

# **ANALYSE DER LOKALISATIONSFÄHIGKEIT UND DES DARBIETUNGSEMPFINDENS DES MENSCHEN FÜR 3D-AUDIO IN EINER MIXED REALITY UMGEBUNG**

HIER KÖNNTE IHRE WERBUNG STEHEN

Hochschule Emden/Leer, Studiengang Medientechnik

## **Bachelorarbeit**

am xxxx. vorgelegt von

**Hendrik Park**  
geboren am 16.10.1996 in Westerstede  
Matrikelnummer: 7008075

# Inhaltsverzeichnis

<b>1 Einleitung</b>	<b>4</b>
<b>2 Menschliche Wahrnehmung</b>	<b>5</b>
2.1 Funktionsweise des menschlichen Hörens . . . . .	5
2.1.1 Verarbeitung von Schall im Ohr . . . . .	5
2.1.2 Verarbeitung akustischer Signale vom Zentralen Nervensystem . . . . .	7
2.2 Räumliche Wahrnehmung . . . . .	7
2.2.1 Kopfbezogene Übertragungsfunktionen . . . . .	7
2.2.2 Größen des binauralen Hörens . . . . .	8
2.2.3 Funktionsweise der Schallquellenortung . . . . .	9
2.2.4 Im-Kopf-Lokalisation . . . . .	11
2.2.5 Schallquellenortung bei mehreren Schallquellen . . . . .	11
2.3 Zusammenspiel mehrerer Sinnesreize . . . . .	14
2.3.1 Einfluss des Visuellen auf die auditive Wahrnehmung . . . . .	14
2.3.2 Hörausrichtung an Blickrichtung . . . . .	14
2.3.3 Bauchredner-Effekt . . . . .	14
2.4 Wohlbefinden bei akustischen Szenen . . . . .	15
2.4.1 Immersion . . . . .	15
2.4.2 Das Uncanny Valley und der Bezug zur Immersion bei VR und AR . . . . .	16
<b>3 Technische Grundlagen</b>	<b>18</b>
3.1 Augmented Reality . . . . .	18
3.1.1 Microsoft Mixed Reality . . . . .	19
3.2 Microsoft HoloLens . . . . .	19
3.2.1 Head-Mounted-Displays . . . . .	20
3.2.2 Kopfhörer der HoloLens . . . . .	20
3.2.3 Head-Tracking-Funktion . . . . .	20
3.3 Binaurale Wiedergabe . . . . .	21
3.3.1 Kopfbezogene Übertragungsfunktionen - MS HRTF Spatializer . . . . .	21
3.3.2 Auralisation . . . . .	21
3.4 Software . . . . .	23
3.4.1 Blender . . . . .	23
3.4.2 Unity . . . . .	23
<b>4 Vorstellung der HoloLens-Applikation und des Hörversuches</b>	<b>25</b>
4.1 Idee hinter dem Versuch . . . . .	25
4.2 Beschreibung der Virtuellen Welt . . . . .	26
4.2.1 Dorf . . . . .	26
4.2.2 Wald . . . . .	27
4.2.3 Gebirge und Fluss . . . . .	28

4.3	Audio-Einstellungen in der Applikation . . . . .	28
4.3.1	Binauralisierung in der Applikation . . . . .	29
4.3.2	Auralisierung in der Applikation . . . . .	29
4.4	Beschreibung der Szenen des Hörversuchs . . . . .	30
4.4.1	Szene 1 - Adlerschrei im Gebirge . . . . .	30
4.4.2	Szene 2 - Wolfsheulen . . . . .	31
4.4.3	Szene 3 - Kirchenglocken . . . . .	32
4.4.4	Szene 4 - Flussrauschen . . . . .	33
4.4.5	Szene 5 - Adlerschrei über der Welt . . . . .	33
4.5	Bewertungskriterien der Darbietung . . . . .	34
4.6	Beschreibung der Benutzeroberfläche . . . . .	35
4.7	Ablauf des Hörversuchs . . . . .	38
4.8	Probanden . . . . .	40
<b>5</b>	<b>Versuchsergebnisse</b>	<b>41</b>
5.0.1	Szene 1: Adlerschrei im Adlernest . . . . .	41
5.0.2	Szene 2: Wolfsheulen . . . . .	44
5.0.3	Szene 3: Kirchenglocken . . . . .	47
5.0.4	Szene 4: Adler in der Luft . . . . .	49
5.0.5	Szene 5: Flussrauschen . . . . .	51
5.1	Ergebnisse kritisch hinterfragen . . . . .	54

# 1 Einleitung

Laserschwertkämpfe, ein Besuch auf dem Mars oder die Fähigkeit ohne Flugzeug fliegen zu können sind Bestrebungen des Menschen, die entweder noch in ferner Zukunft liegen oder wohl niemals Realität werden. Doch auch wenn der einfache Mensch diese Erfahrungen wohl niemals wirklich erleben wird, so ist der Schritt diese Erlebnisse nachzuempfinden mit der heutigen Technik gar kein so großer mehr. Der Fortschritt in der virtuellen Realität lässt uns einen simulierten Mars besuchen, die Entwicklung von Augmented lässt in der realen Welt eine Fernbedienung als Laserschwert erscheinen und auch das Fliegen durch simulierte Welten kann so echt wirken, dass manch ein Mensch mit der Höhenangst zu kämpfen hat.

Um diese virtuellen Erfahrungen so plausibel wie möglich zu gestalten ist es wichtig die Inhalte an den Menschen anzupassen. Die Überlagerung der Realität mit virtuellen Inhalten wird als Augmented Reality bezeichnet und während die Anforderungen an eine visuelle Überlagerungen schon sehr weit erforscht ist, so ist dies für auditive Überlagerung bisher nicht der Fall.

Durch diverse Software ist es bereits möglich Schallquellen in der virtuellen Realität so zu positionieren, dass es so klingen soll als wären diese tatsächlich an der festgelegten Position im Raum. Dies soll es Entwicklern ermöglichen extrem realistische Akustik in virtuellen Welten zu erzeugen. Wie plausibel die räumliche Wiedergabe funktioniert und wie diese vom Nutzer angenommen wird, soll im Rahmen dieser Arbeit überprüft werden. Hierfür wurde ein Hörversuch entwickelt, bei dem Probanden in einer Mixed Reality- Umgebung virtuelle Schallquellen lokalisieren sollten.

Der durchgeführte Hörversuch stellt einen typischen Anwendungsfall für Augmented Reality dar. Es werden virtuelle Inhalte in die reale Welt projiziert, die währenddessen durch binaurale Hörereignisse ergänzt werden. Für die Probanden gilt es die Position der virtuellen Schallquellen anhand des Schalls zu bestimmen und anschließend die Darbietung anhand bestimmter Kriterien zu bewerten. Die Plausibilität der Hörereignisse und das Immersionsempfinden des Probanden stehen hierbei im Vordergrund.

# **2 Menschliche Wahrnehmung**

Die Erstellung multimedialer Anwendungen erfordert präzises Wissen über die menschliche Wahrnehmung, so wäre ohne die Anpassung der Bildwechselsequenz an das Auge ein Kino nur ein Raum mit flackernden Bildern, ohne flüssige Bewegungen. Gleiches gilt für die auditive Wahrnehmung des Menschen. Ohne die Physiologie und die Psychologie des Menschen zu analysieren ist es nicht möglich adäquate Inhalte zu erstellen, daher soll es im folgenden Kapitel darum gehen die Sinneswahrnehmung des Menschen zu verinnerlichen. Der Fokus wird hierbei auf die räumliche Wahrnehmung durch audiovisuelle Reize gelegt.

## **2.1 Funktionsweise des menschlichen Hörens**

Um zu verstehen wie der Mensch seine Umgebung wahrnimmt, ist es zunächst wichtig zu erklären wie der Mensch überhaupt in der Lage ist Informationen aus Schall zu beziehen. Aus diesem Grund wird zunächst geklärt wie der Mensch Schall in für ihn verwertbare Signale umwandelt und wie er diese anschließend verarbeitet.

### **2.1.1 Verarbeitung von Schall im Ohr**

Die Umwandlung von Schall in verwertbare Signale findet im Mittel- und Innenohr des Menschen statt, siehe Abbildung 2.1. Der Schall trifft auf das Trommelfell, wodurch dieses zu schwingen beginnt. Die Gehörknöchelchen übertragen nun angepasst an die höhere Dichte der Gehörfüssigkeit der Cochlea den Schall in das Innenohr. Frequenzabhängig erreichen die übertragenden Wellen ihre maximale Amplitude an verschiedenen Stellen der Basilarmembran. Diese Frequenzdispersion wird als Ortstheorie des Hörens bezeichnet. Die sich an der Basilarmembran befindenen Stereozilien werden durch die Wellenbewegungen der Gehörfüssigkeit so bewegt, dass sich Ionenkanäle öffnen und ein elektrisches Potential durch diffundierende Ladungsträger entsteht. Auf diese Weise entstehen Signale, die anschließend über den Hörnerv an das Zentrale Nervensystem übertragen und dort analysiert werden [12, S.48].

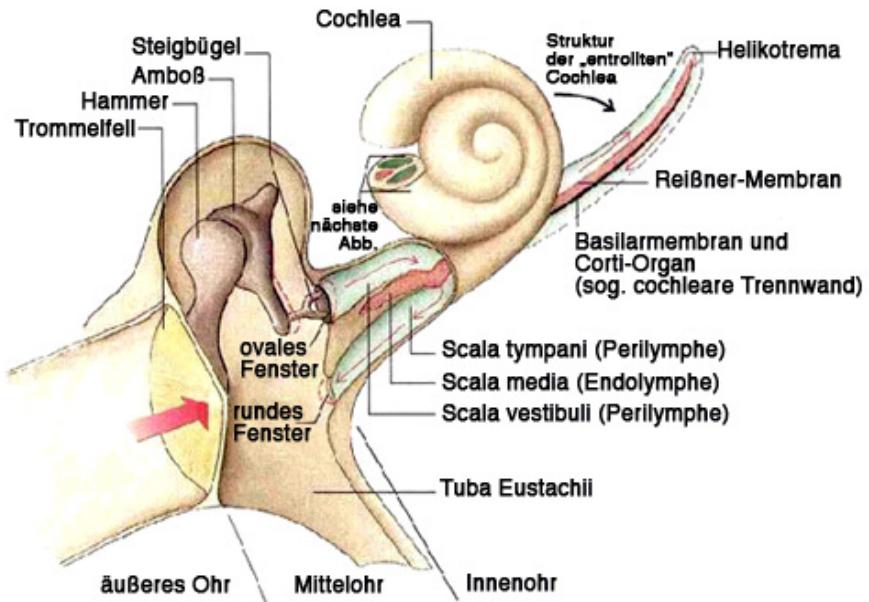


Abbildung 2.1: Schematische Darstellung eines Hörorgans mit abgerollter Kochlea  
Abruf 25.01.19 [http://www.eclim.de/AkuLNew/roomac\\_1n.html](http://www.eclim.de/AkuLNew/roomac_1n.html)

Das menschliche Gehör ist in der Lage Schallereignisse mit Frequenzen von ca. 20Hz bis 16kHz - 20kHz wahrzunehmen. Dabei nimmt die Fähigkeit hohe Frequenzen zu hören um ca. 1kHz pro Lebensdekade, durch Abnutzung der Haarzellen, mit dem Alter ab [4, S.41].

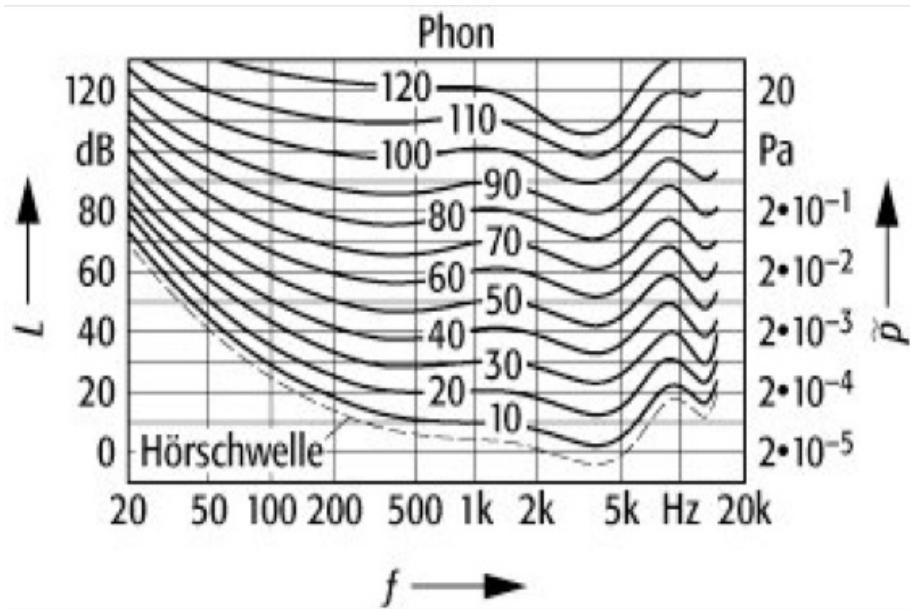


Abbildung 2.2: Kurven gleicher Lautheit  
Abruf 25.01.19 <https://www.spektrum.de/lexikon/physik/kurven-gleicher-lautstaerke/8648>

Der Dynamikbereich des menschlichen Gehörs wird durch die Ruhehörschwelle und die Schmerzschwelle eingegrenzt und kann hierbei frequenzabhängig zwischen 120 dB und 140 dB umfassen. Die Kurven gleicher Lautheit, siehe Abbildung 2.2 zeigen, dass Schallereignisse je nach Fre-

quenz unterschiedlich laut wahrgenommen werden. Diese Frequenzabhängigkeit der empfundenen Lautheit basiert auf der Anatomie des menschlichen Ohres. Im mittleren Frequenzbereich (1 kHz bis 5 kHz) ist das menschliche Ohr am empfindlichsten. Dies ist deutlich daran zu erkennen, dass für die gleiche empfundene Lautheit bei tiefer- bzw. höherfrequenten Schallereignissen ein deutliche höherer Schalldruckpegel erreicht werden muss. Außerdem ist in diesem Bereich die Dynamik zwischen Hörschwelle (0 Phon) und Schmerzgrenze (120 Phon) am höchsten. Dies ist evolutionsbiologisch darauf zurückzuführen, dass sowohl viele Umgebungsgeräusche als auch für das Sprachverständnis wichtige Konsonanten in diesem Frequenzbereich liegen [12, S.47].

### **2.1.2 Verarbeitung akustischer Signale vom Zentralen Nervensystem**

Die im Zentralen Nervensystem eintreffenden Signale bilden nach der Ortstheorie des Hörens nur ein bestimmtes Frequenzband ab. Die elektrischen Signale lassen sich daher als durch Bandpassfilter verschiedener Bandbreite und Güte, gefilterte Signale betrachten, wobei das menschliche Gehör in der Lage ist die Parameter an die zu analysierende Szene adaptiv anzupassen [4, S.59].

Die Verarbeitung im Zentralen Nervensystem erfolgt nach dem Prinzip der Mustererkennung. Die Signale werden spektral analysiert werden, um Merkmale zu extrahieren und anschließend bestimmte Muster zu identifizieren und wiederzuerkennen [12, S.51]. Die Betrachtung der neuronalen Prozesse zur Mustererkennung soll im Rahmen dieser Arbeit nicht weiter erläutert werden. Wichtig zu sagen ist jedoch, dass zur Analyse von Schallereignissen frühere Wahrnehmungen, Erfahrungen, sowie andere Sinnesreize herangezogen werden. Aus diesem Grund variiert die Wahrnehmung von Individuum zu Individuum, weshalb der Aspekt der individuellen Wahrnehmung bei der Erstellung multimedialer Inhalte stets berücksichtigt werden muss.

## **2.2 Räumliche Wahrnehmung**

Zur Erkennung des Ortes und der Ausdehnung werden Schallsignale nach bestimmten Merkmalen ausgewertet. Die am Trommelfell eintreffenden Schallsignale können hierbei in monoaurale und interaurale Ohrsignale kategorisiert werden. Zum Empfang monoauraler Ohrsignale ist ein Ohr ausreichend, während für den Empfang interauraler Ohrsignale beide Ohren benötigt werden. Die räumliche Wahrnehmung erfolgt sowohl durch die Analyse monoauraler als auch interauraler Ohrsignale, wobei das interaurale Hören hauptsächlich für die horizontale und das monoaurale Hören für die vertikale Wahrnehmung verantwortlich ist [12, S.88].

### **2.2.1 Kopfbezogene Übertragungsfunktionen**

Das menschliche Gehör kann im technischen Sinne als Antenne betrachten werden, dessen Richtcharakteristik eine Richtungs-, Entfernungs- und Frequenzabhängigkeit aufweist [12, S.89]. Je nach Ort der Schallquelle ändert sich damit Phasengang, Gruppenlaufzeit und Frequenzgang des Schalls, der letztlich am Trommelfell ankommt. Die Veränderung der Schalls von der Quelle des Schallereignisses bis hin zur Senke, also dem Trommelfell, kann mittels sogenannter kopfbezogener Übertragungsfunktionen oder Head-Related Transfer Function (HRTF) beschrieben werden. Eine HRTF ist abhängig vom Ort der Quelle, dem Ort der Senke und dem Frequenzspektrum des Schallereignisses. Die Parameter einer HRTF  $H(f, r, \phi, \delta)$  sind somit

die Frequenz  $f$ , der Abstand zwischen Trommelfell und Schallereignis  $r$  und die kopfbezogenen Richtungswinkel Azimut  $\phi$  und Elevation  $\delta$ . Alle Parameter variieren zwischen dem linken und dem rechten Ohr, so dass immer ein binaurales Paar von HRTFs bzw. eine sogenannte interaurale Außenohr-Übertragungsfunktion  $H_i$  betrachtet werden muss. Diese setzt sich aus dem Verhältnis der linken HRTF  $H_l$  und der rechten HRTF  $H_r$  zusammen [12, S.91]:

$$H_i(f, r, \phi, \delta) = \frac{H_r(f, r, \phi, \delta)}{H_l(f, r, \phi, \delta)} \quad (2.1)$$

Die kopfbezogenen Übertragungsfunktionen sind auf Grund der unterschiedlichen Formen von Kopf, Schultern und Außenohren des Menschen sehr individuell. In Abbildung 2.3 ist dies besonders im Frequenzbereich über 7kHz zu erkennen, da die Schalldruckpegel hier sehr stark zwischen den bei Probanden gemessenen HRTFs variieren.

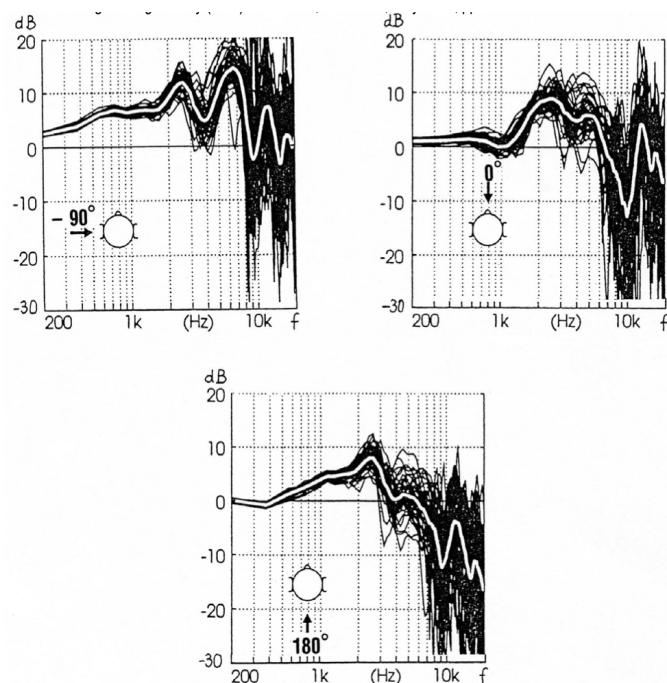


Abbildung 2.3: Kopfbezogene Übertragungsfunktionen eines linken Ohres  
Es wurden HRTFs mehrerer Personen für unterschiedliche Richtungen aufgenommen.  
Die weißen Kurven entsprechen dem bewerteten Mittelwert.

Abruf 25.01.19 <http://www.sengpielaudio.com/KopfbezogeneUebertragungsfunktionHRTF.pdf>

## 2.2.2 Größen des binauralen Hörens

Wichtige Merkmale, die zur Lokalisation eines Schallereignisses analysiert werden, lassen sich aus der interauralen Außenohr-Übertragungsfunktion bestimmen. Nach Brauert und Braasch werden diese Größen des binauralen Hörens wie folgt berechnet [12, S.91]:

### Interaurale Laufzeitdifferenzen

Die interaurale Phasenlaufzeit oder auch interaurale Laufzeitdifferenz  $\tau_{ph}$  berechnet sich aus dem Verhältnis der interauralen Phasendifferenz  $\Delta L_i$ , also der Differenz der Phase von linker

zu rechter HRTF, zur Frequenz  $f$  des Schallereignisses:

$$\tau_{ph}(f, r, \phi, \delta) = \frac{\Delta\delta_i(f, r, \phi, \delta)}{f} \quad (2.2)$$

Anhand der interauralen Laufzeitdifferenz lässt sich bestimmen an welchem Ohr der Schall bestimmter Frequenzen zuerst ankommt. Dies ist ein, für den Menschen, wichtiges Merkmal für die Lokalisierung des Schallereignisses, da von links kommender Schall auch zu erst beim linken Ohr auftritt und umgekehrt.

### **Interaurale Gruppenlaufzeit**

Für die interaurale Gruppenlaufzeit wird das Verhältnis der Ableitung der interauralen Phasendifferenz zur Frequenz  $f$  des Schallereignisses berechnet:

$$\tau_{gr}(f, r, \phi, \delta) = \frac{d\Delta\delta_i(f, r, \phi, \delta)}{f} \quad (2.3)$$

Die interaurale Gruppenlaufzeit gibt im Gegensatz zur interauralen Laufzeitdifferenz an, bei welchem Ohr die Einhüllende des Schallereignisses zu erst ankommt und wie stark verzögert sie das zweite Ohr erreicht.

### **Interaurale Pegeldifferenzen**

Die interaurale Pegeldifferenz  $\Delta L_i$  wird als logarithmisches Maß zur Beschreibung der Intensitätsunterschiede des Schalls verwendet, der an beiden Ohren ankommt:

$$\Delta L_i = 20 * \lg(H_i(f, r, \phi, \delta)) \quad (2.4)$$

### **2.2.3 Funktionsweise der Schallquellenortung**

Das Außenohr, der Kopf und die Schultern wirken alle als richtungs-, entfernungs- und , frequenzabhängige Filter für den am Trommelfell an kommenden Schall. Ist dem Hörenden ein Schallereignis bekannt, so ist er in der Lage die Filtereigenschaften zu bestimmen und so Rückschlüsse auf die Position des Schallereignisses zu ziehen.

