungsmäßig angepassten Raum, d) die Erwartungen des Zuhörers, die möglicherweise von dem ursprünglichen Klangfeld abweichen könnte. (Neidhardt et al., 2022, S. 3).....	35
Abbildung 8: Schallwellen "durch die Wand" welche aus der Quellrichtung gehört werden sind inkorrekt. Offene Portale (Türen, Fenster,...) leiten den klang realistisch um. („Project Triton - Immersive Sound Propagation“, o. J.).....	49
Abbildung 9: Ablauf einer Testapplikation .....	57
Abbildung 10: Platzieren eines Markers (grün) während eines Testdurchlaufs durch die Testperson. Der Passthrough-Modus ist aktiv. ....	58
Abbildung 11: Installation des Oculus XR Plugins.....	61
Abbildung 12: Konfiguration der OVR Manager Komponente am OVR Camera Rig GameObject. ....	62
Abbildung 13: Platzierung des Motion Capturing Labs 3D Models in der Unity Szene .....	63
Abbildung 14: Positionen der zu verortenden Samples im 3D Model .....	64

Abbildung 15: OVRCameraRig mit integrierten OVRController und RayInteractor .....	65
Abbildung 16: Flussdiagramm der Hörtest Applikation .....	66
Abbildung 17: Binauralaufnahmen für 8 Monitorpositionen im virtuellen MoCapLab .....	71
Abbildung 18: Kunstkopfaufnahme Position 3 mit Blickrichtung vorne. ....	71
Abbildung 19: Setup für die raumakustische Messung mit der Software REW... ..	72
Abbildung 20: RT60 mit T20 in rot und T30 in orange. ....	73
Abbildung 21: Routing des Audiosignals von REW nach Unity und wieder zurück mithilfe von VoiceMeeter Banana. ....	75
Abbildung 22: ASIO Routing in REW .....	76
Abbildung 23: VoiceMeeter Banana Konfiguration .....	76
Abbildung 24: Meta XR Audio Spatializer Plugin Auswahl in den Unity Project Settings .....	77
Abbildung 25: Parameter der MetaXRAudioSource und MetaXRAudioSourceExperimentalFeatures.....	78
Abbildung 26: Einstellungen des Meta XR Audio Reflection Effekts.....	79
Abbildung 27: Parameter der MetaXRAudioRoomAcousticProperties Komponente auf dem RoomAcoustics GameObject. ....	80
Abbildung 28: Messeinstellungen in REW für den MetaXR Spatializer .....	80
Abbildung 29: RT60 des MetaXR Spatializers im virtuellen Motion Capturing Lab. ....	81
Abbildung 30: Automatische Raumanalyse mit dem DearVRManager.....	82
Abbildung 31: DearVRSource Parameter .....	83
Abbildung 32: Das DearVR Reverb Plugin.....	83
Abbildung 33: RT60 des Reverbpreset „Room_Large“ mit dem DearVR Spatializer. ....	84
Abbildung 34: RT60 mit Reverbpreset „Recording_Hall_Medium“ mit dem DearVR Spatializer.....	84
Abbildung 35: SteamAudioSource Parameter .....	85