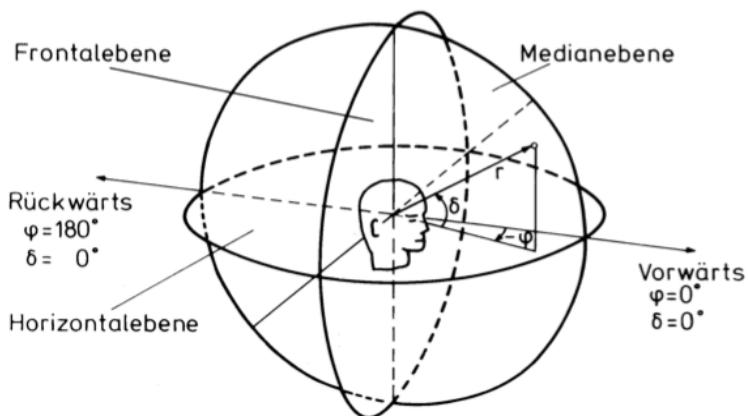


Abbildung 2.4: Kopfbezoogenes Koordinatensystem  
Nach Blauert und Braasch [12, S.88]

### Richtungswahrnehmung

Schallquellen innerhalb der Frontal- und Horizontalebene können anhand von interauralen Laufzeitdifferenzen, interauralen Pegeldifferenzen sowie interauralen Gruppenlaufzeitunterschieden lokalisiert werden. Der Kopf des Menschen weist eine Filtercharakteristik für hohe Frequenzen auf. Wellenlängen, die kleiner als der Kopf selbst sind, werden durch diesen gedämpft. Die Ortung hoher Frequenzen erfolgt daher weitestgehend durch die Analyse der Gruppenlaufzeit und der Pegeldifferenzen, da die Einhüllende des Schalls durch den Kopf weniger stark beeinflusst wird. Die Lokalisierung niederfrequenten Schalls erfolgt dahingehend durch die Verarbeitung der Laufzeit- und Pegeldifferenzen [4, S.45].

Die monoauralen Außenohrübertragungsfunktionen weisen auf Grund der Struktur des Außenohres eine signifikante Abhängigkeit zum Einfallsinkel des Schalls auf. Dadurch ist es dem Menschen erlaubt schon geringe Positionsänderungen der Schallquelle zur Mediane Ebene wahrzunehmen [4, S.46].

Die horizontale Wahrnehmung ist empfindlicher als die vertikale Wahrnehmung. In der Horizontalebene werden Richtungsänderungen ab ca. 4 grad und in der Frontalebene erst ab ca. 10 grad merklich wahrgenommen [12, S.95].

Die Lokalisation von Schallquellen in der Mediane Ebene basiert auf der Verarbeitung monoauraler Merkmale, da durch die Position der Schallquelle der Direktschall für beide Ohren annähernd identisch ist. Auch hier gilt, dass die Filterwirkung von Außenohr und Körper, speziell die Schultern, den Schall so beeinflussen, dass Änderungen des Klangs zur Richtungslokalisation interpretiert werden können [4, S.46].

Wellenlängen, die größer als die Ohrmuschel selbst sind, werden durch das Ohr kaum beeinflusst. Daraus resultiert, dass die Filterwirkung des Außenohres sich auf Frequenzen oberhalb von ca. 3,5 kHz begrenzt. Die Richtungslokalisation von Schallquellen in der Horizontal- und Frontalebene erfolgt daher besonders gut bei höherfrequenten Schallereignissen [4, S.47].

Ein wichtiger Aspekt der Richtungslokalisation ist die Fähigkeit des Menschen seinen Kopf

zu bewegen. Die dadurch entstehenden Änderungen der interauralen Laufzeit- und Intensitätsdifferenzen lassen sich zeitlich analysieren, so dass die Richtung des Schalls leichter erkannt werden kann [12, S.88].

### **Entfernungswahrnehmung**

Wie bei der Richtungslokalisierung gelingt auch das Entfernungshören aufgrund der Verarbeitung von mono- sowie interauralen Merkmalen.

Ist die Schallquelle dem Hörenden sehr nahe (weniger als 25cm entfernt) so bewirkt die Abschirmung durch den Kopf, dass der Schall bei einem Ohr stark gedämpft ankommt. Die interaurale Pegeldifferenz ist daher für die Entfernungseinschätzung naher Schallquellen von herausragender Bedeutung [12, S.98].

Des Weiteren wird auch der allgemeine Signalpegel zur Entfernungseinschätzung genutzt. Ist ein Signal leiser so ist davon auszugehen, dass die Quelle weiter entfernt ist [12, S.98].

Ist die Schallquelle zwischen 25cm und 15m vom Hörer entfernt spricht man vom mittleren Entfernungsreich. Bei Kugelstrahlern 0. Ordnung, im freien Schallfeld, erzielt eine Abstandsverdopplung der Quelle zum Hörer, in diesem Bereich, eine Steigerung des Signalpegels um 6dB [12, S.98]. In einem Raum gilt, dass die Menge an Direktschall, der am Trommelfell ankommt, mit der Entfernung abnimmt. Aus diesem Grund ist das Verhältnis von Direktschall zu diffusem Schall ein weiteres informatives Merkmal zu Einschätzung von Entfernungen [7, S.57].

Die Entfernung zu Hörereignissen, die weiter als 15m entfernt sind, kann ohne Hintergrundwissen, wie z.B visuelle Eindrücke, nicht geschätzt werden. Da Schall über weite Entfernung jedoch frequenzabhängig gedämpft wird, lässt sich anhand des dumpfen Klangs beurteilen, ob die Schallquelle weit entfernt sein muss [12, S.99].

### **2.2.4 Im-Kopf-Lokalisation**

Sind die Schallquellen sehr nah, z.B bei Kopfhörern, kann es zur sogenannten Im-Kopf-Lokalisation kommen. Das Schallereignis wird hierbei im Kopf statt vor den Ohren lokalisiert.

Die Filterwirkung des Außenohrs und die Beschallung der Schädelknochen, die ebenfalls vom Gehör analysiert werden kann, fallen bei der zu nahen Beschallung weg. Da dies natürlicherweise nicht vorkommt ist das Gehör nicht daran gewöhnt Schall zu analysieren der nicht durch das Außenohr entzerrt worden ist. Der Im-Kopf-Lokalisation kann daher durch eine vorherige Entzerrung der Ohrsignale entgegengewirkt werden, so dass auch über Kopfhörer entfernte Schallquellen simuliert werden können [2, S.47].

### **2.2.5 Schallquellenortung bei mehreren Schallquellen**

Bei der Hörereignisbildung mit mehreren Schallquellen treten verschiedenste Effekte auf. Die wichtigsten Effekte sollen im Sinne dieser Arbeit kurz erläutert werden, um ihren Einfluss beim Hörversuch einschätzen zu können.

## Summenlokalisierung - Phantomschallquellen

Bei zwei Schallquellen, die das selbe Signal wiedergeben und gleich weit vom Hörenden entfernt sind, siehe Abbildung 2.5, wird das Hörereignis zentral zwischen den beiden Schallquellen gebildet. Diese Summenlokalisierung der Ohrsignale führt zur Bildung einer imaginären Schallquelle, welche zentral zwischen den realen Schallquellen lokalisiert wird. Diese imaginären Schallquellen werden als Phantomschallquellen bezeichnet [12, S.101]. Die Richtungslokalisierung ist jedoch bei Phantomschallquellen in der Regel fehleranfälliger als bei realen Schallquellen. Dies gilt besonders bei der seitlichen Bildung von Phantomschallquellen mit Surround-Sound [12, S.103].

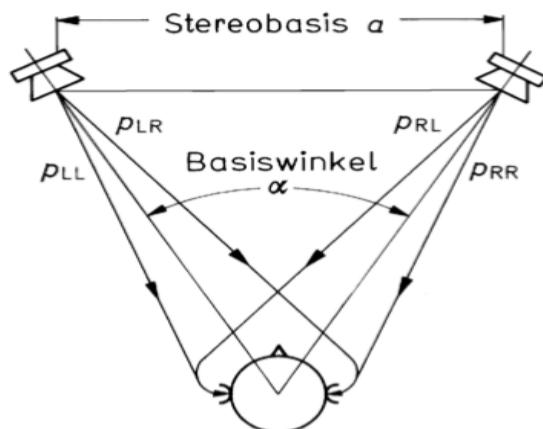


Abbildung 2.5: Stereolautsprecher-Anordnung  
[12, S.101]

Bei der Summenlokalisierung werden Laufzeit- und Pegeldifferenzen zwischen den beiden Lautsprechersignalen analysiert, um die Phantomschallquelle zu orten. Ist ein Signal stärker als das andere wird die Phantomschallquelle eher in Richtung der lauteren, realen Quelle lokalisiert. Gleichzeitig wird die Phantomschallquelle eher in Richtung der realen Schallquelle lokalisiert, dessen Schall früher beim Hörenden ankommt. Bei größeren Verzögerungen kommt es zum so genannten Präzedenzeffekt.

Anhand der Abbildung 2.6 lässt sich erkennen, dass die Richtungszuordnung in keinem linearen Zusammenhang zu wahrgenommenen Laufzeit- und Pegeldifferenzen steht. Des Weiteren ist zu erwähnen, dass sich die Auswirkung der Laufzeit- und Pegeldifferenzen gegenseitig beeinflussen und aufheben können [12, S.103].

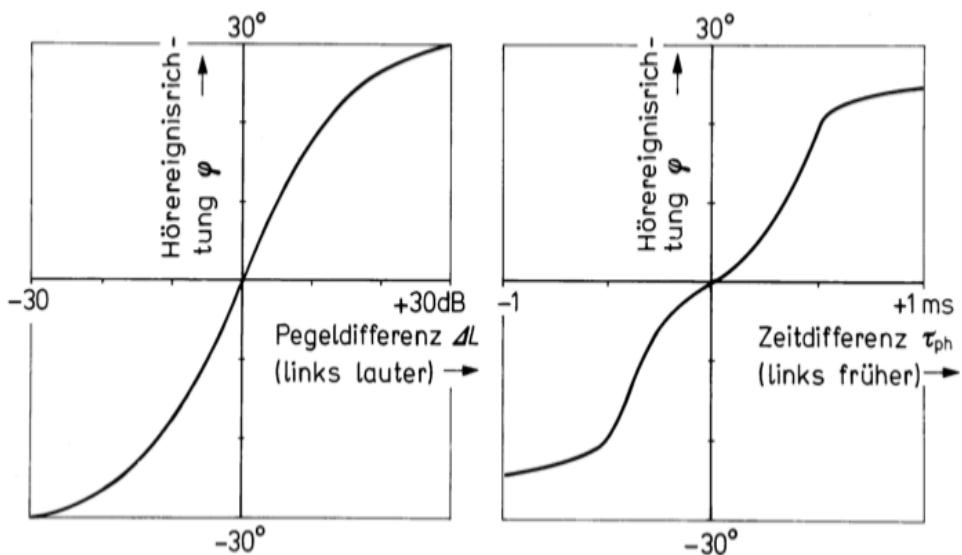


Abbildung 2.6: Summenlokalisationskurven für Pegel- und Laufzeitdifferenzen der beiden Lautsprechersignale

[12, S.101]

### Präzedenzeffekt

Bei der Wiedergabe eines Signals über mehrere Lautsprecher wird „nach dem Gesetz der ersten Wellenfront, die Hörereignisausrichtung fast ausschließlich anhand des ersten eintreffenden Schalls vorgenommen. Die Signale, die mit mehr als einer Millisekunde Verzögerung eintreffen werden dann als indirekter Schall gewertet, so dass dieser nur zur Lautstärke beiträgt und nicht zur Richtungslokalisierung. Dies ermöglicht es auch in geschlossenen Räumen ein Hörereignis trotz mehrfach reflektierten Schalls zu orten [12, S.103].

Treffen Signale mit einer Verzögerung von mehr als 50 ms auf wird der Schall nicht als indirekter Schall sondern als Echo erkannt. Dies bewirkt, dass der Schall nicht in die Summenlokalisierung für die Position der ersten Quelle mit einbezogen wird, sondern stattdessen als einzelnes Hörereignis mit eigener Quelle erkannt wird. Besitzt ein Signal einen deutlichen höheren Pegel als die vorherigen Signale kann es auch bei einer geringeren Verzögerung als 50ms schon als Echo gedeutet werden [4, S.49].

### Binaurale Störgeräuschunterdrückung/ Cocktailparty-Effekt

Der Mensch ist in der Lage den Schall in seiner Umgebung zu analysieren und zwischen Nutzschall und Störgeräuschen zu unterscheiden. Durch die Lokalisation der Nutzschallquelle, z.B. einer sprechenden Person, kann bewusst die Wahrnehmung anderer Störschallquellen unterdrückt werden (Cocktailparty-Effekt). Die Analyse interauraler Laufzeit- und Pegeldifferenzen ermöglicht eine Dämpfung des Störschalls zwischen 9 und 15 dB. Die Effektivität der Störschallunterdrückung ist abhängig von der Richtung des einfallenden Schalls [4, S.48].

## **2.3 Zusammenspiel mehrerer Sinnesreize**

Für die Deutung der akustischen Signale vom Zentralen Nervensystem werden, neben den akustischen, weitere Sinnesreize herangezogen. Im Rahmen der Versuches dieser Arbeit werden Probanden sowohl optischen als auch akustischen Sinnesreizen unterzogen. Wie sich die Wahrnehmungen dieser Reize gegenseitig beeinflussen und welche Effekte hierbei auftreten können soll im Folgenden erläutert werden.

### **2.3.1 Einfluss des Visuellen auf die auditive Wahrnehmung**

Die Einordnung des Schallereignisses mittels einer Mustererkennung im Gehirn des Menschen basiert darauf, dass Merkmale aus den aktuell eintreffenden Reizen gebildet werden. Solch ein Merkmal könnte beispielsweise eine bestimmte spektrale Zusammensetzung sein. Die Menge aller Merkmale wird daraufhin mit Merkmalen vergangener Ereignisse verglichen. Korreliert hierbei eine Merkmalsmenge mit der eines vergangen Ereignisses, so kann das Zentrale Nervensystem das Schallereignis kategorisieren und einordnen [12, S.51]. Neben akustischen Merkmalen werden auch aus optischen Reizen Merkmale gebildet [12, S.88]. Die Wahrnehmungsfähigkeit des Menschen steigert sich durch den Einsatz mehrerer Sinnesreize enorm, da es dem Zentralen Nervensystem so ermöglicht eine größere Menge an Merkmalen zu bilden, irrelevante Merkmale leichter auszusortieren und auf einen größeren Erfahrungsschatz zurückzugreifen.

### **2.3.2 Hörausrichtung an Blickrichtung**

Neben diesem psychologischen Aspekt, muss der physiologische Aspekt der Anpassung der Richtcharakteristik des Gehörs an die Blickrichtung mit herangezogen werden. Der Mensch ist nachweislich in der Lage sowohl die Ausrichtung des Trommelfells als auch die Stellung der Haarzellen in der Kochlea an die Blickrichtung anzupassen, um Hörereignisse, auf die er sich fokussiert, besser hören zu können. Dabei geschieht die Ausrichtung der Gehörkomponenten sogar vor dem Wechsel des Blickpunkts[11]. Eine Visualisierung der akustischen Szene kann also dafür sorgen, dass gezielt bestimmte Inhalte besser gehört werden können, in dem diese visuell Aufmerksamkeit auf sich ziehen.

### **2.3.3 Bauchredner-Effekt**

Treffen akustischer und optischer Reiz synchron, also mit einer Verzögerung von maximal ca. 50- 100ms, auf können sich die individuellen Wahrnehmungen gegenseitig beeinflussen. Für die Deutung der auditiven Merkmale ist der optische Reiz hierbei von größerer Bedeutung als es umgekehrt der Fall ist. Bauchredner versuchen auf diese Weise dem Publikum die Lokalisation der wahren Schallquelle zu erschweren. Der sich bewegende Mund der Puppe stellt einen optischen Reiz dar, der mit dem Hörereignis des sprechenden Redners übereinstimmt. In Folge dessen wird die wahrgenommene Schallquelle auf den Mund der Puppe projiziert [4, S.163].

Obwohl visueller und akustischer Reiz nicht zusammen passen wird die Gesamtsituation vom Konsumenten als plausibel wahrgenommen. Dies bedeutet, dass ein perfektes Panning der auditiven Signale für eine plausible Wiedergabe nicht immer von Nöten ist. Andererseits ist aber auch zu beachten, welchen visuellen Reizen der Konsument im Moment der auditiven Wiedergabe unterstellt ist. Schaut der Betrachter im Moment der auditiven Wiedergabe auf eine

potentielle, alternative Schallquelle kann es zu ungewollter Fehllokalisierung der Hörereignisse stattfinden kann.

## 2.4 Wohlbefinden bei akustischen Szenen

Damit virtuell erstellte Medien bei den Kunden Anklang finden müssen diese sich beim Konsum vor allen Dingen wohlfühlen. Gerade im Bereich AR und VR geht es oft darum, dass akustische und visuelle Szenen so realistisch wie möglich dargestellt werden, dabei ist bereits bewiesen, dass zu realistische, visuelle Szenen eher Unbehagen als Begeisterung beim Konsumenten bewirken. Im Rahmen dieser Arbeit soll dahingehend auch geprüft werden inwiefern bei akustischen Szenen solch ein Unbehagen ausgelöst werden kann und wie sich die Versuchspersonen bei der Darbietung der audiovisuellen Szenen fühlen.

### 2.4.1 Immersion

Virtual Reality wird häufig mit dem Wort Immersion in Verbindung gebracht. Als Immersion bezeichnet man die Einhüllung in Medien. Im Folgenden soll kurz der Zustand der Immersion erläutert und im Anschluss der Zusammenhang zu den Themen dieser Bachelorarbeit geschlossen werden.

Nach Grau äußere sich die Immersion in dem Moment in dem der Betrachter die kritische Distanz zum Medium verliere und die emotionale Involvierungen ansteige [6, S.13]. Der Konsument wird im Zustand der Immersion also Teil der medialen Szene und ist in der Lage seine reale Umgebung komplett auszublenden. Auch wenn dies schon bei älteren Medien wie Fernsehen oder Theater der Fall gewesen ist, so ist dieser Zustand mit dem Einsatz von VR und AR noch intensiver zu verwirklichen.



Abbildung 2.7: Immersives Erlebnis durch Achterbahnsimulation mit der Samsung Gear  
Abruf 29.01.19: <http://forums.cgsociety.org/t/samsung-gear-vr-fully-immersive-4d-experience/1810570>

Virtual Reality biete hierbei Techniken, die traditionelle Medien polysensorisch erlebbar machen würden, behauptet Grau [6, S.15] In diesem Zusammenhang wurde besonders der Aspekt der freien Kopfbewegung genannt, der es ermögliche die Szene aus unendlich vielen Perspektiven zu beobachten. Auch der Einsatz von räumlichem Audio sowie die Interaktion mit der virtuellen Welt erlaube ein noch stärkeres Gefühl der Immersion [6, S.16].

Damit die synthetisch erschaffene Welt eine Immersion zulasse, müsse diese nicht die Umgebung komplett umfassen, beziehungsweise realistisch sein, sondern stattdessen plausibel[6, S.17]. Die Stärke der Immersion hänge dabei zum einen von der Glaubwürdigkeit, aber auch von der Genießbarkeit der Illusion ab[6, S.17].

Die Immersion ist im Zusammenhang mit dem Einsatz von räumlichem Audio ein erstrebenswerter Zustand für den Konsumenten. Dieser soll den Eindruck vermittelt bekommen, dass die ihm dargebotene virtuelle Szene Wirklichkeit sein könnte, so dass dieser die Realität ausblendet.

In den Versuchen dieser Arbeit soll eine virtuelle Welt in die reale Welt integriert werden. Hierbei werden visuelle und auditive Inhalte im Raum platziert und über eine Microsoft Mixed Reality Brille wiedergegeben. In diesem Zusammenhang würde eine Immersion bedeuten, dass der Konsument die dargebotenen, virtuellen Inhalte als Teil seiner realen Umgebung wahrnimmt.

Nicht nachvollziehbare Reize können eine Immersion beim Konsumenten verhindern[6, S.17]. Im Zusammenhang mit den Versuchen bedeutet dies, dass sowohl auditive und visuelle Inhalte, als auch reale und virtuelle Welt zueinander kompatibel sein müssen und plausible Sinneseindrücke bewirken müssen.

Um das Immersionsempfinden der Probanden während des Hörversuchs zu überprüfen werden, die Probanden gebeten die Darbietung nach den Kriterien der Plausibilität, dem Realismus der auditiven Inhalte und ihrem eigenen Wohlbefinden für jede Szene zu bewerten.

## 2.4.2 Das Uncanny Valley und der Bezug zur Immersion bei VR und AR

Das Uncanny Valley ist ein Begriff der Robotik, der zum ersten mal vom japanischen Robotiker Masahiro Mori 1970 verwendet worden ist. Dieser beschreibe die verminderte Vertrautheit des Menschen gegenüber synthetisch erstellten Figuren, wenn diese eine bestimmte Intensität des Anthropomorphismus (Menschenähnlichkeit) aufweisen [10].

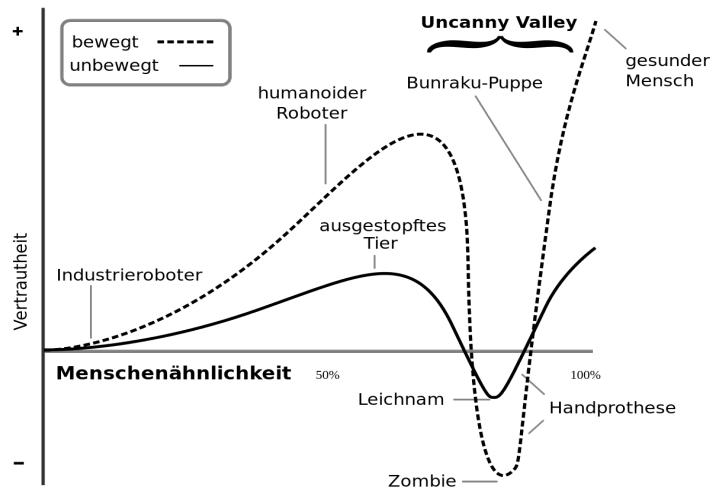


Abbildung 2.8: Darstellung des Uncanny Valley mit Einordnung verschiedener Figuren  
Abbildung von Mori 1970 –Übersetzung durch Tobias K.

Abruf 23.01.19: [https://de.wikipedia.org/wiki/Datei:Mori\\_Uncanny\\_Valley\\_de.svg](https://de.wikipedia.org/wiki/Datei:Mori_Uncanny_Valley_de.svg)

Wie in Abbildung 2.8 zu erkennen ist steigt die Akzeptanz des Betrachters gegenüber der visuellen Erscheinung zunächst mit steigender Menschenähnlichkeit an, bis sie abrupt ab einem bestimmten Realitätsgehalt in eine starke Befremdlichkeit übergeht. Anschließend steigt diese Vertrautheit jedoch wieder mit der Menschenähnlichkeit an. Bei bewegten Figuren ist dieser Effekt noch stärker zu beobachten. Der Bereich des starken Abfalls und Wiederanstieg der Vertrautheitskurve wird als Uncanny Valley bezeichnet [10].