Abbildung 36: SteamAudioListener zur Aktivierung des Reverbs mit dem Steam Audio Spatializer.....	86
Abbildung 37: RT60 mit dem Steam Audio Spatializer .....	87
Abbildung 38: Acoustics Material Tab mit zugewiesenen Absorptionsgraden pro Material in der Szene.....	89
Abbildung 39: Berechnete Probe Points (blaue Würfel) und Voxels (grüne Linien) innerhalb der Unity Szene.....	89
Abbildung 40: RT60 des virtuellen Motion Capturing Labs mit dem Microsoft Acoustics Spatializer.....	90
Abbildung 41: Einführung des Online-Fragebogen.....	91
Abbildung 42: Fragebogen zu der Hörtestapplikation.....	92
Abbildung 43: Hörtest mit vorheriger Lokalisierungsprüfung .....	94
Abbildung 44: Testteilnehmer bei der Ausführung des Hörtest mit einer der Hörtest Applikationen.....	95
Abbildung 45: Möglicher Nullpunkt, Ursprung und zwei Marker in einem Koordinatensystem.....	96
Abbildung 46: Ordner zu den abgelegten .json Dateien des Hörtests .....	97
Abbildung 47: Hörtestergebnisse mit Teilnehmeridentifikation für die MSAcoustic Applikation.....	98
Abbildung 48: Veranschaulichung der Berechnung des Punktes P2“ durch Transformation des Punktes P2 in GeoGebra.....	99
Abbildung 49: Streudiagramm der durchschnittlichen Markerpositionen des DearVR Spatializers .....	105
Abbildung 50: Streudiagramm der durchschnittlichen Markerpositionen des Steam Spatializers.....	106
Abbildung 51: Streudiagramm der durchschnittlichen Markerpositionen des MSAcoustics Spatializers.....	107
Abbildung 52: Streudiagramm der durchschnittlichen Markerpositionen des Meta Spatializers.....	108
Abbildung 53: Boxplot - Vergleich der Distanzdifferenz zwischen Audio Spatializer .....	109

Abbildung 54: Boxplot - Vergleich der Azimuthdifferenz zwischen Audio Spatializer .....	110
Abbildung 55: Gruppiertes Balkendiagramm der Mittelwerte des Online- Fragebogens nach Spatializern .....	115
Abbildung 56: Boxplot – Vergleich der Lokalisationsgenauigkeit. Ein hoher Wert bedeutet eine hohe Lokalisationsgenauigkeit (Selbsteinschätzung).....	116
Abbildung 57: Boxplot – Vergleich der Natürlichkeit der Stimme. Ein hoher Wert bedeutet eine hohe Natürlichkeit.....	117
Abbildung 58: Boxplot – Vergleich der Lokalisationsschwierigkeit. Ein hoher Wert bedeutet eine einfache Lokalisierbarkeit. ....	117
Abbildung 59: Boxplot – Vergleich der Distanzdifferenzen fern .....	122
Abbildung 60: Boxplot – Vergleich der Distanzdifferenzen nah .....	122
Abbildung 61: Boxplot – Vergleich der Azimuthdifferenzen fern .....	123
Abbildung 62: Boxplot - Vergleich der Azimuthdifferenzen nah .....	123

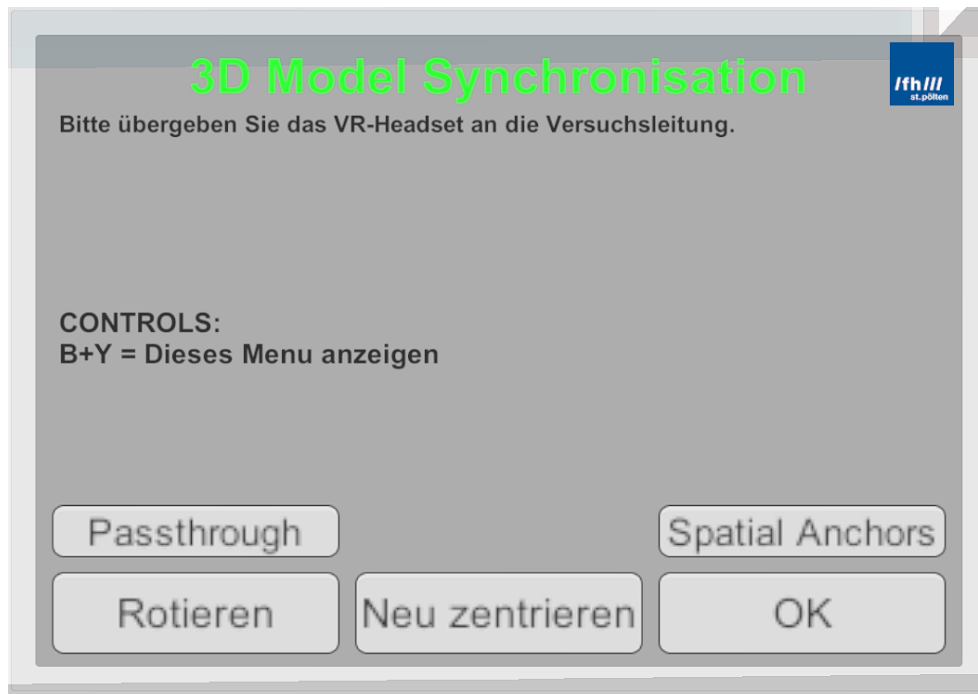
# Tabellenverzeichnis

Tabelle 1: Zeitpunkte und zugehörige TeilnehmerIDs der Hörtests .....	95
Tabelle 2: Übersicht der Mittelwerte der Distanzdifferenzen und Azimuthdifferenzen und Einordnung in Toleranzbereiche .....	113
Tabelle 3: Mittelwerte (M) und Standardabweichung (SD) der nahen und fernen Distanzdifferenzen pro Audio Spatializer .....	122
Tabelle 4: Mittelwerte (M) und Standardabweichung (SD) der nahen und fernen Azimuthdifferenzen pro Audio Spatializer .....	124

# Anhang

## UI Screenshots

### Model Synchronisation




### Spatial Anchor Panel



## Listening Test Introduction Panel

### Hörtest Einführung



Willkommen und vielen Dank, dass Sie bei diesem Hörtest teilnehmen!

Sie können mit dem Controller in ihrer Hand mit dem Programm interagieren. Zwei Buttons sind zur Durchführung des Tests relevant: Der Trigger, welcher sich auf der Hinterseite des Controllers befindet und der A Button.

Im Nachfolgenden werden Sie gebeten 21 Klangquellen zu verorten. Zielen sie mit dem Controller auf den Boden, wo Sie die Klangquelle vermuten und setzen Sie den Marker mit dem Trigger Button. Sie können den Marker durch Gedrückt halten des Triggers weiter verschieben. Bestätigen Sie die gewählte Position mit dem A Button um mit dem nächsten Klangsample fortzufahren.


Sollten Sie diesen Hörtest zum ersten Mal durchführen, wählen sie bitte die Schaltfläche "Warm-Up" mit den Motion Controllern. Bestätigen Sie ihre Wahl mit dem Trigger.

Sollten Sie bereits mit dem Hörtest vertraut sein, wählen sie "Hörtest".  
Bei Fragen können Sie sich stets an die Versuchsleitung wenden!

**Warm-Up** **Hörtest**

## Warm-Up Panel 1

### Warm-Up




Sie hören nun eine virtuelle Klangquelle. Verorten Sie die Stimme und zielen Sie mit dem Controller auf die Position, an der Sie die Klangquelle vermuten.

Setzen Sie den Marker mit dem Trigger des Controllers.

## Warm-Up Panel 2

**Warm-Up**




**Sehr gut. Sollten Sie sich umentscheiden, können Sie den Marker mit gedrücktem Trigger an den gewünschten Ort verschieben.**

**Versuchen Sie das Verschieben des Markers zum Fortfahren.**

## Warm-Up Panel 3

**Warm-Up**




**Wunderbar.**

**Bestätigen Sie ihre Wahl nun mit dem A Button auf dem Controller.**

**Die nächste Klangquelle wird sofort ertönen.**

#### Warm-Up Panel 4

## Warm-Up



Führen Sie das Setzen und Bestätigen des Markers mit einigen Klangquellen so oft durch, bis Sie sich sicher mit dem System fühlen.

Klicken Sie auf den Button “Zurück”, um zum Einstieg zurückzukehren.