Die Akzeptanzlücke gegenüber realistischen,künstlich entstandenen, visuellen Reizen, die beim Menschen von Mori beobachtet worden ist, lässt vermuten, dass die Befremdlichkeit gegenüber synthetisierten Szenen, sowohl visuell als auch auditiv, mit dem Realitätsgehalt der Medien zunehmen könnte.

Große Firmen der Unterhaltungsindustrie wie Sony und Oculus rieten Entwicklern daher schon auf der Game Developers Conference in Köln 2014, dass diese ihre VR- und AR-Anwendungen nicht zu realistisch gestalten dürfen, da der Effekt des Uncanny Valley drohe. Des Weiteren dürfe man die Blickrichtung des Benutzers nicht einschränken und müsse Phobien wie Klaustrophobie oder Höhenangst beachten. VR-Erlebnisse sollen demnach eher als Analogie zu einem Ritt durch einen Vergnügungspark begriffen werden als zu einem Kinofilm" [5].

Im Zuge dieser Arbeit werden Audioszenen betrachtet die extremen Realismus implizieren könnten. Es ist nicht auszuschließen, dass auch mit steigendem Realitätsgehalt auditiver Szenen bzw. räumlichem Audio eine Art Üncanny ValleyEffekt beim Hörenden entsteht. In Anbetracht dessen solltet es umso wichtiger sein, während des Hörversuchs, das Wohlempfinden der Probanden zu beobachten. Die Erforschung, ob es wirklich ein Art Üncanny Valley"für räumliches Audio gibt, ist ein wichtiges Thema für die Weiterentwicklung von AR und VR. Es übersteigt jedoch den Rahmen dieser Bachelorarbeit und wird deshalb nur angerissen.

# 3 Technische Grundlagen

Der im Rahmen dieser Arbeit durchzuführende Hörversuch beinhaltet einen typischen Anwendungsfall für eine AR-Anwendung. Eine virtuelle Welt wird mit Hilfe einer Microsoft Mixed Reality Brille auf einen realen Tisch projiziert. Es folgt die binaurale Wiedergabe von Schallereignissen über virtuelle Punktschallquellen im Raum, die es von den Probanden zu lokalisieren gilt. Ziel hierbei soll sein, die Lokalisationsfähigkeit und das Darbietungsempfinden der Probanden bei räumlichem Audio für einen realistischen Anwendungsfall zu analysieren. Dafür ist eine AR-Applikation entwickelt worden, mit der die Durchführung und Dokumentation des Hörversuchs automatisiert wird.

Die zur Erstellung der Applikation notwendigen technischen Grundlagen sowie die verwendete Hard- und Software sollen im Folgenden veranschaulicht werden.

## 3.1 Augmented Reality

Die Erweiterte Realität , engl. Augmented Reality (AR), ist eine Variation der Virtuellen Realität, engl. Virtual Reality (VR). Während bei VR-Anwendungen versucht wird den Benutzer komplett in einer synthetisierten Welt eintauchen zu lassen, geht es bei AR darum die physische Welt des Benutzers virtuell zu ergänzen [1].



Abbildung 3.1: Zeitschrift dessen Inhalt durch virtuelle Objekte ergänzt wird

Abruf 22.01.19:

<https://www.dpc-consulting.org/augmented-reality-aktuelle-beispiele-für-verlage-und-medienhäuser/>

Im Hörversuch dieser Bachelorarbeit sollen die Probanen in eine AR-Umgebung eintauchen, bei der der reale Raum mit einer audiovisuellen, virtuellen Welt überlagert wird.

### 3.1.1 Microsoft Mixed Reality

Microsoft verwendet im Zusammenhang mit AR und VR häufig den Begriff Microsoft Mixed Reality und bezeichnet auch ihre Head Mounted Displays als Microsoft Mixed-Reality Brillen. Nach der Dokumentation von Microsoft beinhaltet Microsoft Mixed Reality neben der virtuellen Erweiterung der Realität auch die Interaktion zwischen Mensch, Maschine und Umgebung [3].

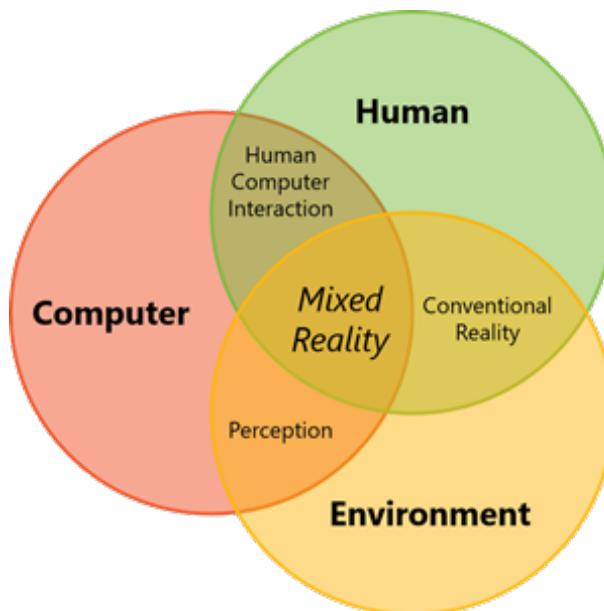


Abbildung 3.2: Zusammensetzung von Mixed Reality

Abruf 07.02.19: <https://docs.microsoft.com/en-us/windows/mixed-reality/mixed-reality>

Microsoft bietet mit Microsoft Mixed Reality Gestensteuerung, Spracherkennung, Kopfverfolgung (Head-Tracking), Kopfbezogene Übertragungsfunktionen(HRTF) und weitere Funktionen an. Dabei können sogar auditive und visuelle Inhalte an die physischen Gegebenheiten des Raums angepasst werden. Entwicklern werden diese Funktionen umsonst mit dem MixedRealityToolkit zur Verfügung gestellt, dessen Funktionalitäten wiederum in Unity3D importiert werden können. Auf diese Art und Weise können AR- und VR-Anwendungen erstellt werden, die sowohl räumliches Audio als auch eine an Mixed Reality Brillen angepasste Steuerung ermöglichen. Die wichtigsten verwendeten Funktionen des Toolkits und von Unity3D werden im Verlauf der Arbeit noch individuell betrachtet.

## 3.2 Microsoft HoloLens

Als Wiedergabegerät während des Hörversuchs wird eine Microsoft HoloLens verwendet, siehe Abbildung 3.3. Bei dieser handelt es sich um eine Microsoft Mixed Reality Brille, die mit einem Head-Mounted-Display, Kopfhörern, Head-Tracking-Funktion sowie einem Natural-User-Interface ausgestattet ist.



Abbildung 3.3: Mixed-Reality Brille - Microsoft HoloLens

Abruf 23.01.19:

<https://www.macerkopf.de/2016/10/12/microsoft-hololens-kommt-im-november-nach-deutschland>

### 3.2.1 Head-Mounted-Displays

Für die Umsetzung von AR- und VR-Anwendungen werden sogenannte Head-Mounted-Displays verwendet, welche dem Nutzer ermöglichen sollen möglichst realitätsnah visuelle Reize wahrzunehmen. Die Bildschirme in der Brille projizieren virtuelle Inhalte direkt in das Auge des Benutzers, so dass dieser denken soll Teil der virtuellen Umgebung zu sein. Bei AR-Brillen bzw. der Mixed Reality Brille Microsoft HoloLens wird hierbei ein transparenter Bildschirm verwendet, der es ermöglicht neben der virtuellen auch die reale Welt zu sehen.

### 3.2.2 Kopfhörer der HoloLens

Die Kopfhörer befinden sich an der Unterseite des Haltemechanismus der Brille. Sie sind über den Ohren positioniert, bedecken diese aber nicht. Auf diese Weise lassen sie die Umgebungsgeräusche ungedämpft durch. Dies ist vor allem dann wichtig wenn es um die Lokalisationsfähigkeit von Hörereignissen in der Umgebung des Nutzers geht. Wie bereits im Unterkapitel "räumliche Wahrnehmung" beschrieben wird, bewirkt die individuelle Außenohrform eine spektrale Verzerrung des Schall und ist für die spektrale Analyse zur Lokalisation von Hörereignissen von herausragender Bedeutung. Ohrenumschließende Kopfhörer sollten daher theoretisch für eine AR-Anwendung mit auditiver Überlagerung nur bedingt geeignet sein, da diese frequenzabhängig dämpfen und die Form der Ohrmuschel durch Druck verändern können.

### 3.2.3 Head-Tracking-Funktion

Die HoloLens bietet eine Head-Tracking-Funktion, die es erlaubt die Hörsignale an die Ausrichtung des Kopfes anzupassen. Dies ist wichtig, da die Rotation des Kopfes neben der offensichtlichen Veränderung des Blickfeldes auch eine Veränderung des an den Ohren an kommenden Schalls bewirkt. Die Analyse der damit einhergehenden Änderungen der interauralen Laufzeit- und Pegeldifferenzen tragen maßgeblich dazu bei Hörereignisse zu lokalisieren. Die für den Hörversuch entwickelte HoloLens-App erlaubt die freie Bewegung des Probanden im Raum und die

freie Bewegung des Kopfes. Auf diese Art und Weise sollte die Lokalisation von virtuellen Schallquellen im Vergleich zu statischen Wiedergabesystemen ohne Head-Tracking, stark vereinfacht werden und den Realismus der audiovisuellen Szenen verstärken.

### 3.3 Binaurale Wiedergabe

Das Problem der binauralen Wiedergabe räumlicher auditiver Szenen über Kopfhörer ist, dass die Beeinflussung des Schalls durch die Umgebung auf dem Weg zum Trommelfell nicht gegeben ist. Es werden also weder Schallreflexionen, noch die Filterwirkung des Körpers und der Außenohren auf den Schall wahrgenommen. Die Lokalisation von Schallereignissen über Kopfhörer ist dementsprechend nur dann möglich wenn die binauralen Signale angepasst werden.

#### 3.3.1 Kopfbezogene Übertragungsfunktionen - MS HRTF Spatializer

Die bereits im Kapitel "Menschliche Wahrnehmung" behandelten kopfbezogenen Übertragungsfunktionen beschreiben das System zwischen der Quelle eines Schallereignissen und der Senke, also dem Trommelfell. Die Faltung des Audio-Signals mit der zugehörigen HRTF verändert den eintreffenden Schall so, als ob er durch den Körper und die Außenohren gefiltert worden wäre.

Technisch gesehen wird räumliches Audio dadurch realisiert, dass je nach Position der virtuellen Schallquelle zum Kopf ein HRTF-Paar aus einer Datenbank ausgewählt und mit dem wiederzugebenden Audio-Signal digital gefaltet wird. Auf diese Art und Weise werden interaurale Laufzeit- und Pegeldifferenzen dem trockenen Signal hinzugerechnet, so dass auch über binaurale Wiedergabe eine Schallquellenlokalisierung stattfinden kann. Für die Applikation wurde der von Microsoft entwickelte MS HRTF Spatializer verwendet, der ebenfalls nach diesem Prinzip arbeitet [8].

Damit die Lokalisationsfähigkeit bei binauraler Wiedergabe vollständig gegeben ist muss für jeden Menschen ein individueller Datensatz an HRTFs verwendet werden. Dies spiegelt sich besonders im Bereich der Lokalisation hochfrequenter Schallereignisse wieder, da hier, durch die individuelle Außenohrform, die HRTFs am stärksten variieren. Ein weiterer Ansatz der personalisierten HRTF-Datensätze besteht darin über neuronale Netze und Methoden des Machine Learnings die HRTF-Datensätze automatisch an den Benutzer anzupassen [8].

Da die Aufnahme individueller HRTF-Datensätze für jeden Probanden der Versuche den Rahmen dieser Bachelorarbeit übersteigen würde, wird für die Durchführung der Versuche ein einzelner HRTF-Datensatz verwendet. Bei der Auswertung der Versuchsergebnisse muss daher berücksichtigt werden, dass der genutzte HRTF-Datensatz unterschiedlich stark mit den wirklichen HRTF-Datensätzen der Probanden korreliert. Dies gilt besonders bei der Betrachtung hochfrequenter Schallereignisse über 7kHz.

#### 3.3.2 Auralisation

Auralisation beschreibt Verfahren die es ermöglichen Schallsignale so zu verändern, dass es so klingt als wären diese in einem bestimmten Raum aufgenommen worden. Die Faltung eines Audio-Signals mit der monoaural-aufgenommenen Raumimpulsantwort ermöglicht es den individuellen Klang des gewünschten Raumes in das Schallsignal zu übertragen und so ein plausible

Schallereignisse zu erzeugen [12, S.521].

Für AR-Anwendungen in Echtzeit ist dieser Ansatz jedoch eher bedingt geeignet, da sich die Position des Betrachters während der Wiedergabe verändern kann und damit auch eine Anpassung der Raumimpulsantwort in Echtzeit notwendig wäre.

## **Microsoft - Spatial Audio**

Microsoft bietet mit Spatial Audio aus dem MixedRealityToolkit eine alternative Möglichkeit für die Entwicklung von AR-Anwendungen an, die Audio-Signale an die Umgebung anzupassen. Diese arbeitet mit drei Instanzen, den sogenannten Audio Emittern, den Audio Occludern und einem Audio Listener.

Spatial Audio nutzt den Effekt, dass hohe Frequenzen des Schalls von Objekten stärker absorbiert werden als niedrige Frequenzen. Aus diesem Grund werden Wände und andere Objekte die Einfluss auf die Schallausbreitung haben mit Audio Occludern versehen. Diese Skripte steuern Tiefpassfilter mit einer bestimmten Grenzfrequenz und Lautstärkeanpassung.

Die Audio Occluder arbeiten Hand in Hand mit den Audio Emittern und dem Audio Listener. In Unity wird jeder Schallquelle ein Audio Emitter zugewiesen, während der Audio Listener nur dem Kamera-Objekt, also dem Nutzer des Head-Up-Displays, übergeben wird. Bei Wiedergabe eines Audio-Signals werden von den Audio Emittern Strahlen an die Umgebung abgegeben. Trifft ein Strahl einen Audio Occluder, dann werden von diesem aus ebenfalls Strahlen ausgesendet, die Informationen über die Tiefpasscharakteristik des Objekts enthalten. Beim Audio Listener werden die ankommenden Strahlen ausgewertet. Treffen Strahlen direkt vom Audio Emitter aus dem Audio Listener dann wird dieser als Direktschall gewertet und das Audio-Signal unverändert beim Betrachter wiedergegeben. Bei Strahlen, die von einem Audio Occluder aus gesendet worden sind, wird das Audio-Signal tiefpassgefiltert und lautstärkeangepasst wiedergegeben. Auf diese Weise kann diffuser Nachhall imitiert werden [9].

Neben den offensichtlich in der Unity-Szene enthaltenden Objekten, die den Schall mittels Occluder beeinflussen können, müssen bei AR-Anwendung für eine plausibles Hörergebnis die im Raum enthaltenden Objekte wie den Boden, Wände, Tische und Schränke mit Occludern versehen werden. Hierbei gibt es zum einen die Möglichkeit mit Spatial Mapping, eine Funktion aus dem MixedRealityToolkit die unmittelbare Umgebung direkt von der HoloLens scannen zu lassen und sich als 3D-Modell ausgeben zu lassen. Die andere Möglichkeit besteht darin ein eigenes Raummodell mittels 3D-Software zu entwerfen. In beiden Fällen entsteht ein 3D-Modell, welches sich in die Unity-Szene integrieren lässt und durch einen Audio Occluder ergänzen lässt.

Für die Applikation des Hörversuchs wurde der zweite Ansatz gewählt, da der Raum zu viele kleine Objekte enthält, bei denen der Scanner der HoloLens Probleme bei der Erstellung eines realitätsnahen Raummodells bekam.

## 3.4 Software

Für die Umsetzung des Hörversuchs der Bachelorarbeit wurde eine Applikation für die Microsoft HoloLens entwickelt. Im Folgenden werden die dafür verwendeten Programme kurz vorgestellt, sowie erklärt welche Funktion diese bei der Erstellung des Hörversuchs-Applikation gespielt haben.

### 3.4.1 Blender

Blender ist eine freie, Open-Source 3D-Grafiksuite der Blender Foundation zur Erstellung von visuellen Inhalten mittels Animations-, Texturierungs- und Modellierungstools. In den Versuchen dieser Arbeit sollen auditive Szenen visuell unterstützt werden. Die dafür erstellten Animationen wurden alle mit Blender erstellt.

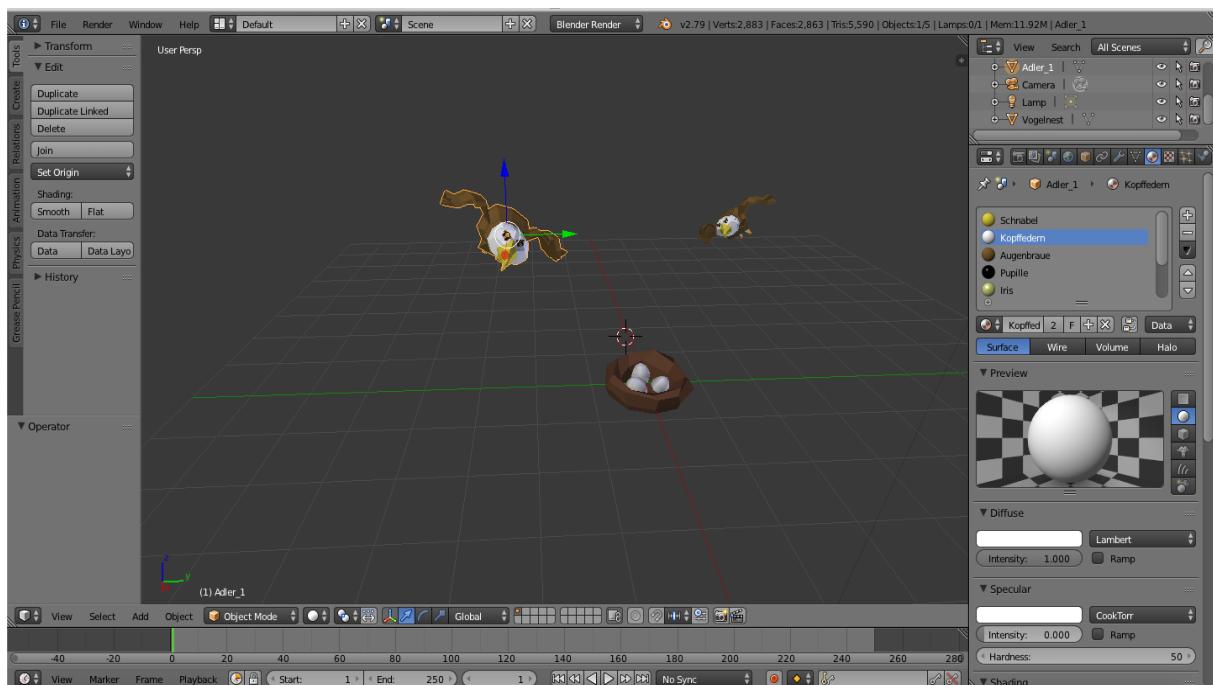


Abbildung 3.4: Blendersoberfläche bei der Erstellung von 3D-Modellen für die Applikation

### 3.4.2 Unity

Unity ist eine Laufzeit- und Entwicklungsumgebung für Spiele, die vom Unternehmen Unity Technologies entwickelt worden ist. Unity bietet Funktionen, die schon seit langem von Unternehmen für professionelle Visualisierungen genutzt werden. Im Rahmen dieser Arbeit bietet Unity die Schnittstelle für die Nutzung der Microsoft HoloLens und den erstellten Animationen (Blender). Des Weiteren wird über Unity ein Userinterface für die Durchführung der Versuche entwickelt, welches die Effektivität bei der Durchführung der Versuche steigern soll.

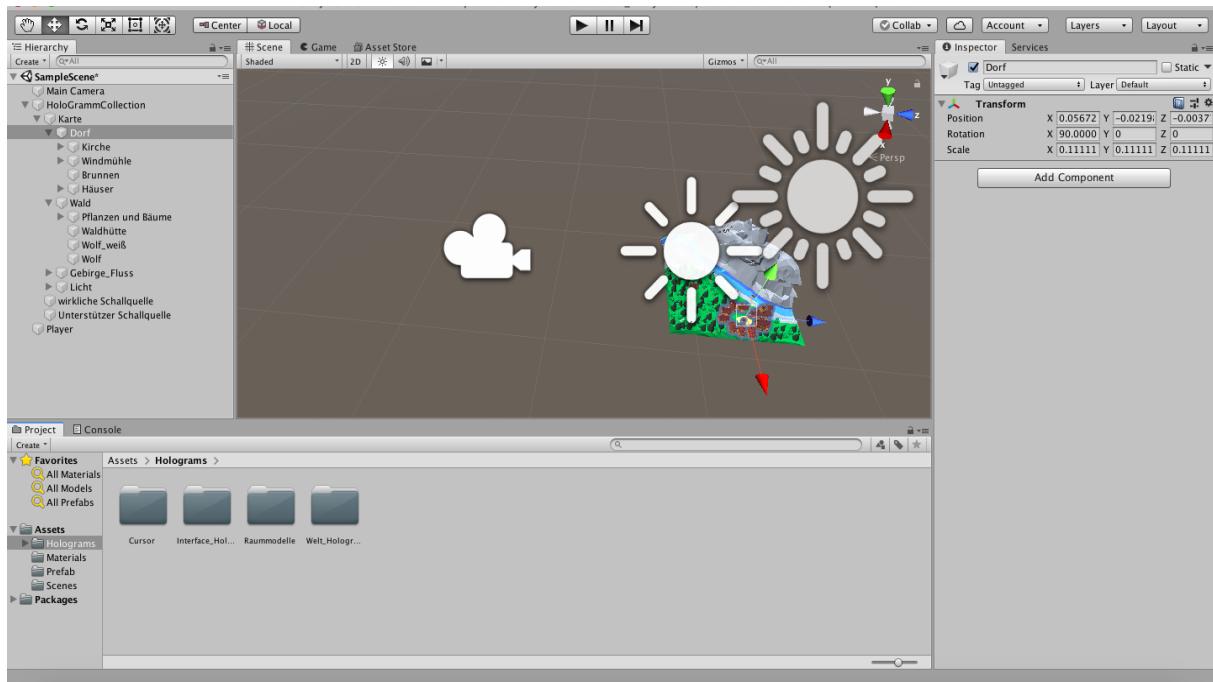


Abbildung 3.5: Unityoberfläche bei der Erstellung der Applikation

# **4 Vorstellung der HoloLens-Applikation und des Hörversuches**

Ziel dieser Arbeit ist es zu überprüfen wie gut bei VR- und AR-Anwendung virtuelle Schallquellen mit räumlichem Audio lokalisiert werden können und wie 3D-Audio bei solchen Anwendungen allgemein angenommen wird. Hierfür wurde ein Hörversuch in einer Mixed-Reality-Umgebung durchgeführt. Die dafür verwendete HoloLens-Applikation wurde mit Unity entwickelt. Im folgenden sollen die App und der Versuchsaufbau vorgestellt werden.