#### Hörtest Panel

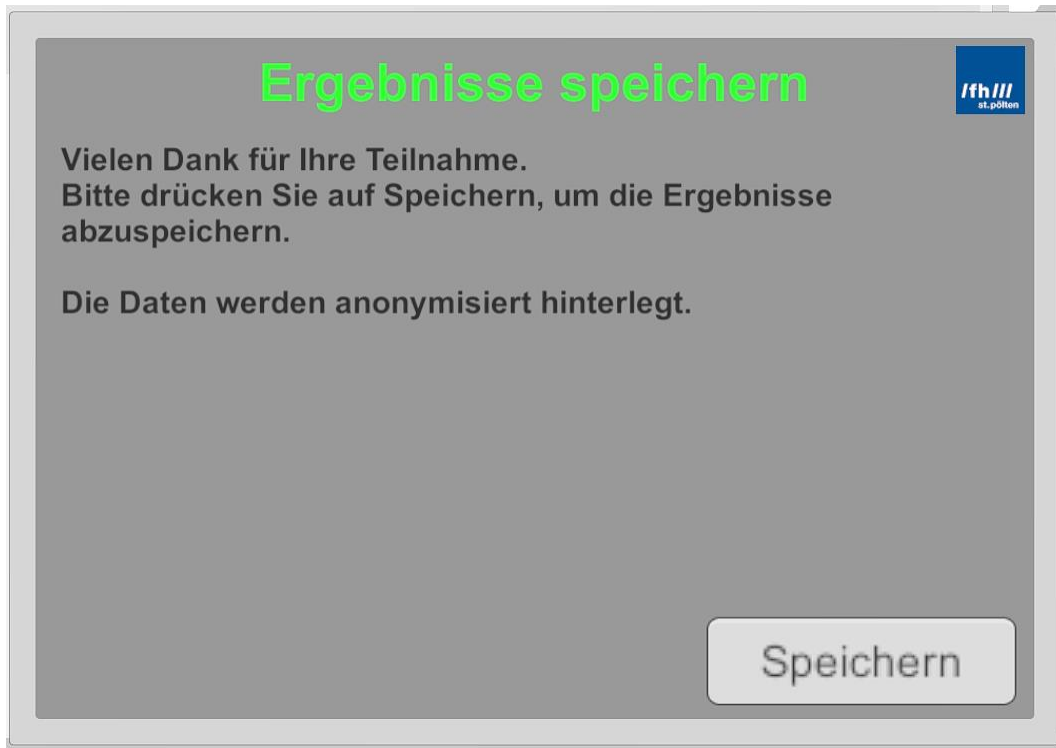
## Hörtest



Es folgen nun 21 Klangsamples. Setzen Sie die Marker zu den geschätzten Positionen mit dem Trigger und bestätigen Sie mit dem A Button die Position, um die nächste Klangquelle zu verorten. Die Ergebnisse werden am Ende gespeichert.

Bei Fragen wenden Sie sich bitte an die Versuchsleitung. Klicken Sie auf “Start”, um mit dem Hörtest zu beginnen.

## Ende Panel



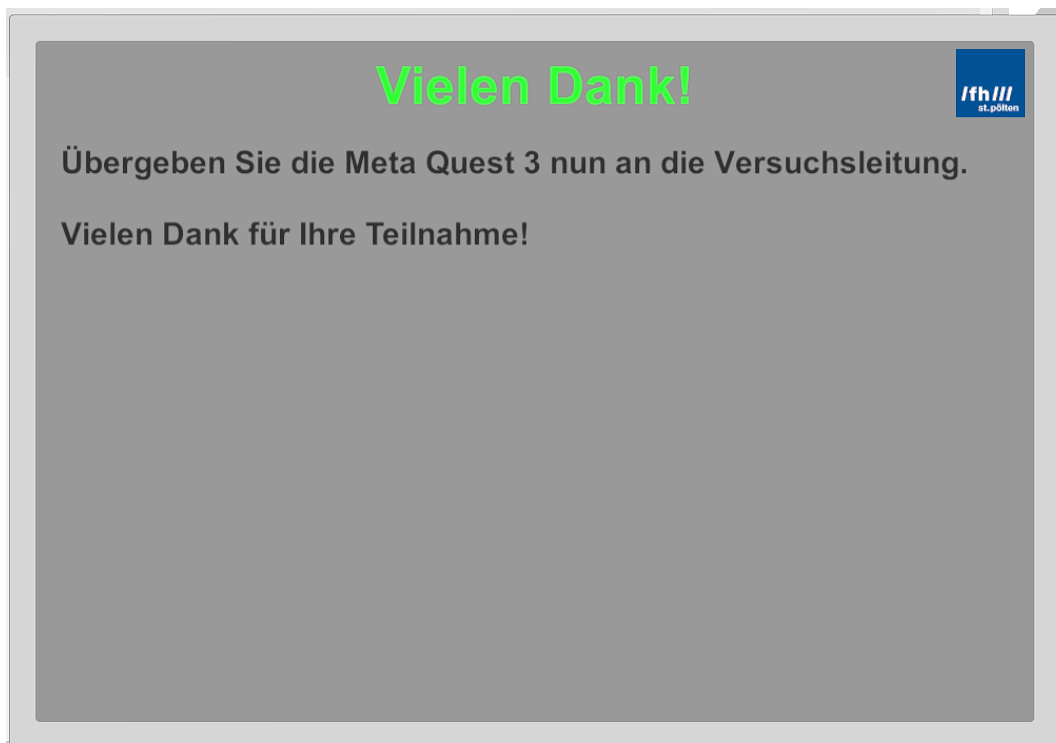
**Ergebnisse speichern**

Vielen Dank für Ihre Teilnahme.  
Bitte drücken Sie auf Speichern, um die Ergebnisse abzuspeichern.

Die Daten werden anonymisiert hinterlegt.

Speichern

## Ende Panel 2



**Vielen Dank!**

Übergeben Sie die Meta Quest 3 nun an die Versuchsleitung.

Vielen Dank für Ihre Teilnahme!

# Questionnaire

## Einleitung

### Spatializing Sound

Liebe\*r Teilnehmer\*in,

vielen Dank, dass Sie sich bereit erklärt haben, an meinem Hörtest teilzunehmen. Ihr Beitrag ist von unschätzbarem Wert für meine Masterarbeit im Bereich der auditiven Wahrnehmung, insbesondere im Zusammenhang mit der Verwendung von Mixed-Reality-Technologie.


Bevor Sie mit dem Hörtest beginnen, möchten ich Sie bitten, einige grundlegende Informationen mit mir zu teilen. Diese Informationen helfen mir, die Ergebnisse besser zu interpretieren und die Qualität meiner Studie zu verbessern. Bitte nehmen Sie sich einen Moment Zeit, um die folgenden Fragen sorgfältig zu beantworten. Alle Ihre Antworten werden vertraulich behandelt, und Ihre persönlichen Daten werden pseudonymisiert, um Ihre Privatsphäre zu schützen.

Es ist wichtig zu beachten, dass dieser Hörtest mit einer Mixed-Reality-Applikation und der Meta Quest 3 durchgeführt wird. Ich bin daran interessiert, nicht nur Ihre auditive Wahrnehmung, sondern auch Ihre Erfahrungen mit dieser speziellen Technologie zu verstehen.

Vielen Dank für Ihre Teilnahme und Ihr Engagement. Ich schätze Ihre ehrlichen und detaillierten Antworten. Wenn Sie während des Hörtests oder im Anschluss weitere Fragen haben, stehe ich Ihnen gerne zur Verfügung. Bitte beginnen Sie nun mit dem Ausfüllen des Fragebogens.

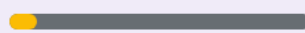
Herzlichen Dank,

Manuel Schumach

 Not shared



Next



Page 1 of 11

Clear form

## Pseudonymisierte demographische Datenerhebung

### Allgemeine Informationen

Bitte geben Sie die von der Testleitung vergebene Teilnehmeridentifikation ein.  
Sie werden gebeten einige kurze Fragen zu Ihrer Person zu beantworten.