## **4.1 Idee hinter dem Versuch**

Bei dem im Hörversuch nachgestellten Szenario handelt es sich um einen typischen Anwendungsfall für Augmented Reality. Über das Head-Up-Display der HoloLens werden virtuelle Objekte auf einen Tisch vor dem Nutzer projiziert. In der Industrie findet sich so ein Anwendungsfall zum Beispiel in der Vorstellung von virtuellen Prototypen. Für die Vorstellung solcher Prototypen wird neben dem Visuellen auch die auditive Komponente immer wichtiger.

Hierbei wird von Software, wie z.B Unity, die Möglichkeit geboten virtuelle Schallquellen an bestimmten Positionen im Raum zu platzieren. Diese sollen dann nach Möglichkeit so klingen als wären diese tatsächlich an der angegebenen Position im Raum. Bei der Vorstellung eines Motorprototypen über AR könnte so z.B eine visuelle aber auch akustische Kostprobe gegeben werden.

Wie plausibel diese Tools virtuelle Schallquellen in den Raum projizieren und ob die Konsumenten diese Darbietung überhaupt so annehmen wie gewünscht, soll anhand eines Hörversuches ermittelt werden.

## 4.2 Beschreibung der Virtuellen Welt

In dem Versuch wird eine virtuelle Welt im lowpoly-Grafikstil auf einen Tisch vor den Probanden projiziert. Es handelt sich hierbei um eine Landschaft mit einem Gebirge, einem Fluss, einem Dorf und einem Waldgebiet. Die Welt ist ca. 1m breit wie tief und bis zu 70cm hoch. Im Folgenden werden die einzelnen Gebiete genauer beschrieben. Die dort platzierten Hörereignisse werden in einem einzelnen Kapitel vorgestellt.



Abbildung 4.1: Hologramm der virtuellen Welt von oben  
Rechts, über der Welt und dem Probanden, befindet sich zusätzlich ein weiterer Adler.

### 4.2.1 Dorf

Das Dorf befindet sich am Rand des Waldes und wird durch den Fluss vom Gebirge abgetrennt. Im Zentrum des Dorfes befindet sich ein Brunnen, der auf einer Sandfläche steht. Darum herum sind sieben einfache Häuser, sowie eine Mühle und eine Kirche platziert. Die Kirche besitzt einen Glockenturm mit einer großen, goldenen Glocke. Diese könnte als potentielle virtuelle Schallquelle verwendet werden, wobei das Hörereignis der erschallenden Kirchenglocken ohne räumliches Audio wiedergegeben wird und somit keine Position in der Welt einnimmt.

TODO: Besseres Bild machen.



Abbildung 4.2: Dorf: Die Kirchenglocke dient als Ort für potentielle Schallquelle

#### 4.2.2 Wald

Der Wald besteht aus zum größten Teil aus Pflanzen und Bäumen. Im Wald befindet sich noch ein einfaches Haus und zwei Wölfe, wobei einer weiß und einer grau ist. Zu Beginn des Hörversuchs befindet sich der Wald im Vordergrund. Der weiße Wolf sticht hierbei durch seine Farbe, seine Größe und seine Position heraus, während der graue Wolf zunächst eher durch Bäume und Büsche verdeckt wird. Eine virtuelle Schallquelle wird an der Position des grauen Wolfes platziert.



Abbildung 4.3: Wald: Schwarzer Wolf dient als Ort für Schallquelle

### 4.2.3 Gebirge und Fluss

Das Gebirge ist das höchste Gebiet der Welt und wird vom Fluss von der restlichen Welt abgegrenzt. Die Bergspitze des größten Berges ist hierbei ca. 70cm hoch. In dem Gebirge befindet sich ein Adlernest, über dem zwei Adler fliegen. Eine virtuelle Schallquelle wird im Adlernest platziert.

Ein weiterer Adler befindet sich weit über der Welt, über dem Probanden. An dieser Position befindet sich ebenfalls eine virtuelle Schallquelle. Diese spielt das gleiche Signal ab wie die virtuelle Schallquelle im Adlernest, siehe Abbildung 4.1.

Am Fuße des Gebirges befindet sich ein Fluss. Dieser fließt von einer erhöhten Quelle abwärts in Richtung des Dorfes und geht hierbei von einem Kante der Karte bis zur gegenüberliegenden.

An der Stelle wo der Fluss an das Dorf angrenzt befindet sich ein Steg mit einem kleinen Boot. Eine virtuelle Schallquelle wird an der Quelle des Flusses platziert und eine weitere im Boot. Beide virtuellen Schallquellen spielen das gleiche Audio-Signal ab.

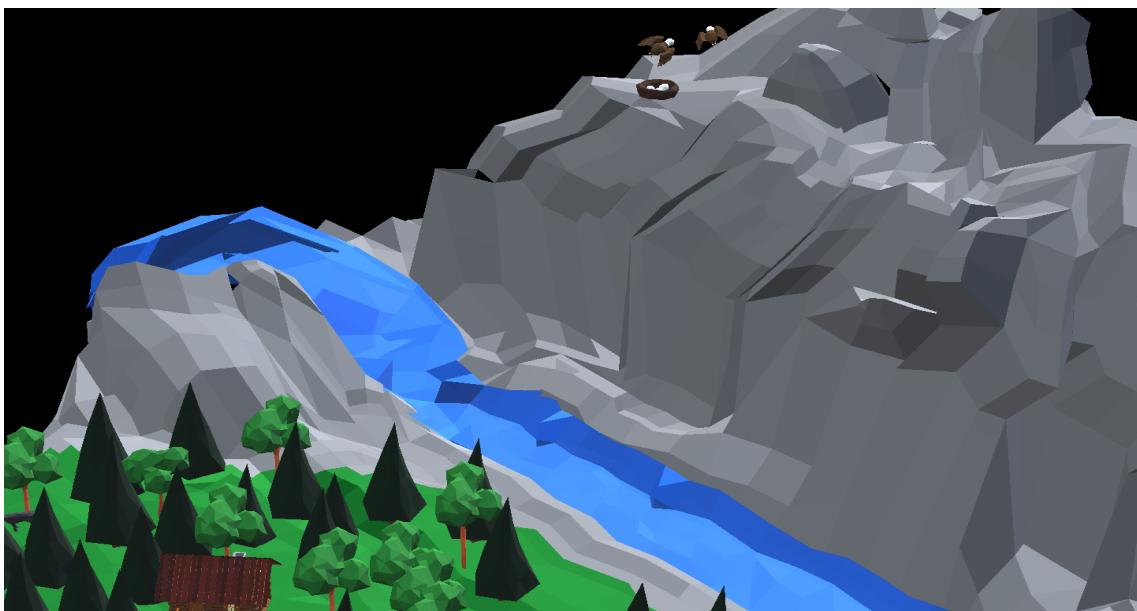


Abbildung 4.4: Gebirge und Fluss: Adler, Adlernest und Fluss dienen als Ort für Schallquellen

## 4.3 Audio-Einstellungen in der Applikation

Die virtuelle Schallquelle ist in Unity ein Audio Source - Objekt. Die Einstellungen für das Audio Source Objekt sind nach dem Microsoft Tutorial für Spatial Audio für Mixed Reality vorgenommen worden [9]. Diese Einstellungen gelten im Folgenden als Standart-Einstellungen. Abweichungen der Einstellungen werden in den Beschreibung für die Einzelszenen genannt. Bei den abgespielten Audio-Signalen handelt es sich ausschließlich um MP3 codierte Geräusche die mit 48 kHz abgetastet werden.

TODO: Bild der Einstellungen aus Unity-Screenshot oder Liste ausführen

**Update Intervall:** 0.15

**Max Distance:** 20

**Max Objects:** 10

### 4.3.1 Binauralisierung in der Applikation

Für das räumliche Audio wird der MS HRTF Spatializer und damit ein einheitlicher HRTF-Datensatz verwendet. Der MS HRTF Spatializer berechnet für die Szene jeweils die benötigten kopfbezogenen Übertragungsfunktionen für jedes Ohr des Probanden. Auf diese Weise gelingt eine binaurale Wiedergabe der Audio-Signale über die Kopfhörer der HoloLens.

### 4.3.2 Auralisierung in der Applikation

Für eine plausible Wiedergabe wird eine Anpassung der Geräusche an den Versuchsraum mittels "Microsoft Spatial Audio" vorgenommen. Hierfür wird ein abstraktes Raummodell des Versuchsraumes entworfen und in die Unity-Szene integriert 4.5. Dem Raummodell wird anschließend ein Audio OccluderScript hinzugefügt. Microsoft empfiehlt in seinem Tutorial bei den Parametern des Audio Occluders eine Tiefpass-Grenzfrequenz von 1500Hz und eine Lautstärkereduzierung um den Faktor 0.75 einzustellen, um plausiblen diffusen Nachhall zu imitieren[9].

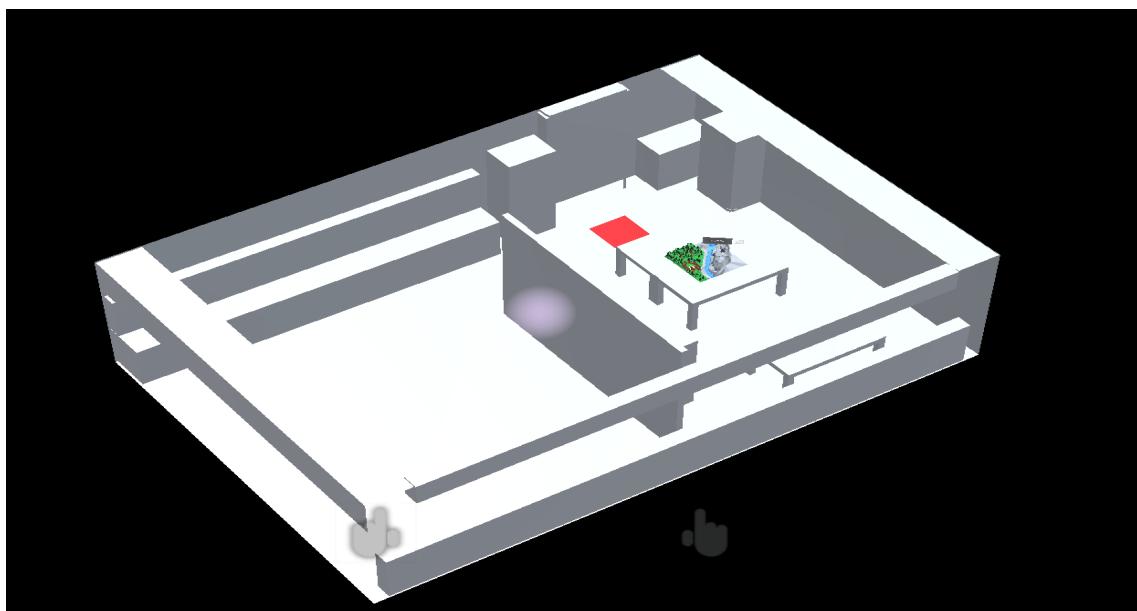


Abbildung 4.5: Abstraktes Raummodell des Versuchsraumes E113 der HS Emden/Leer

Jedem Objekt mit einer Audio Source muss zusätzlich ein Audio Emitter zugewiesen werden. Die Parameter der Emitter werden wie folgt festgelegt:

**Update Intervall:** 0.15

**Max Distance:** 20

**Max Objects:** 10

Durch die gewählten Parameter wird alle 0.15 Sekunden eine Neuberechnung der Raumakustik vorgenommen, was für die langsamen Bewegungen des menschlichen Kopfes genügen sollte.

Zusätzlich werden nur maximal 10 Objekte das Audio-Signal beeinflussen können. Diese dürfen dabei eine Distanz von maximal 20m entfernt sein. Auf diese Art und Weise kann sichergestellt werden, dass die Anpassung der Audio-Signale einen natürlichen Klang hervorbringt.

## 4.4 Beschreibung der Szenen des Hörversuchs

Der Hörversuch umfasst fünf verschiedene Szenen. In jeder Szene gibt es ein Hörereignis, das der Proband zu lokalisieren hat. Die Welt ist statisch und verändert sich während des Versuches nicht. Die virtuelle Schallquelle wird für jede Szene neu positioniert.

### 4.4.1 Szene 1 - Adlerschrei im Gebirge

In der ersten Szene befindet sich die virtuelle Schallquelle im Adlernest, mitten im Gebirge. Bei dem Geräusch handelt es sich um das Schreien eines Adlers. Die Einstellungen weichen nicht vom gewählten Standard ab.

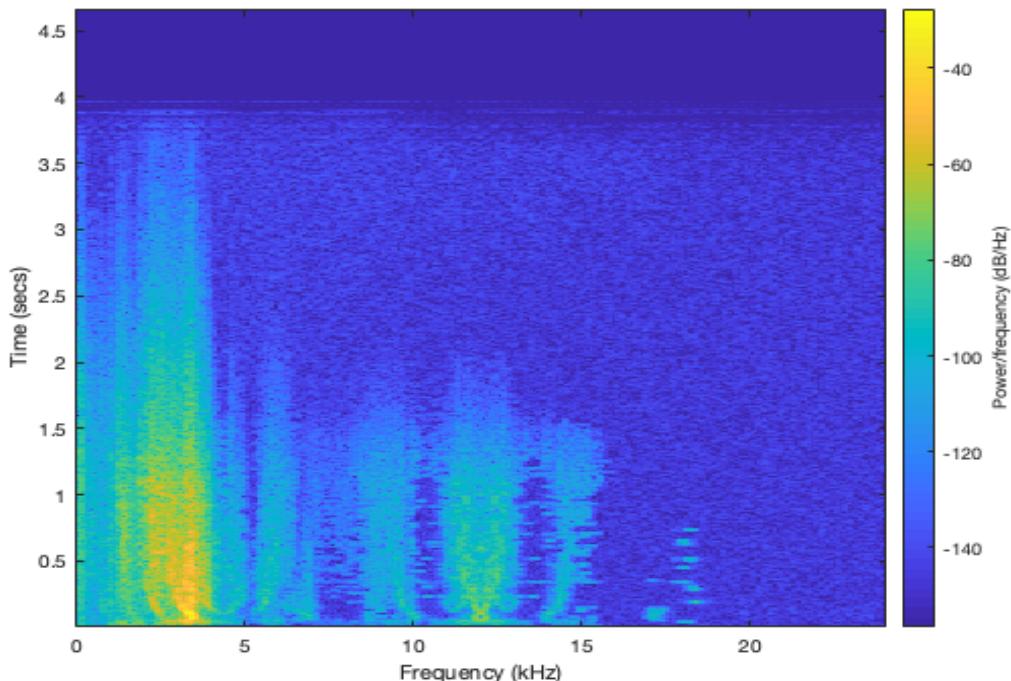


Abbildung 4.6: Spektrogramm des Adlerschreis

Abbildung 4.6 zeigt das Spektrogramm des Adlergeräusches. Es ist zu erkennen, dass das Geräusch ein breites Frequenzspektrum besitzt, jedoch die meisten Frequenzen zwischen 2 und 4 kHz liegen. Über den zeitlichen Verlauf lässt sich sagen, dass das Signal sich vor allem durch ein kurzes intensives Geräusch am Anfang auszeichnet, dass im Frequenzbereich zwischen 2 kHz und 4 kHz noch einige Sekunden ausklingt während die höheren Frequenzen schneller nachlassen.

Der Schrei eines Adlers ist ein Geräusch, dass vermutlich jedem Probanden aus Fernsehen oder anderen Medien bekannt sein sollte. Gleichzeitig werden die meisten Probanden diesem Geräusch vermutlich noch nie in ihrem echten Leben begegnet sein.

#### 4.4.2 Szene 2 - Wolfsheulen

In der zweiten Szene befindet sich die virtuelle Schallquelle auf dem grauen Wolf im Wald. Das Geräusch beinhaltet das Heulen eines einsamen Wolfes. Das Spezielle an der Szene ist, dass es im unmittelbaren Blickfeld des Probanden eine alternative, potentielle Schallquelle mit dem weißen Wolf gibt. Dieser wird in der Regel vom Betrachter auf Grund seiner Größe, seiner Erscheinung und seiner ungeschützten Position leichter entdeckt. Interessant wird es daher zu beobachten, ob die Probanden in der Lage sind, auf Grund des Geräusches, den schwarzen Wolf zu erblicken und diesen als virtuelle Schallquelle zu bestimmen. Die Einstellungen für die Wiedergabe weichen erneut nicht vom gewählten Standard ab.

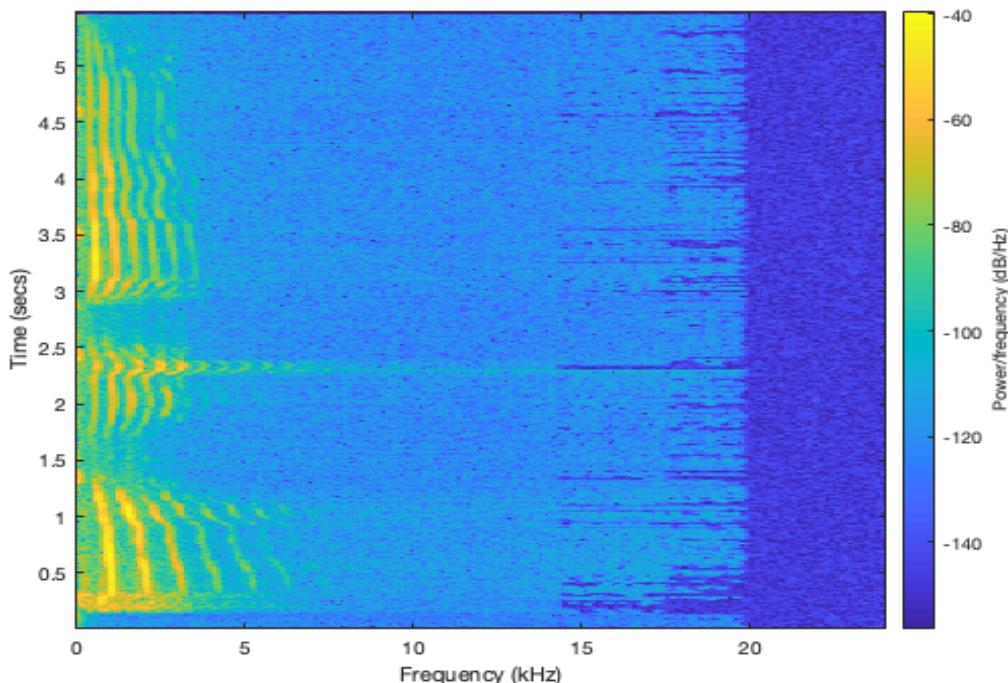


Abbildung 4.7: Spektrogramm des Wolfsheulens

In Abbildung 4.7 ist das Spektrogramm des Wolfsheulens zu sehen. Es ist zu erkennen, dass es sich in regelmäßigen, zeitlichen Intervallen auf ähnliche Art und Weise wiederholt. Das Geräusch besteht hauptsächlich aus Frequenzen zwischen 1 Hz und 5 kHz. Auffällig ist dabei, dass es zur Bildung schmäler Frequenzkeulen kommt, die sich bei vielfachen Frequenzen wiederholen. Es ist daher anzunehmen, dass sich das Geräusch harmonisch anhört.

Das Heulen eines Hundes könnte aufgrund der Verbundenheit des Menschen zu ihren Hunden ein Gefühl von Empathie auslösen. Wird das Geräusch eher als Wolfsheulen wahrgenommen, so

könnte dieses Geräusch eher ein Angstgefühl beim Probanden bewirken. In beiden Fällen sollte das Geräusch ein befremdliches Gefühl auslösen und nicht als angenehm empfunden werden.

#### 4.4.3 Szene 3 - Kirchenglocken

Das Erklingen von Kirchenglocken in der dritten Szene wird ohne räumliches Audio durchgeführt und soll mit als Referenz dienen. Aus diesem Grund gibt es in dieser Szene keine virtuelle Schallquelle mit festgelegter Position. Es ist bei dieser Szene zu beobachten, ob die Probanden die Darbietung trotz des fehlenden, räumlichen Audios als plausibel und angenehm empfinden. Außerdem stellt sich die Frage, ob die Schallquelle rein auf Grund der visualisierten Kirchenglocke auch dort von den Probanden lokalisiert werden wird.

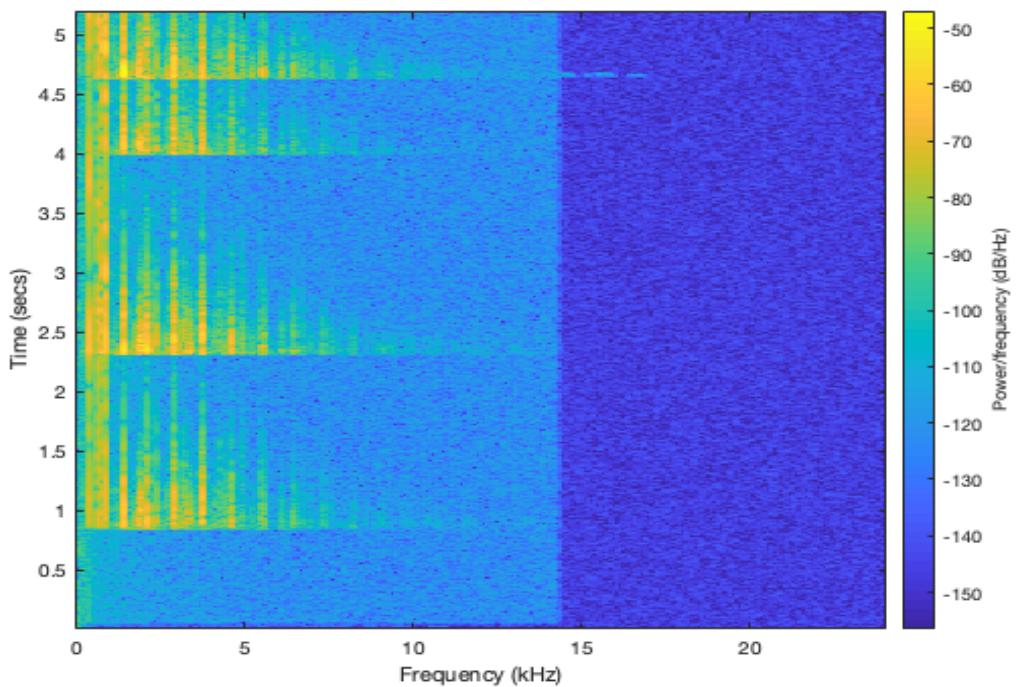


Abbildung 4.8: Spektrogramm der Kirchenglocken

Anhand Abbildung 4.8 lässt sich erkennen, dass das Geräusch der erklingenden Kirchenglocken ebenfalls eindeutige Frequenzkeulen aufweist, die sich bei vielfachen Frequenzen wiederholen. Aus diesem Grund handelt es sich ebenfalls um ein harmonisches Klangereignis. Das Frequenzspektrum umschließt Frequenzen zwischen 100 Hz und 10 kHz, wobei besonders die tiefen Frequenzen zwischen 100 Hz und 1kHz sehr dominant sind.

Zeitlich verhält sich das Audio-Signal so, dass es immer zu einem wiederholenden Anschlag der Glocken neu ertönt. Hierbei lassen die hohen Frequenzen ebenfalls schneller nach als die tieferen. Besonders auffällig ist dabei, dass die sehr tiefen Grundtöne zwischen zwei Glockenschlägen nicht vollends abklingen, so dass diese Frequenzen dauerhaft zu hören sind.

Das erklingen von Glocken sollte in der Regel als angenehmes Geräusch wahrgenommen werden und wird häufig auch mit einem Friedensgefühl assoziiert.

#### 4.4.4 Szene 4 - Flussrauschen

In der vierten Szene des Hörversuchs werden im Vergleich zu den anderen Szenen zwei virtuelle Schallquellen im Raum platziert, die beide das gleiche Geräusch wiedergeben. Bei diesem handelt es sich um das Rauschen eines Flusses. Eine virtuelle Schallquelle wird hierbei an der Quelle des Flusses, im Gebirge und eine neben dem Boot, am anderen Ende des Flusses platziert. Je nach Position des Probanden dürfte sich eine Phantomschallquelle entweder zwischen den beiden Schallquellen oder an der nächstgelegenen Schallquelle bilden.

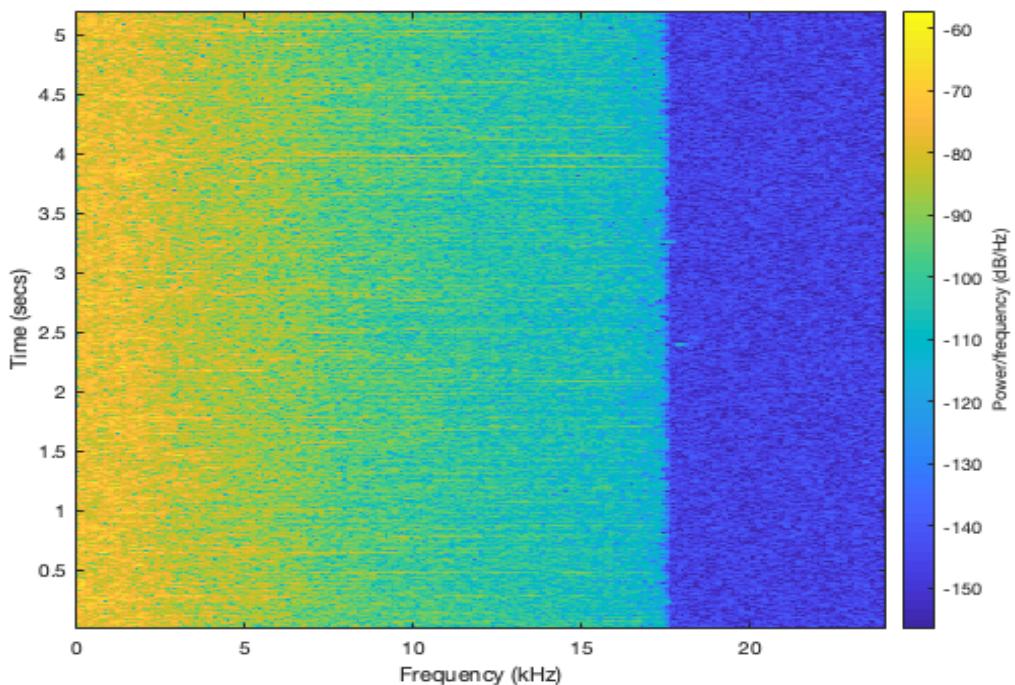


Abbildung 4.9: Spektrogramm des Flussrauschenes

Das Spektrogramm des Flussrauschenes in Abbildung 4.9 ähnelt sehr dem eines rosa Rauschens. Von den tiefen zu den hohen Frequenzen nimmt die Intensität ab. Die beruhigende Wirkung von rosa Rauschen könnte dementsprechend auch dem Flussrauschen zugesprochen werden.

Auch inhaltlich sollte das Rauschen eines Flusses eher als angenehm wahrgenommen werden.

#### 4.4.5 Szene 5 - Adlerschrei über der Welt

In der letzten Szene wird die virtuelle Schallquelle bei dem Adler über dem Probanden platziert. Das Geräusch ist hierbei das selbe wie bei der ersten Szene.

In der Realität handelt es sich bei dem Adlerschrei um eine Geräusch, dass eigentlich über dem Hörenden erwartet wird. Dadurch, dass sich das Geräusch nun tatsächlich über dem Probanden befindet müsste es eigentlich ein plausibles Hörereignis darstellen. Es bleibt jedoch abzuwarten ob die Probanden anhand des Schalls den Adler über sich überhaupt lokalisieren können oder ob die offensichtlicheren Adler im Gebirge von der eigentlichen virtuellen Schallquelle ablenken.

## 4.5 Bewertungskriterien der Darbietung

Zur Bewertung der Darbietung werden dem Probanden Fragen zum Darbietungsempfinden gestellt. Dieser hat der Proband mit Hilfe von jeweils zwei dualistischen Antwortmöglichkeiten zu beantworten. Dabei kann der Proband in zehn Stufen einordnen, welche der beiden Antworten für ihn eher zu trifft. Diese Art der Befragung erlaubt es den Probanden bei der Bewertung ihren Freiraum zu lassen, aber gleichzeitig die Probandendaten untereinander vergleichbar zu machen.

Mit Hilfe der Fragen soll herausgestellt werden, ob der Proband die Szene als plausibel wahrnimmt, inwiefern bei ihm ein Immersionsgefühl zustande kommt und ob er sich bei der Darbietung wohlgefühlt hat.

Wird eine Szene als plausibel bezeichnet, dann bedeutet dies, dass die Summe aller Sinneseindrücke für den Betrachter glaubhaft und nachvollziehbar ist. Dies wird vor allem dann erreicht wenn die Sinneseindrücke natürlich wirken und den Konsumenten nicht verwirren.

Ein wichtiges Kriterium für Plausibilität ist, dass die wichtigsten Sinneseindrücke, also das Gehörte und das Gesehene, zusammen passen. Ist beispielsweise ein Hörereignis ohne ersichtlichen Grund weit von der Schallquelle entfernt, dann widerspricht dies dem natürlich Verständnis des Menschen für Kausalität. Mit der ersten Frage soll daher geprüft werden, ob für den Probanden das Visuelle und das Auditiv plausibel zusammenarbeiten.

Ein weiteres Kriterium Plausibilität ist der Realismus der Sinneseindrücke. Hierbei geht es weniger darum, dass visuelle und auditive Reize als komplett realistisch wahrgenommen werden, sondern vielmehr darum, dass diese sich in das Gesamtbild einfügen. Da das Visuelle in dem Hörversuch auf einem Low-Poly-Grafikstil basiert, wird niemand von den Geräuschen absoluten Realismus erwarten. Wichtig ist jedoch trotzdem, dass die Geräusche als Komponente der unmittelbaren Umgebung wahrgenommen werden, anstatt als alleinstehendes Audio-Signal, welches über die Kopfhörer wiedergegeben wird. Die zweite Frage soll daher Klarheit darüber geben, inwiefern die Probanden die Geräusche als realen Bestandteil ihrer Umgebung wahrnehmen.

Bei der Erstellung von AR- und VR-Anwendungen ist das Wohlbefinden des Konsumenten ein zu beachtender Faktor. Das stärkere Immersionsgefühl bei AR und VR ermöglicht es noch realistischere, mediale Inhalte zu erstellen. Dies setzt aber auch voraus darauf zu achten unter welchen Umständen sich der Konsument wohlfühlt. Mit der Entdeckung des Uncanny Valley wurde bereits bewiesen, dass zu realistische visuelle Reize Unwohlsein auslösen können. Inwiefern dies auch mit dem steigernden Realismus von räumlichem Audio bewirkt werden könnte ist noch nicht vollends erforscht. Aus diesem Grund soll mit der dritten Frage das Wohlbefinden

des Probanden während der Darbietung überprüft werden.

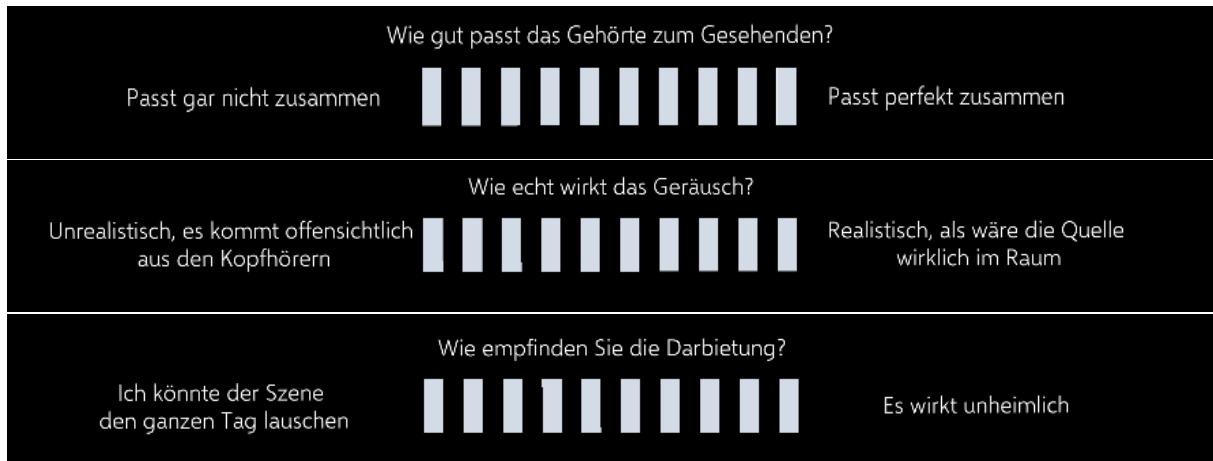


Abbildung 4.10: Fragen zur Bewertung der Szenen im Hörversuch

Anhand der Bewertungen des Probanden soll ermittelt werden, wie stark das Immersionsgefühl bei der Darbietung der Szenen gewesen ist. Dabei wird davon ausgegangen, dass die gewichtete Antwort der ersten beiden Fragen mit dem Immersionsgefühl des Probanden korreliert. Nimmt ein Proband das Geräusch als Teil des Raumes wahr und empfindet gleichzeitig das Zusammenspiel aus Visuellem und Auditivem als plausibel, dann ist auch davon auszugehen, dass das Immersionsgefühl des Probanden recht intensiv ist. Auf diese Weise kann überprüft werden bei welcher Szene das stärkste Immersionsgefühl bewirkt wird.

Im Zweifel wegfallen lassen: Um das Immersionsgefühl statistisch erfassen zu können, wird der Durchschnitt der gewichteten Antworten für die ersten beiden Fragen im Folgenden als Immersionsgefühl festgehalten.

Die zu Anfang gestellte These, dass es eine Art Uncanny-Valley-Effekt für zu realistisches Audio geben könnte, soll anhand der Versuchsergebnisse ebenfalls überprüft werden. Lässt sich ein Zusammenhang zwischen dem Immersionsgefühl bzw. dem bewerteten Realismus der Szene aus Frage 2 und dem geprüften Wohlbefinden aus Frage 3 beobachten, so könnte dies Aufschluss über die gestellte These geben. Natürlich sind die durchgeführten Versuche für die Bestätigung der These nicht vollends aussagekräftig, da der Hörversuch nicht speziell auf die Erforschung dieses Themengebiets ausgelegt ist.

## 4.6 Beschreibung der Benutzeroberfläche

Die Bedienung der HoloLens-App funktioniert über Head-Tracking, Gesten- und Sprachsteuerung, wobei die Probanden für den Hörversuch ausschließlich auf die Head-Tracking-Funktion und die Gestensteuerung zurückgreifen mussten.

Mittels eines Cursors, der wie eine Computermaus funktioniert, lassen sich in der App Aktionen durchführen. Der Cursor befindet sich hierbei immer in der Mitte des Sichtfeldes des

HoloLens-Nutzers. Über die Bewegung des Kopfes kann der Benutzer nun den Cursor bewegen. Der Cursor besitzt drei Darstellungsformen, um dem Nutzer seinen aktuellen Zustand zu signalisieren, siehe Abbildung 4.11.



Abbildung 4.11: Cursor-Zustände: Default(links), Hand erkannt (Mitte), Tap erkannt (rechts)

Der Cursor lässt sich über Handgesten steuern. Hebt man die Hand vor sich hoch, so dass die Sensoren der HoloLens diese erkennen können, dann verändert sich der Cursor in eine Hand, siehe Abbildung 4.11. Hebt man von der geschlossenen Hand aus den Zeigefinger und drückt diesen ruckartig wieder nach unten, wird dies als Tap-Geste bezeichnet. Wurde die Tap-Geste erfolgreich von der HoloLens erkannt, dann verändert sich der Cursor erneut, siehe 4.11. Die Tap-Geste ist ein Mittel um mit Objekten in der virtuellen Welt zu interagieren.

Auch die Beantwortung der Fragen zur Bewertung der Darbietung gelingt über die Tap-Geste. In dem der Proband auf einen der Balken zwischen den Antwortmöglichkeiten klickt kann dieser die Szene bewerten.

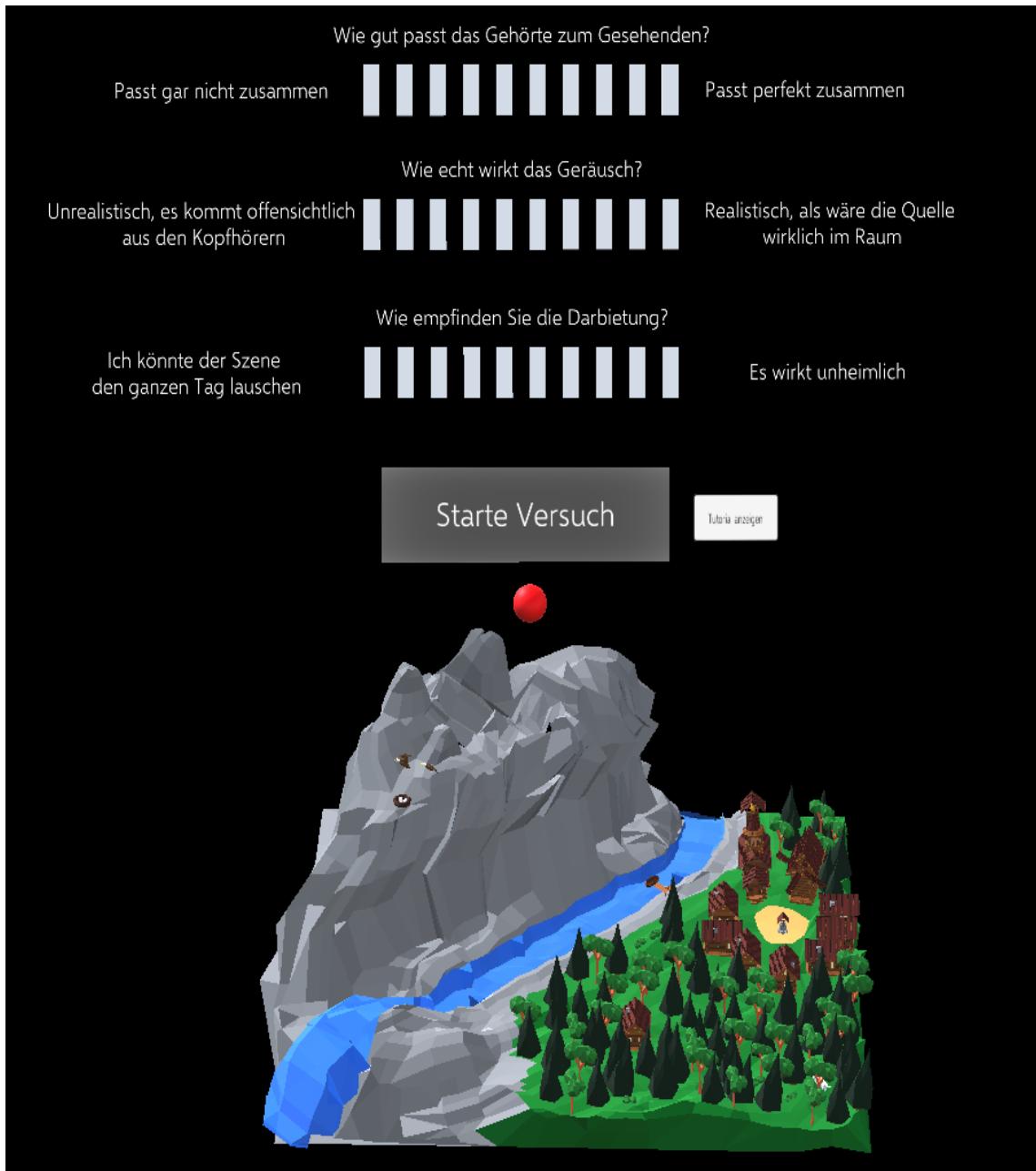


Abbildung 4.12: Welt mit Benutzeroberfläche und roter Kugel

### Position der Schallquelle lokalisieren

Eine Aufgabe des Probanden im Hörversuch ist es die Position der virtuellen Schallquelle zu lokalisieren. Damit dies spielerisch gelingt hat der Proband in der HoloLens-App die Möglichkeit eine rote Kugel dort zu platzieren, wo er denkt, dass sich die virtuelle Schallquelle befindet. Die rote Kugel befindet sich am Anfang jeder Szene direkt über der virtuellen Welt, siehe Abbildung ???. Der Proband hat nun die Möglichkeit die Kugel mit dem Cursor anzuvisieren und mit der Tap-Geste die Kontrolle über die rote Kugel zu erhalten. Hat er die vermeintliche Position der virtuellen Schallquelle lokalisiert, kann er mit Hilfe der Kopfbewegung die rote Kugel dorthin schieben und mit einer weiteren Tap-Geste an der gewünschten Position platzieren. So kann

der Proband leicht und sofort visualisiert angeben, wo er die Schallquelle hört.

### **Sprachbefehle und Positionierung der Welt**

Für die Positionierung der Welt wurden Sprachbefehle implementiert mit denen sich die Welt bewegen lässt. Mit dem Start der App befindet sich die Welt im Rotationsmodus (Sprachbefehl: "Rotate Mode"), wodurch sich die Welt mit der Tap-Geste um die Y-Achse rotieren lässt. Mit dem Sprachbefehl "Move Mode" lässt sich in den Bewegungsmodus gewechselt. In diesem lässt sich die Welt frei im Raum bewegen. Ist die gewünschte Rotation und Transformation eingestellt, so lässt sich mit dem "Lock Mode" Sprachbefehl die Welt feststellen. Die Hörversuche sollten im "Lock Mode" durchgeführt werden, damit der Proband nicht fälschlicherweise die Welt während des Hörversuchs bewegt. Die Sprachbefehle werden den Probanden nicht beigebracht und dienen ausschließlich zur Kontrolle der Welt.

## **4.7 Ablauf des Hörversuchs**

Zu aller erst wird die App auf der HoloLens im Versuchslabor gestartet. Wichtig dabei ist, dass beim Start der App der Träger der HoloLens sich an der durch das Raummodell festgelegten Stelle befindet. Nur dann stimmt das in Unity festgelegte Raummodell mit dem realen Versuchsräum überein, so dass ein plausibles akustisches Ergebnis durch Spatial Audio erzielt werden kann.

Im Anschluss wird das Hologramm der Welt, mit Hilfe der Sprachbefehle, auf dem dafür vorgesehenen Tisch platziert und festgestellt. Daraufhin kann die HoloLens an den ersten Probanden übergeben werden.

Zu Beginn des Hörversuchs stellt sich jeder Proband zunächst an die vorgesehene Stelle im Raum. Damit die Ergebnisse des Hörversuchs nicht durch eine variierende Wortwahl bei der Erklärung verfälscht werden, erhalten alle Probanden die gleiche Einweisung über ein Tutorial-Fenster, welches über die HoloLens dargeboten wird.



Abbildung 4.13: Einheitliches Tutorial für die Probanden zur Erklärung der Steuerung im Hörversuch

Sind nach der Einweisung noch Fragen offen, so kann der Proband diese vor dem Beginn des Hörversuchs klären. Mit dem Beenden des Tutorials sieht der Proband die Welt wie in Abbildung 4.12. Hier kann der Proband testen, ob er das Tutorial verstanden hat und sowohl die Balken der Fragen markieren, als auch die Kugel im Raum platzieren.

Hat der Proband alles verinnerlicht ist er bereit den Hörversuch über den "Versuch startenButton" zu starten. Daraufhin sollte das Geräusch der ersten Szene abgespielt werden. Die Aufgabe des Probanden ist es nun die rote Kugel an der Stelle zu platzieren, an der er die virtuelle Schallquelle hört und anschließend die Fragen über der Welt zu beantworten. Ist dies erledigt kann der Benutzer mit dem "Nächste SzeneButton" die nächste Szene laden. Dieser Vorgang wiederholt sich für alle fünf Szenen. Mit dem Beenden der fünften Szene wird dem Probanden über ein Pop-Up-Fenster signalisiert, dass der Versuch abgeschlossen ist.

Die Daten des Probanden werden für jede Szene einzeln, lokal auf der HoloLens in einer Text-Datei gespeichert. Hierbei werden die zur Welt relativen Koordinaten und die Bewertungen als Zahlen von -5 bis 5 gespeichert.

## **4.8 Probanden**

Bei den für den Hörversuch ausgewählten Probanden handelt es sich hauptsächlich um Studierende der Hochschule Emden/Leer. Der Hörversuch wurde hierbei mit 15 unterschiedlichen Probanden durchgeführt, wobei leider einige aus Zeitgründen nicht alle Szenen des Versuchs durchführen konnten.

Das Alter der Probanden lag hierbei zwischen 20 und 50 Jahren, wobei die meisten zwischen 20 und 30 Jahren alt waren. Anzumerken ist hierbei, dass die abgespielten Geräusch aller eher niederfrequent waren und daher eine Beeinflussung der Ergebnisse durch das Alter in der Auswertung vernachlässigt wird.

Zwölf der getesteten Personen waren männlich und drei weiblich . Einige Probanden trugen unter der HoloLens eine Brille, dies soll aber bei der Auswertung trotz des möglichen Einflusses auf die Ergebnisse nicht weiter berücksichtigt werden.

Der Hörversuch erfolgte anonym und ohne Erfassung persönlicher Daten. Die Auswertung der Probandendaten erfolgt ohne Hintergrundinformationen über die Probanden und basiert ausschließlich auf den durch die App gespeicherten Daten. Für sicherere Versuchsergebnisse sollte der Hörversuch mit einer größeren Menge an Probanden, sowie einer Erfassung der persönlichen Probandendaten, wiederholt werden.

# 5 Versuchsergebnisse

Im Rahmen dieser Bachelorarbeit wurde ein Hörversuch mit binauralem, räumlichem Audio durchgeführt. Ziel des Hörversuches ist es herauszufinden inwiefern die natürliche Lokalisationsfähigkeit von Hörereignissen in einer Mixed Reality Umgebung gegeben ist. Bei der Durchführung des Hörversuches haben die Probanden für fünf verschiedene Audio-Szenen jeweils versucht die virtuelle Schallquelle zu lokalisieren. Zusätzlich haben sie jeweils Fragen zum Darbietungsempfinden anhand bestimmter Kriterien beantwortet. Die Ergebnisse des durchgeföhrten Hörversuches sollen im Folgenden analysiert werden.