Teilnehmeridentifikation \*

Your answer

Alter \*

Your answer

Geschlecht \*

- Weiblich
- Männlich
- Divers

Beruf/Studium \*

Your answer

Haben Sie bekannte Hörschwierigkeiten? \*

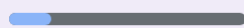
- Ja
- Nein

Haben Sie bereits Erfahrung mit Virtual bzw Mixed Reality Anwendungen? \*

- Nein, überhaupt keine Erfahrung (noch nie genutzt)
- Geringe Erfahrung (einmalig oder unregelmäßig sporadische Nutzungen)
- Moderate Erfahrung (gelegentliche Nutzung)
- Erfahren (im Besitz eines VR/MR Headsets und regelmäßig genutzt)
- Experte (Tätig in der Entwicklung von VR/MR Anwendungen)

Back

Next



Page 2 of 11

Clear form

## Einleitung Hörtest 1

### Hörtest 1

Sie haben nun die einführenden Fragen zum Hörtest beantwortet. Es folgt der erste von vier Hörtests, der mithilfe einer Mixed-Reality-Applikation und dem Meta Quest 3 Headset durchgeführt wird. Ich möchte Ihre Erfahrungen mit der räumlichen Wahrnehmung von Klängen und Ihrer Fähigkeit, virtuelle, nicht sichtbare Klangquellen zu lokalisieren, genauer verstehen.

Während des Hörtests werden Sie gebeten, sich auf die Lokalisierung von Klangquellen zu konzentrieren, die in der Mixed Reality präsentiert werden. Ihr Feedback ist entscheidend, um die Leistungsfähigkeit der Technologie zu bewerten und mögliche Verbesserungen zu identifizieren.

Bitte nehmen Sie sich die Zeit, den Hörtest sorgfältig durchzuführen. Nach Abschluss fahren Sie bitte mit dieser Umfrage fort und beantworten die gestellten Fragen.

[Back](#)

[Next](#)



Page 3 of 11

[Clear form](#)

## Fragen Hörtest 1

### Hörtest 1 Fragen

Beantworten Sie nun folgende Fragen zum eben durchgeführten ersten Hörtest.

Wie genau konnten Sie die Stimme im Raum lokalisieren? \*

	1	2	3	4	5	
Sehr ungenau	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Genau

Inwiefern empfanden Sie die Wahrnehmung der Stimme als natürlich im Vergleich \* zur realen Stimme?

	1	2	3	4	5	
Überhaupt nicht natürlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sehr natürlich

Wie würden Sie den Schwierigkeitsgrad bewerten, die Stimme im Raum zu lokalisieren? \*

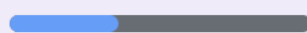
	1	2	3	4	5	
Sehr schwierig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Sehr einfach

Ist Ihnen während der Durchführung des Tests etwas Besonderes aufgefallen? (Zu laut/leise, Störgeräusche,...)

Your answer

Back

Next



Page 4 of 11

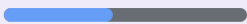
Clear form

## Einleitung Hörtest 2

### Hörtest 2

Es folgt der zweite von vier Hörtests, der mithilfe einer Mixed-Reality-Applikation und dem Meta Quest 3 Headset durchgeführt wird.

Bitte nehmen Sie sich die Zeit, den Hörtest sorgfältig durchzuführen. Nach Abschluss fahren Sie bitte mit dieser Umfrage fort und beantworten die gestellten Fragen.

[Back](#) [Next](#)  Page 5 of 11 [Clear form](#)

## Fragen Hörtest 2

### Hörtest 2 Fragen

Beantworten Sie nun folgende Fragen zum eben durchgeführten zweiten Hörtest.

Wie genau konnten Sie die Stimme im Raum lokalisieren? \*

Sehr ungenau   1   2   3   4   5   Genau

1   2   3   4   5

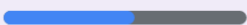
Überhaupt nicht natürlich   1   2   3   4   5   Sehr natürlich

Wie würden Sie den Schwierigkeitsgrad bewerten, die Stimme im Raum zu lokalisieren? \*

Sehr schwierig   1   2   3   4   5   Sehr einfach

Ist Ihnen während der Durchführung des Tests etwas Besonderes aufgefallen? (Zu laut/leise, Störgeräusche,...)

Your answer


[Back](#) [Next](#)  Page 6 of 11 [Clear form](#)

## Einleitung Hörtest 3

### Hörtest 3

Es folgt der dritte von vier Hörtests, der mithilfe einer Mixed-Reality-Applikation und dem Meta Quest 3 Headset durchgeführt wird.