Während des Versuchs hatten die Probanden für jede Szene die Aufgabe eine rote Kugel genau dort zu platzieren, wo sie die virtuelle Schallquelle anhand des Gehörten vermuteten. Die Position der roten Kugel wurde hierbei für jeden Probanden, für jede Szene dokumentiert. Es gilt nun die Daten die gesammelten Daten zu analysieren.

Zur Visualisierung der Probandenergebnisse wird die virtuelle Welt zusammen mit der wirklichen, virtuellen Schallquelle und den von den Probanden lokalisierten, virtuellen Schallquellen dargestellt. Die echte, virtuelle Schallquelle wird, sofern es eine gibt, durch eine rosa Kugel und die von den Probanden lokalisierte virtuelle Schallquelle durch kleinere, rote Kugeln dargestellt. Hierbei ist anzumerken, dass die Platzierung der roten Kugel aufgrund fehlender Übung mit dem Umgang der HoloLens nicht komplett genau erfolgen konnte. Es sollte daher bei der Analyse berücksichtigt werden, dass die Probanden die Kugel vielleicht nicht genau da platzieren konnten, wo sie die Schallquelle ganz genau gehört haben.

## 5.0.1 Szene 1: Adlerschrei im Adlernest

In der ersten Szene wurden die in den Microsoft-Tutorials für Spatial Audio empfohlenen Einstellungen für räumliches Audio direkt übernommen. Die virtuelle Schallquelle mit dem Geschrei eines Adlers wurde für die Szene direkt in das Adlernest gelegt. Anhand der Abbildung 5.1 lässt sich erkennen, dass die Probanden die virtuelle Schallquellen fast ausnahmslos auf der Oberfläche des Berges neben dem fliegenden Adler unmittelbar über dem Berg lokalisiert haben. Es ist anzunehmen, dass die Probanden versucht haben die Kugel genau auf dem Adler zu platzieren und nicht im Nest, wie vorgesehen. Einige Schallquellen wurden außerdem weiter rechts im Gebirge und auch einige über dem Gebirge lokalisiert. In Abbildung 5.2 ist zu sehen, dass ein Proband die Schallquelle sogar weit über sich beim dritten Adler lokalisierten konnte.

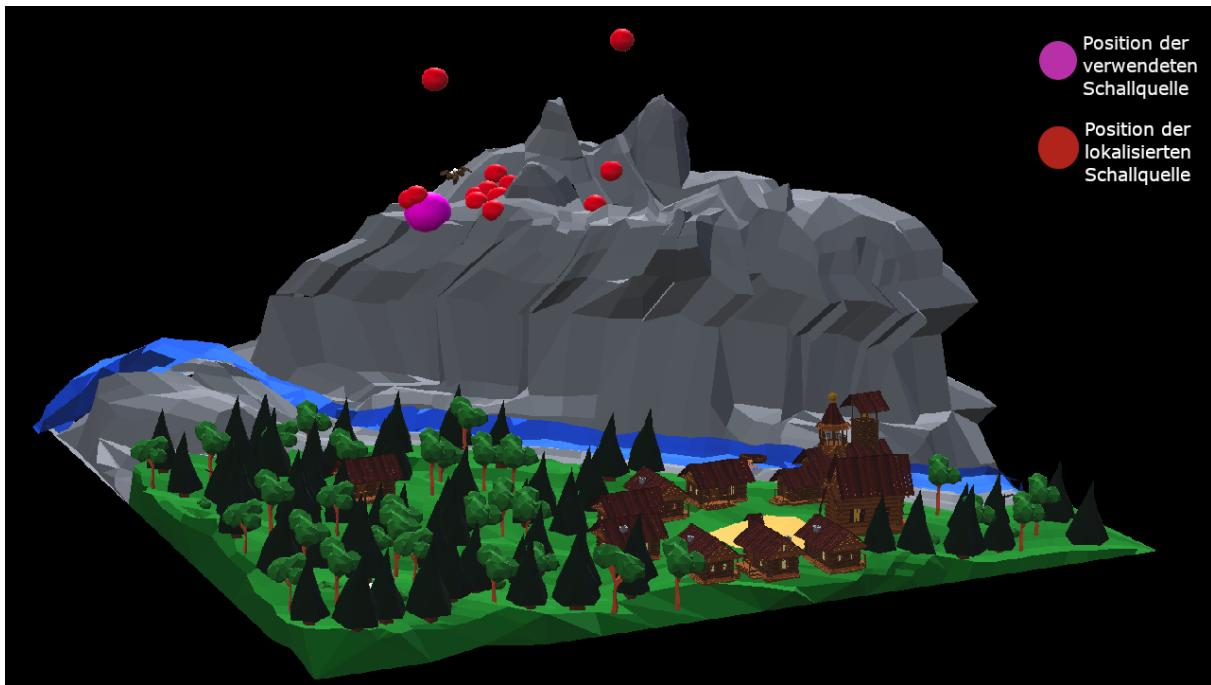


Abbildung 5.1: Visualisierung der lokalisierten, virtuellen Schallquellen für Szene 1

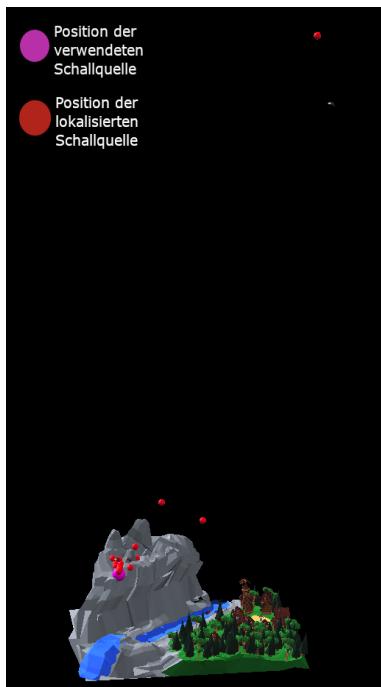


Abbildung 5.2: Visualisierung der lokalisierten, virtuellen Schallquellen für Szene 1 aus der Seitensicht

Anhand der Lokalisationen lässt sich erschließen, dass die Probanden zum größten Teil die ungefähre Position der virtuellen Schallquelle hören konnten. Das Wissen, dass in Wirklichkeit die Schallquelle direkt auf dem Adler liegen müsste ließ die meisten Probanden die virtuelle Schallquelle jedoch direkt auf dem Adler vermuten. Anzumerken ist jedoch, dass auch einige die Schallquelle genau richtig im Nest platziert haben, obwohl dies nicht direkt zum inhaltlichen Kontext passt.

Es ist ebenfalls festzuhalten, dass bis auf drei Probanden alle die virtuelle Schallquelle in der gleichen Höhe lokalisiert haben. Die Verteilung in der Horizontalebene ist hingegen breiter verteilt.

Die Platzierung der Schallquelle über dem Probanden könnte ein Indiz dafür sein, dass das Visuelle einen höheren Einfluss auf die Lokalisation der Höreignisse hat als gedacht. Obwohl der Schall eigentlich nicht so klingen sollte, als ob er von oben käme hat der Proband die Schallquelle trotzdem dort lokalisiert. Jedoch sollte berücksichtigt werden, dass es sich hierbei um einen einzelnen Ausreißer handelt .

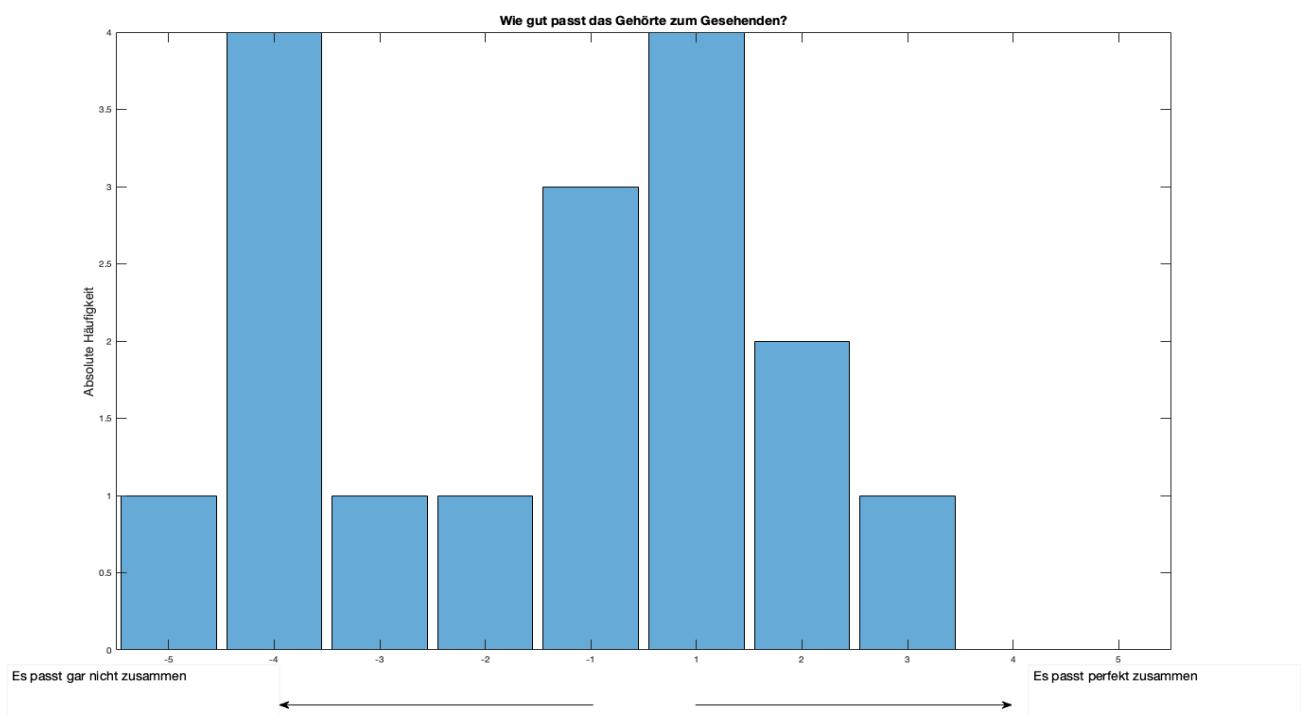


Abbildung 5.3: Histogramm der ersten Frage des Versuches für Szene 1

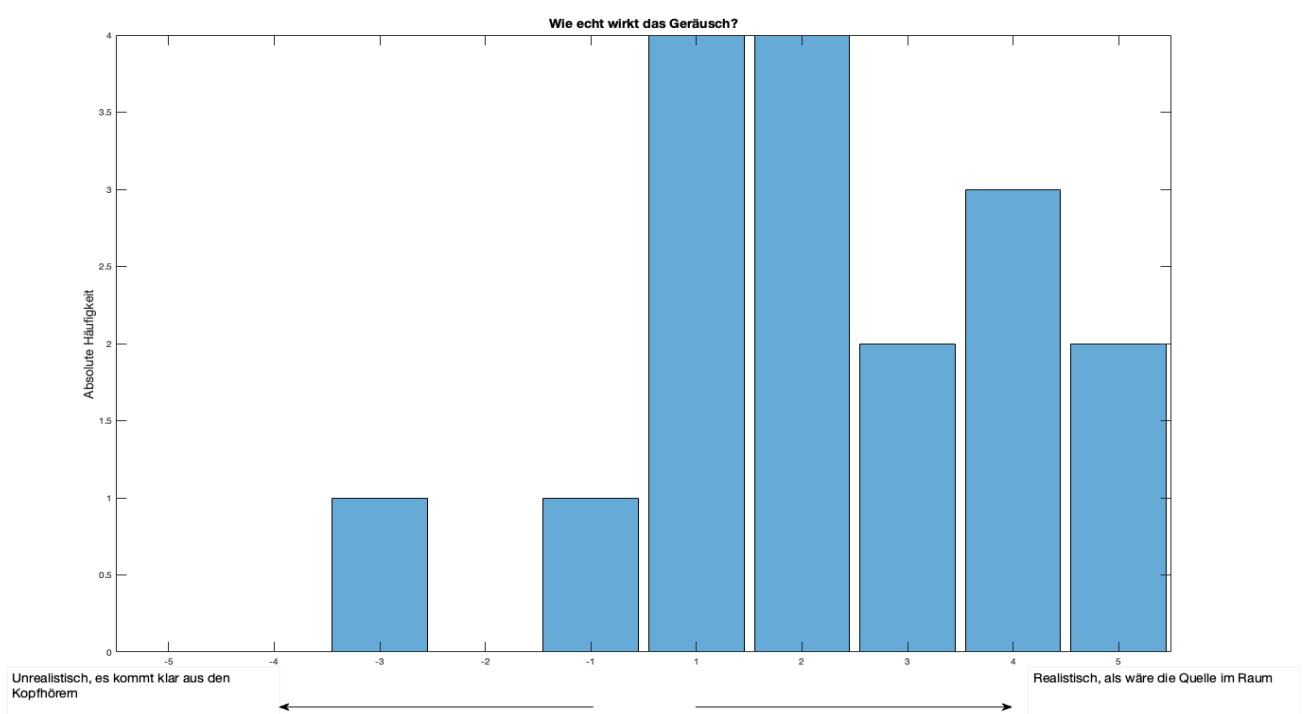


Abbildung 5.4: Histogramm der zweiten Frage des Versuches für Szene 1

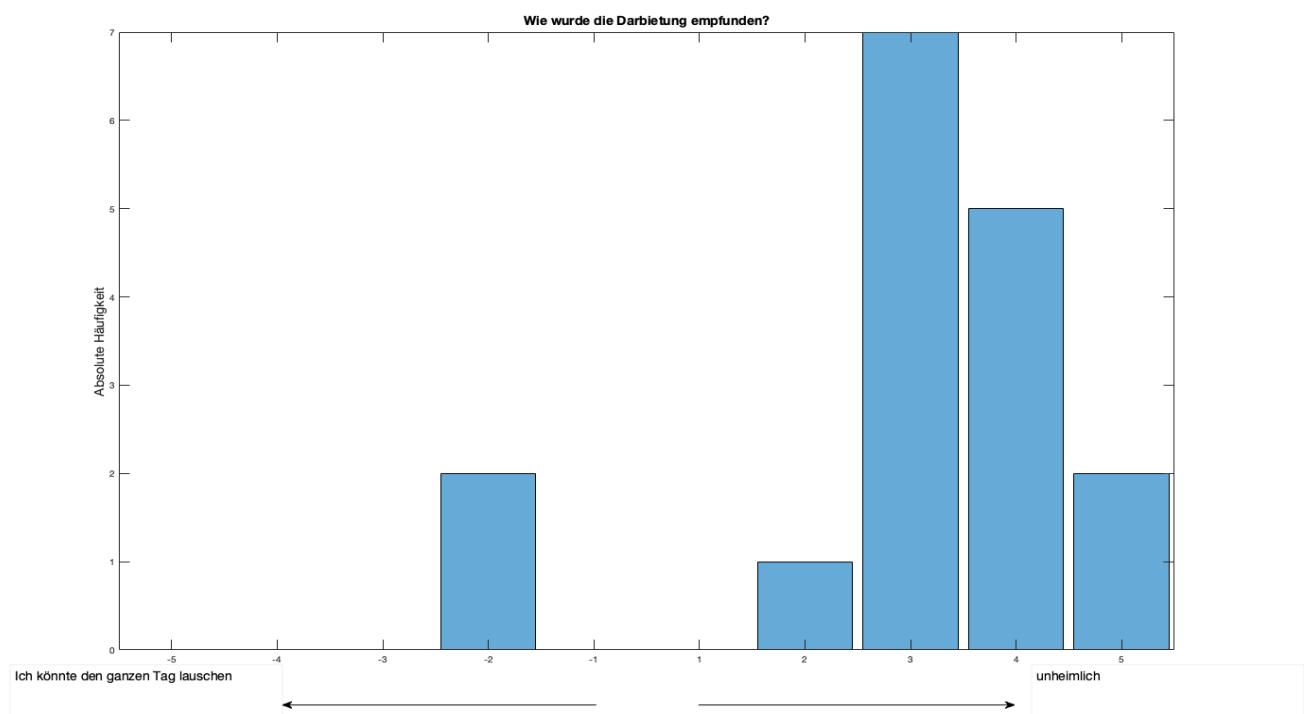


Abbildung 5.5: Histogramm der dritten Frage des Versuches für Szene 1

SZENE 1 ENDE

## 5.0.2 Szene 2: Wolfsheulen

SZENE 2 ANFANG

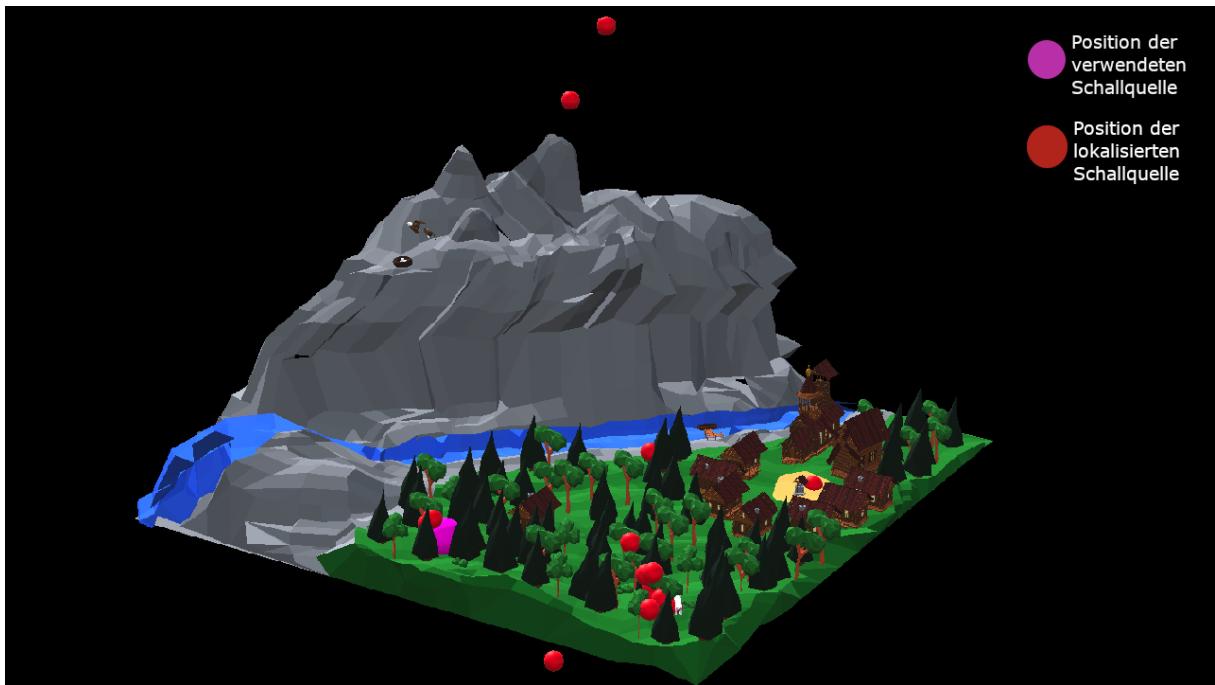


Abbildung 5.6: Visualisierung der lokalisierten Schallquellen für Szene 2

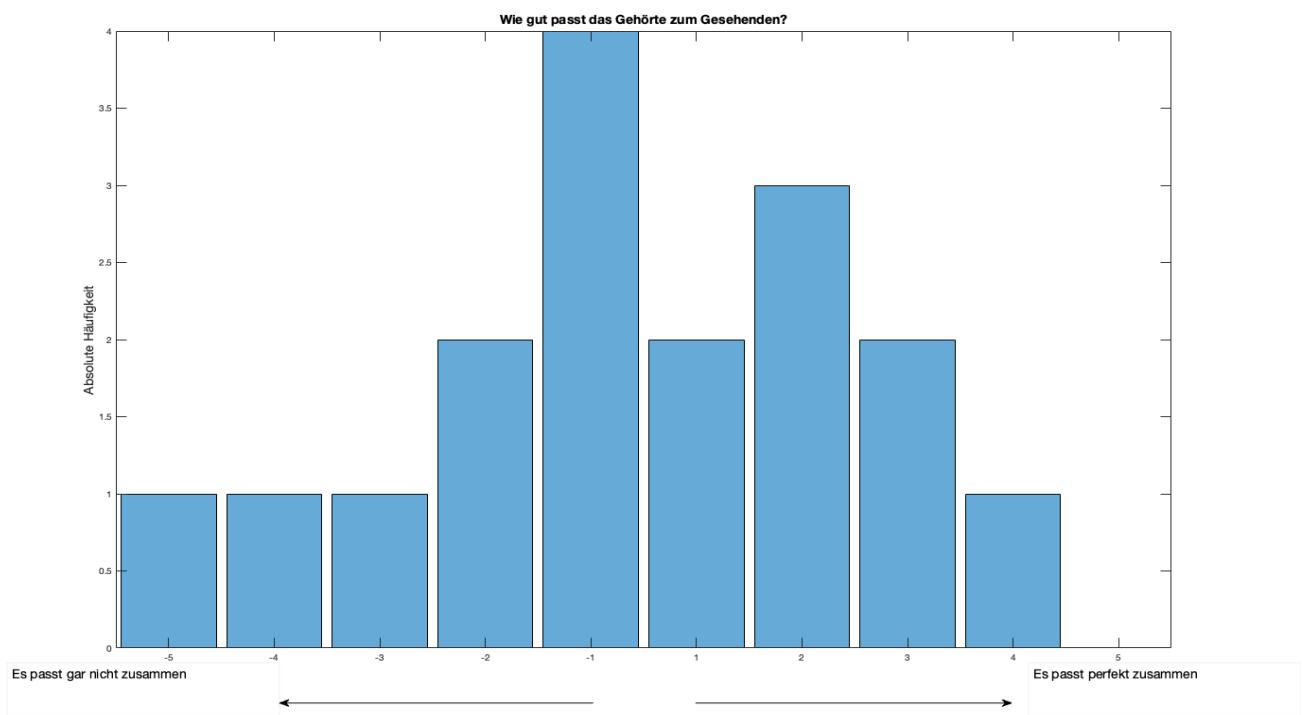


Abbildung 5.7: Histogramm der ersten Frage des Versuches für Szene 2

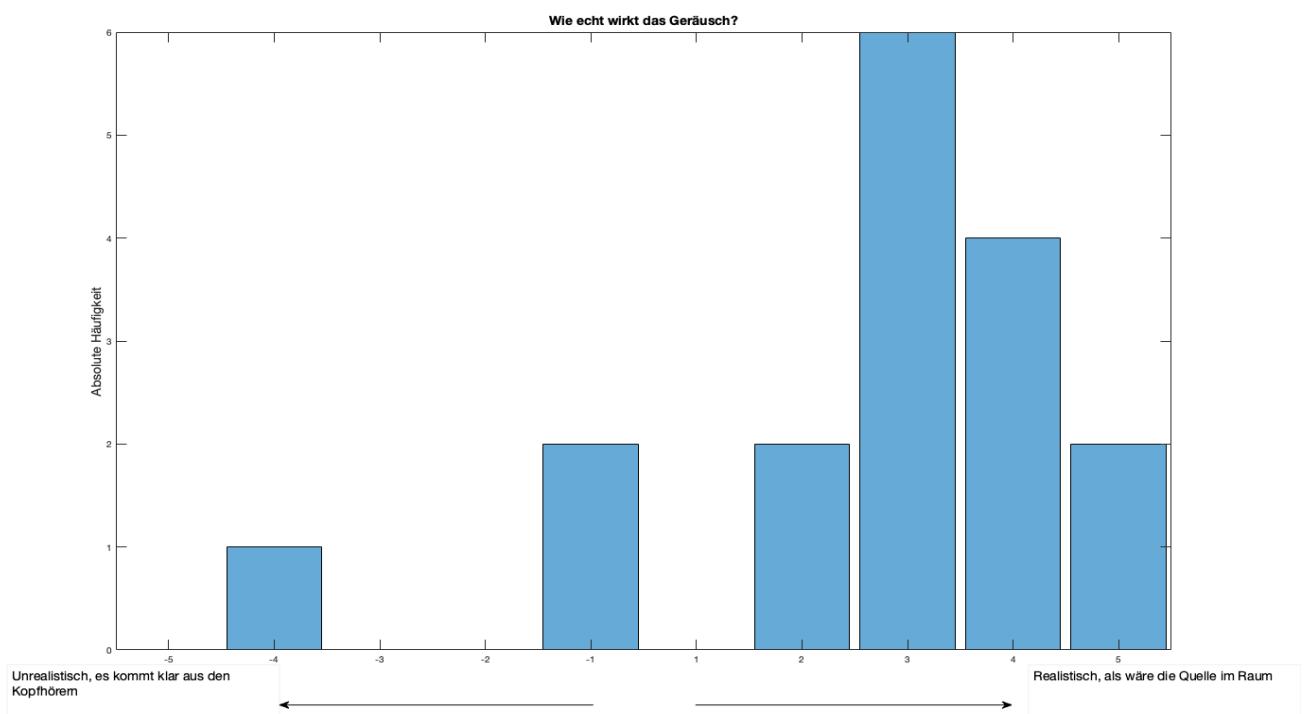


Abbildung 5.8: Histogramm der zweiten Frage des Versuches für Szene 2

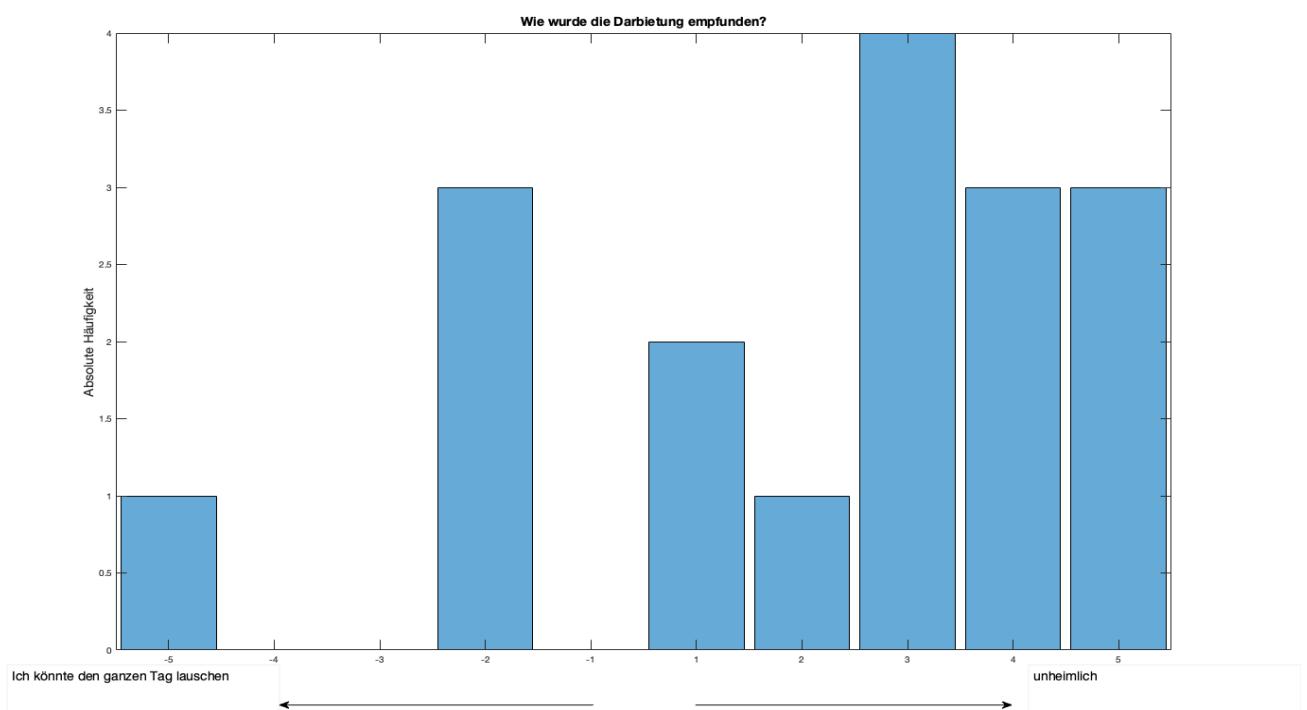


Abbildung 5.9: Histogramm der dritten Frage des Versuches für Szene 2

SZENE 2 ENDE

### 5.0.3 Szene 3: Kirchenglocken

SZENE 3 ANFANG

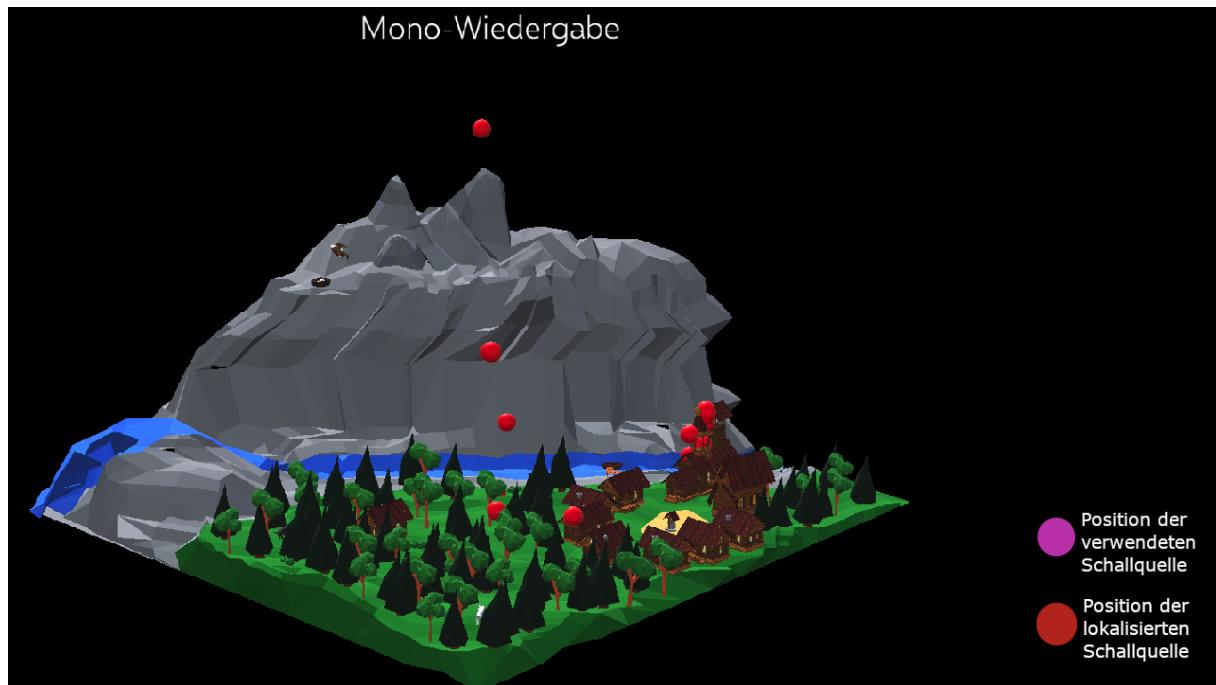


Abbildung 5.10: Visualisierung der lokalisierten Schallquellen für Szene 3

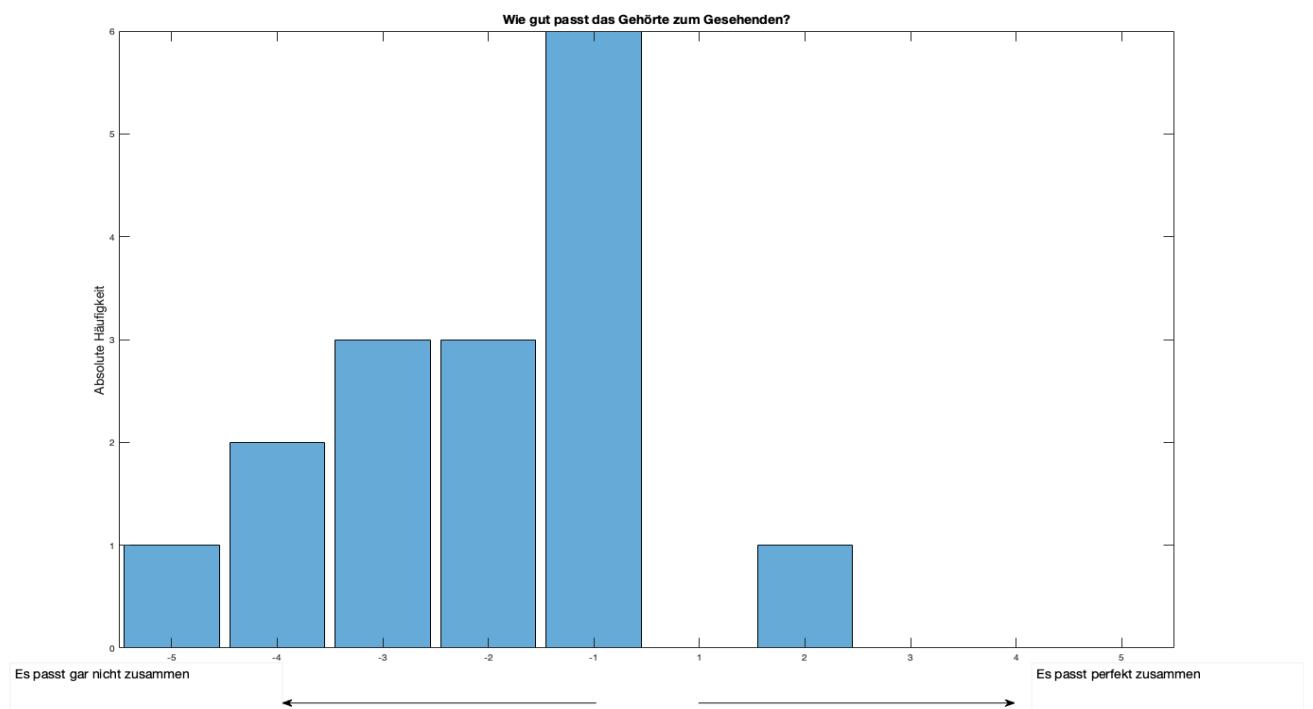


Abbildung 5.11: Histogramm der ersten Frage des Versuches für Szene 3

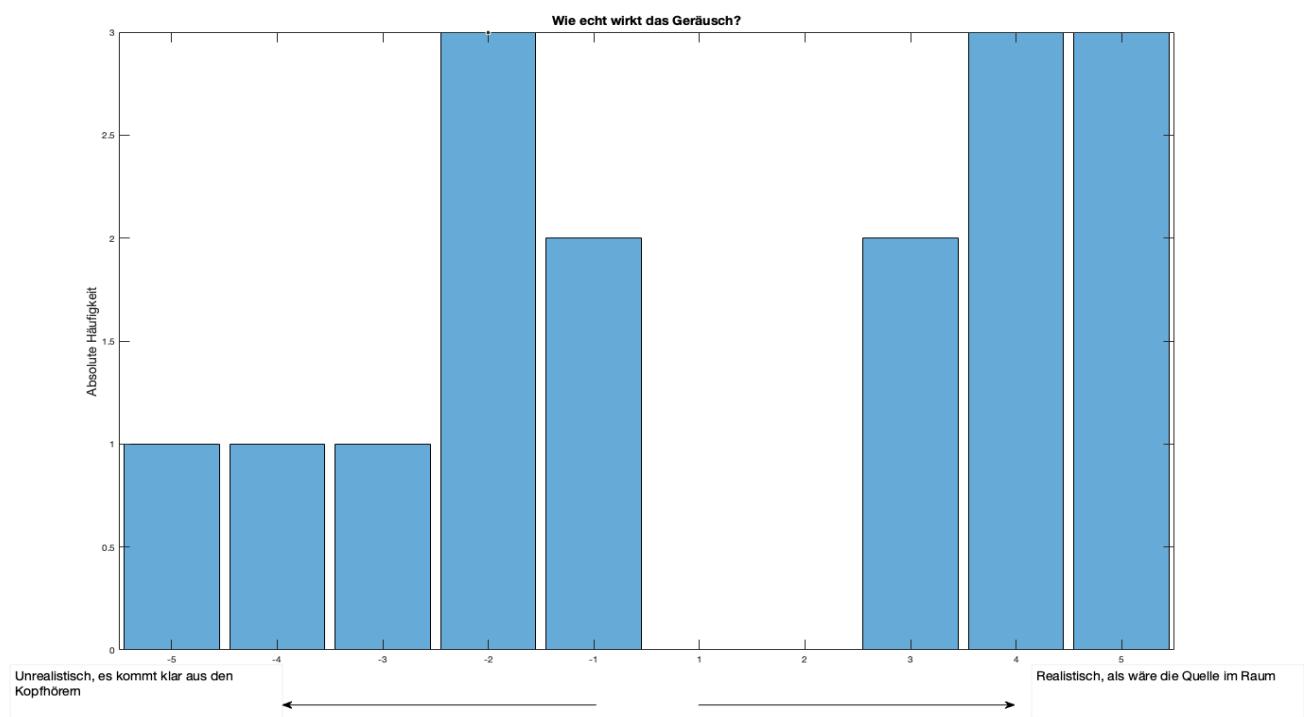


Abbildung 5.12: Histogramm der zweiten Frage des Versuches für Szene 3

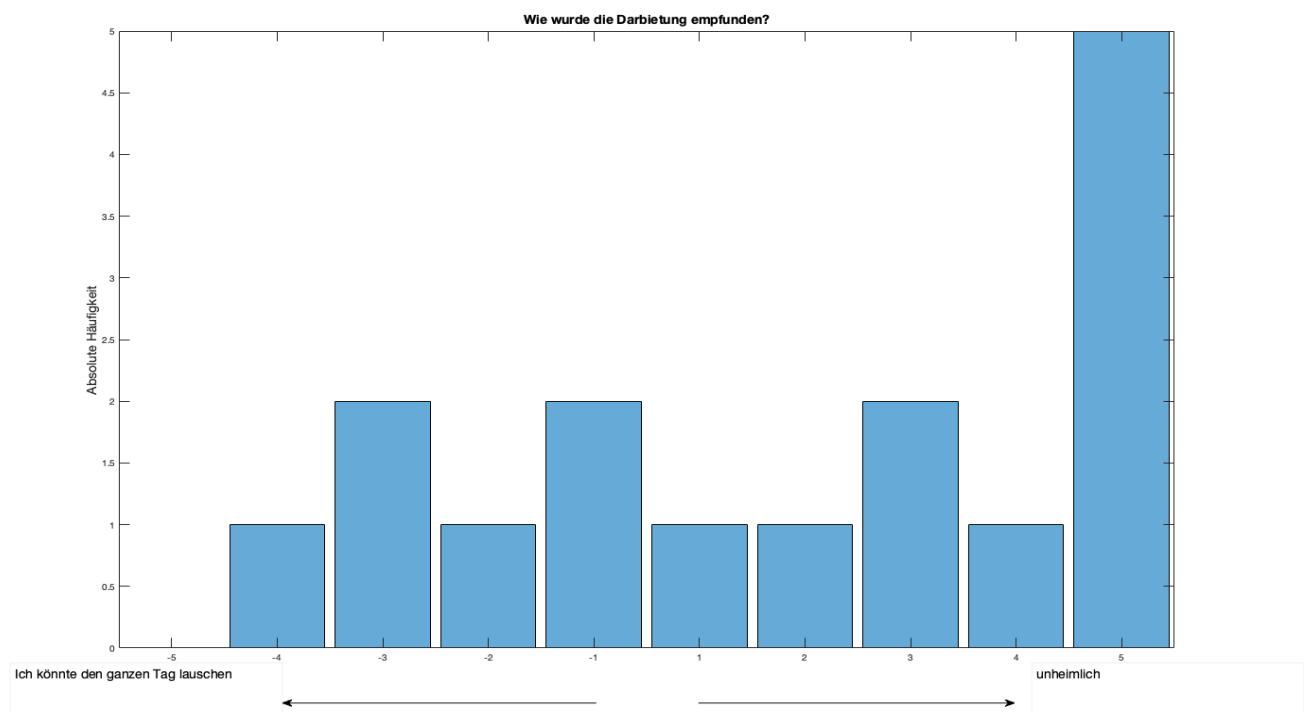


Abbildung 5.13: Histogramm der dritten Frage des Versuches für Szene 3

SZENE 3 ENDE

## 5.0.4 Szene 4: Adler in der Luft

SZENE 4 ANFANG

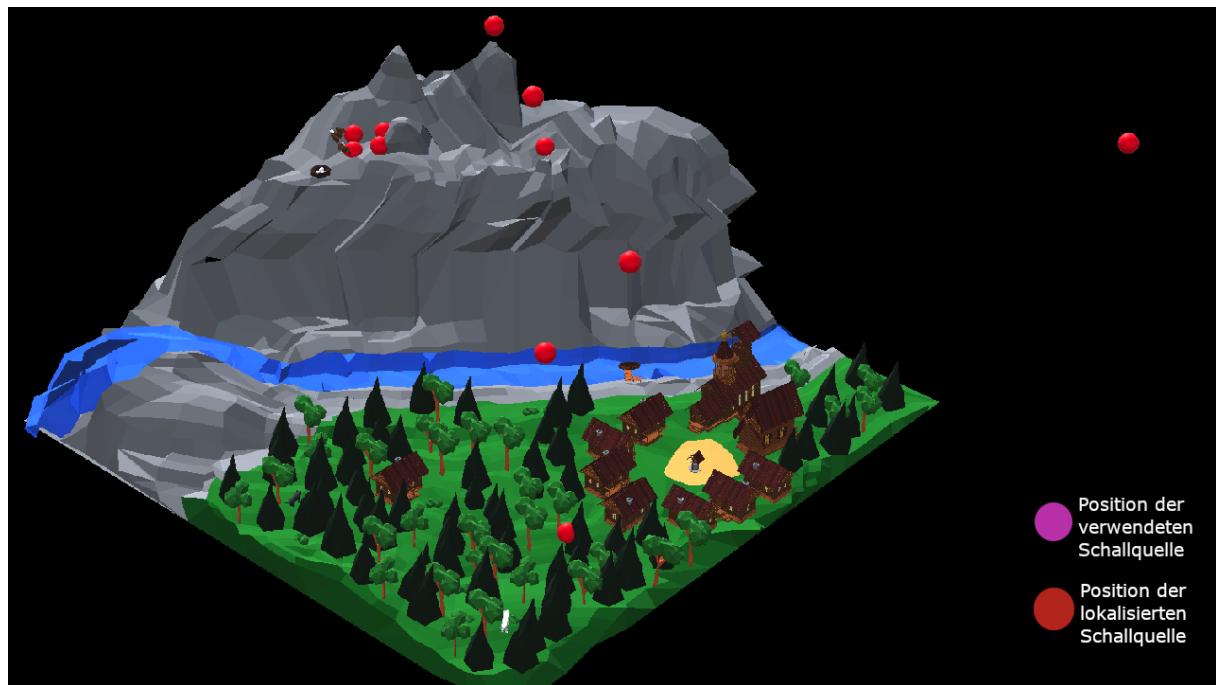


Abbildung 5.14: Visualisierung der lokalisierten Schallquellen für Szene 4

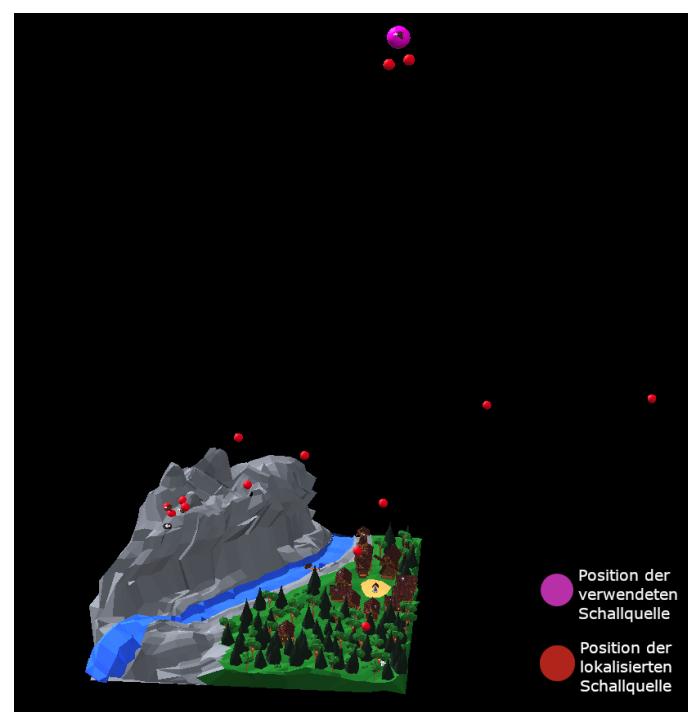


Abbildung 5.15: Visualisierung der lokalisierten Schallquellen für Szene 4

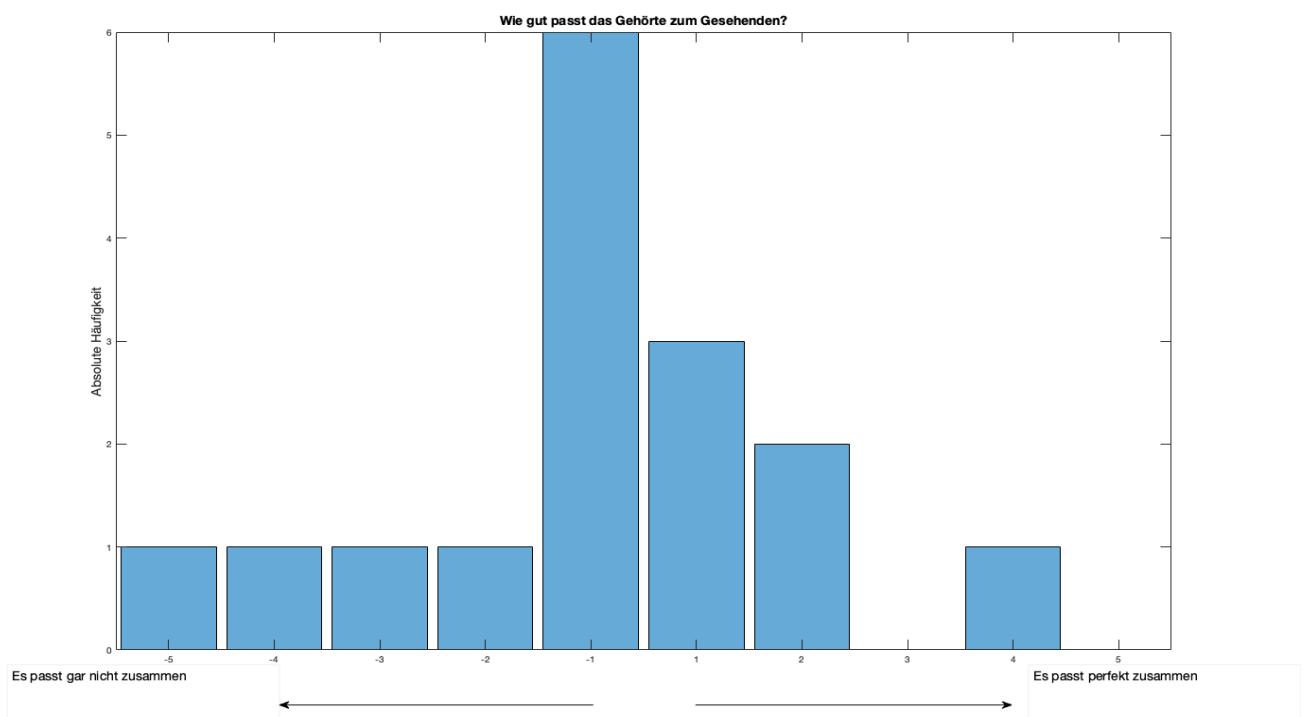


Abbildung 5.16: Histogramm der ersten Frage des Versuches für Szene 4

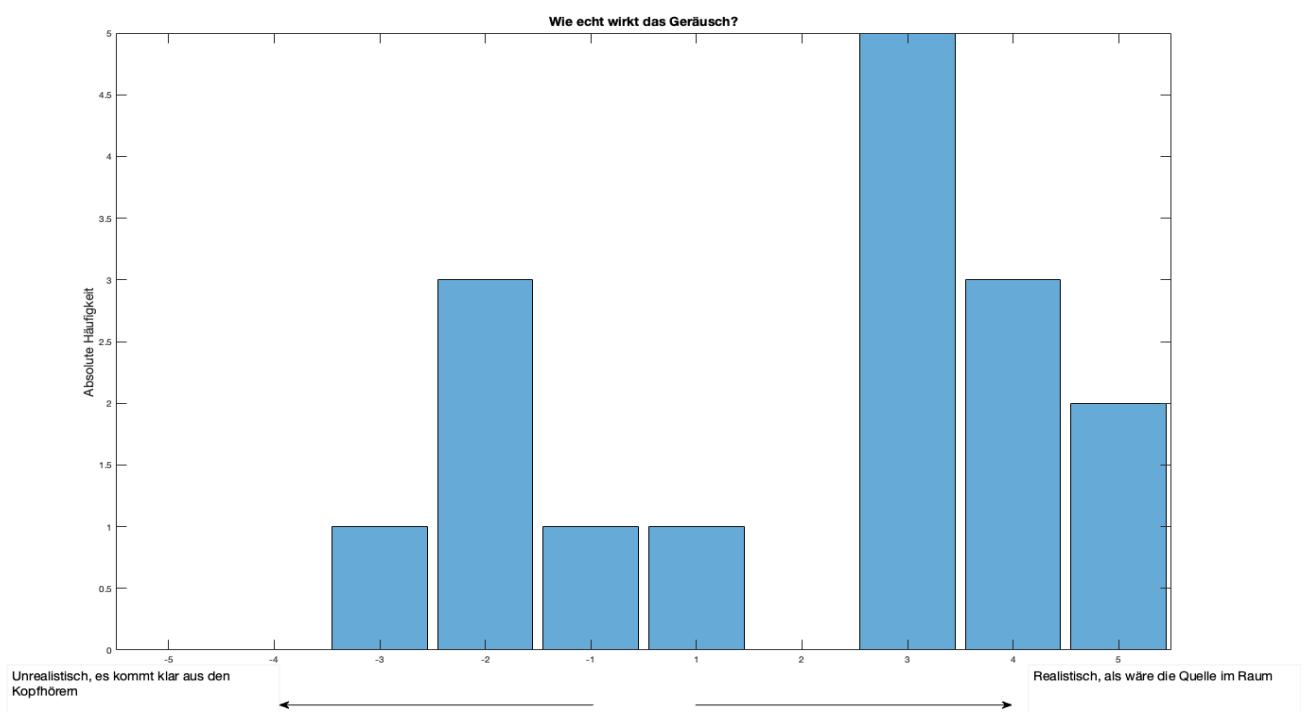


Abbildung 5.17: Histogramm der zweiten Frage des Versuches für Szene 4

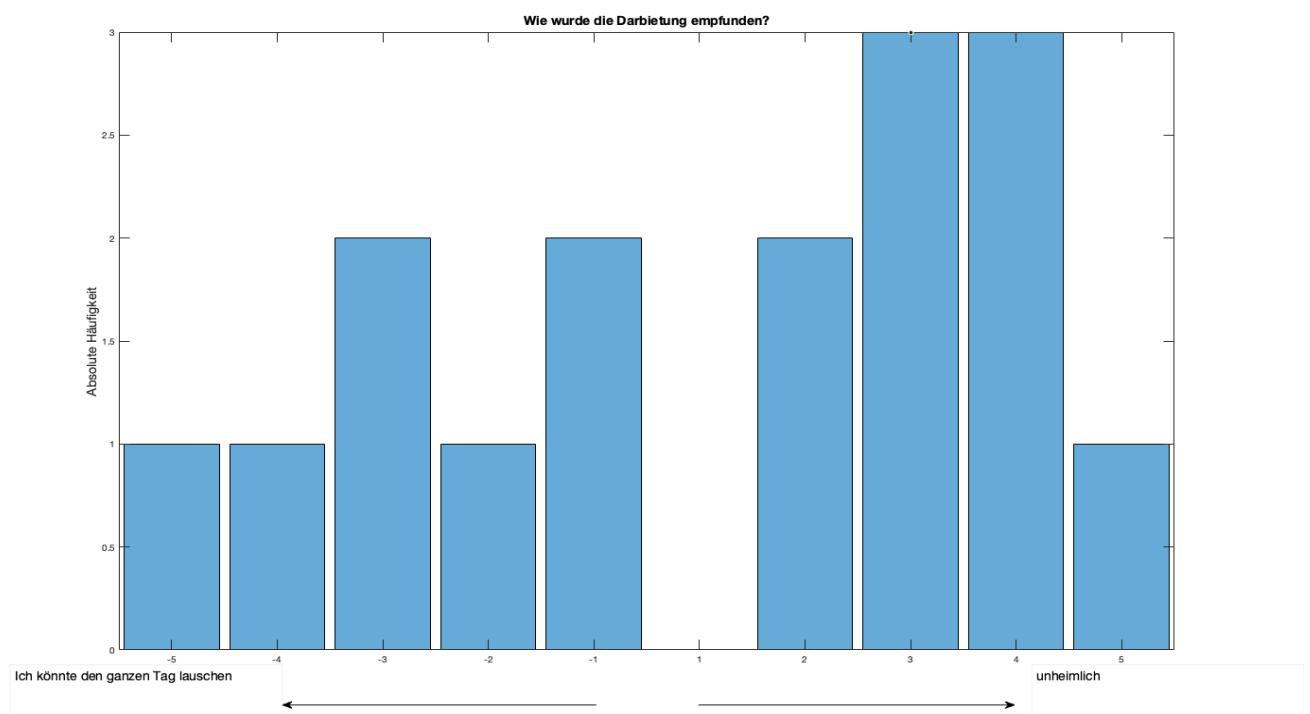


Abbildung 5.18: Histogramm der dritten Frage des Versuches für Szene 4

SZENE 4 ENDE

## 5.0.5 Szene 5: Flussrauschen

SZENE 5 ANFANG

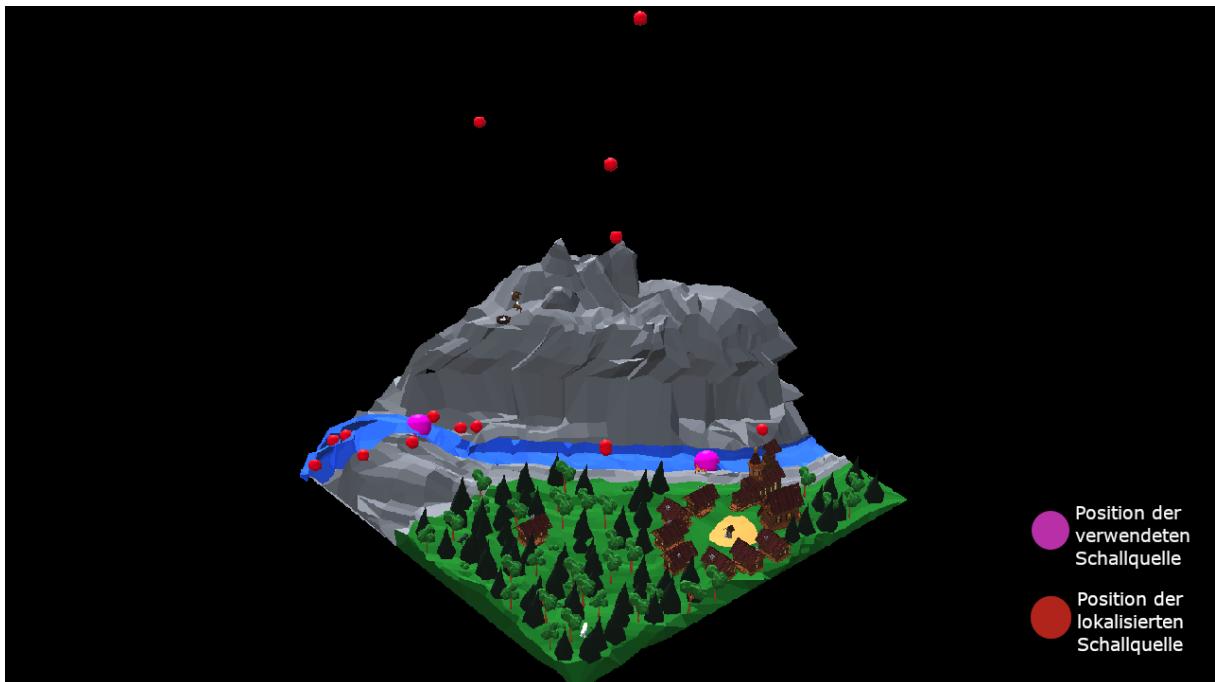