Bitte nehmen Sie sich die Zeit, den Hörtest sorgfältig durchzuführen. Nach Abschluss fahren Sie bitte mit dieser Umfrage fort und beantworten die gestellten Fragen.

[Back](#) [Next](#)  Page 7 of 11 [Clear form](#)

## Fragen Hörtest 3

### Hörtest 3 Fragen

Beantworten Sie nun folgende Fragen zum eben durchgeführten dritten Hörtest.

Wie genau konnten Sie die Stimme im Raum lokalisieren? \*

1 2 3 4 5

Sehr ungenau      Genau

Inwiefern empfanden Sie die Wahrnehmung der Stimme als natürlich im Vergleich \* zur realen Stimme?

1 2 3 4 5

Überhaupt nicht natürlich      Sehr natürlich

Wie würden Sie den Schwierigkeitsgrad bewerten, die Stimme im Raum zu lokalisieren? \*

1 2 3 4 5

Sehr schwierig      Sehr einfach

Ist Ihnen während der Durchführung des Tests etwas Besonderes aufgefallen? (Zu laut/leise, Störgeräusche,...)

Your answer

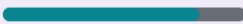
[Back](#) [Next](#)  Page 8 of 11 [Clear form](#)

## Einleitung Hörtest 4

### Hörtest 4

Es folgt der vierte von vier Hörtests, der mithilfe einer Mixed-Reality-Applikation und dem Meta Quest 3 Headset durchgeführt wird.

Bitte nehmen Sie sich die Zeit, den Hörtest sorgfältig durchzuführen. Nach Abschluss fahren Sie bitte mit dieser Umfrage fort und beantworten die gestellten Fragen.

[Back](#) [Next](#)  Page 9 of 11 [Clear form](#)

## Fragen Hörtest 4

### Hörtest 4 Fragen

Beantworten Sie nun folgende Fragen zum eben durchgeführten vierten Hörtest.

Wie genau konnten Sie die Stimme im Raum lokalisieren? \*

1 2 3 4 5

Sehr ungenau      Genau

Inwiefern empfanden Sie die Wahrnehmung der Stimme als natürlich im Vergleich \* zur realen Stimme?

1 2 3 4 5

Überhaupt nicht natürlich      Sehr natürlich

Wie würden Sie den Schwierigkeitsgrad bewerten, die Stimme im Raum zu lokalisieren? \*

1 2 3 4 5

Sehr schwierig      Sehr einfach

Ist Ihnen während der Durchführung des Tests etwas Besonderes aufgefallen? (Zu laut/leise, Störgeräusche,...)

Your answer \_\_\_\_\_

[Back](#) [Next](#)  Page 10 of 11 [Clear form](#)

## Abschließende Fragen

### Abschließende Fragen

Liebe\*r Teilnehmer\*in,

Herzlichen Glückwunsch! Sie haben nun erfolgreich den Hörtest mit allen vier Applikationen abgeschlossen. Ihre Teilnahme war von unschätzbarem Wert für meine Masterarbeit. Bevor Sie den Fragebogen abschließen, möchten wir gerne Ihre abschließenden Gedanken zum Hörtest erfahren.

Ihre ehrlichen Rückmeldungen sind für mich von großer Bedeutung und tragen dazu bei, die Qualität meiner Forschung und Anwendungen zu verbessern.

Ich danke Ihnen herzlich für Ihre Teilnahme an diesem Experiment.

Welche Herausforderungen sind Ihnen während des Hörtests begegnet? Gab es bestimmte Aspekte, die Ihnen besonders leicht oder schwer fielen?

Your answer

---


Bitte teilen Sie mir Ihre allgemeinen Eindrücke vom Hörtest mit. Gibt es etwas, das Sie besonders beeindruckt hat oder das Ihrer Meinung nach verbessert werden könnte?

Your answer

---

[Back](#)

[Submit](#)

 Page 11 of 11

[Clear form](#)