Abbildung 5.19: Visualisierung der lokalisierten Schallquellen für Szene 5

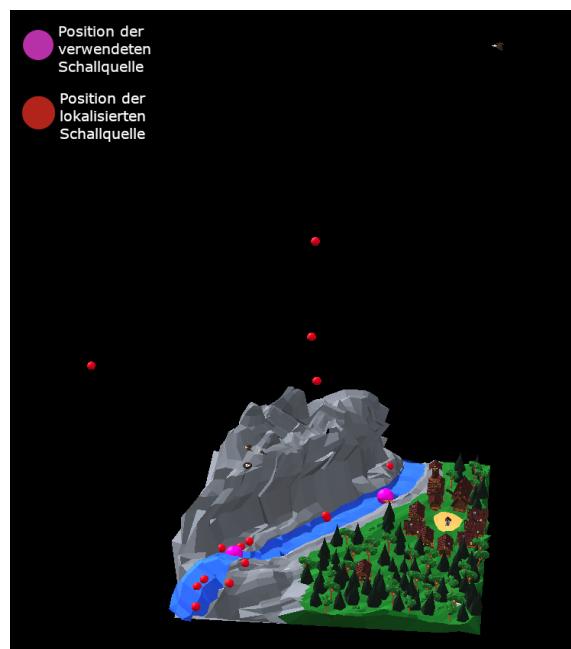


Abbildung 5.20: Visualisierung der lokalisierten Schallquellen für Szene 5

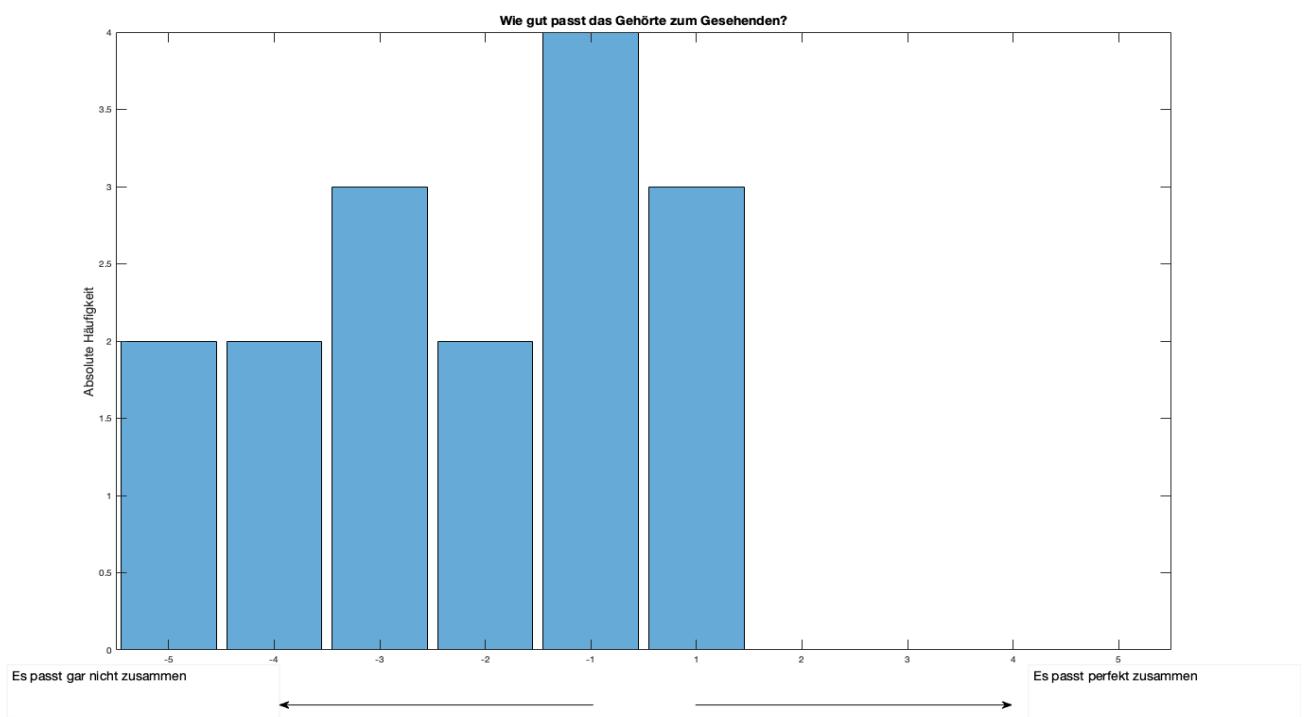


Abbildung 5.21: Histogramm der ersten Frage des Versuches für Szene 5

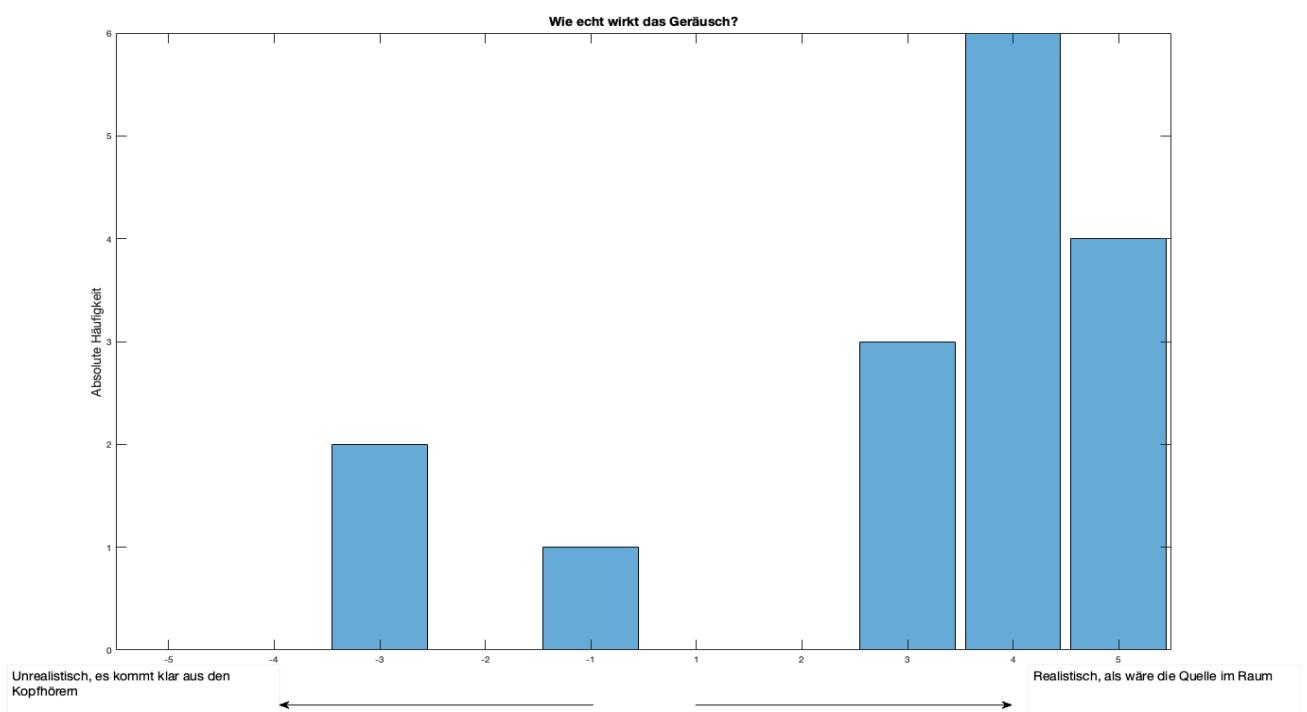


Abbildung 5.22: Histogramm der zweiten Frage des Versuches für Szene 5

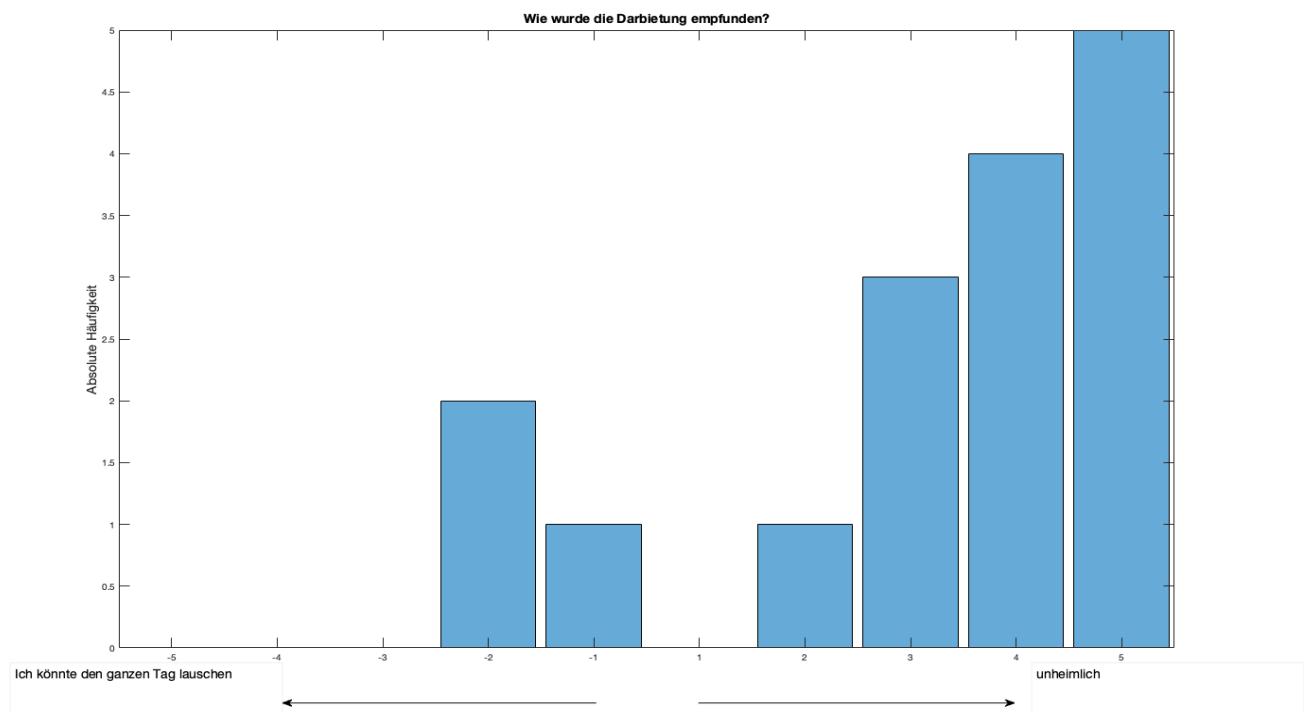


Abbildung 5.23: Histogramm der dritten Frage des Versuches für Szene 5

SZENE 5 ENDE

## 5.1 Ergebnisse kritisch hinterfragen

# Literaturverzeichnis

- [1] R. T. Azuma. A surey of augmented reality. *Hughes Research Laboratories*, 1997.
- [2] J. Blauert. Neuere ergebnisse zum räumlichen hören. *Westdeutscher Verlag GmbH Opladen*, 1980.
- [3] B. Bray, J. McCulloch, N. Schonning, and M. Zeller. What is mixed reality? Microsoft, <https://docs.microsoft.com/en-us/windows/mixed-reality/mixed-reality>, Mar. 2018. abgerufen am 07.02.19.
- [4] P. D.-I. K. Genuit. *Sound-Engineering im Automobilbereich*. Springer-Verlag Berlin Heidelberg, 2010.
- [5] H. Gieselmann. Gdc: Vr-entwickler sony und oculus warnen vor der "uncanny valley". Heise Online, <https://www.heise.de/newsticker/meldung/GDC-VR-Entwickler-Sony-und-Oculus-warnten-vor-der-Uncanny-VR-alley-2291230.html>, Aug. 2014. abgerufen am 05.02.19.
- [6] O. Grau. *From illusion to immersion*. Massachusetts Institute of Technology, 2003.
- [7] P. D.-I. K. Heidermanns. *Elektroakustik*. B. G. Teubner Studienskripten Stuttgart, 1979.
- [8] D. Johnston and H. Gamper. Spatial audio. <https://www.microsoft.com/en-us/research/project/spatial-audio/>, Dec. 2015. abgerufen am 11.04.19.
- [9] keveleigh. Mr spatial 220: Spatial sound. <https://docs.microsoft.com/en-us/windows/mixed-reality/holograms-220>, Mar. 2018. abgerufen am 12.04.19.
- [10] K. F. MacDorman and N. Kageki. The uncanny valley. *IEEE Robotics & Automation Magazine*, 2012.
- [11] P. L. Strick. The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing. *University of Pittsburgh, PNAS*, 2017.
- [12] S. Weinzierl. *Handbuch der Audiotechnik*. Springer-Verlag Berlin Heidelberg, 2008.