

BACHELORARBEIT

Entwicklung eines Produktionsverfahrens für Audio Augmented Reality in einer urbanen Umgebung

vorgelegt am 22. April 2025
Gunther Greim

Erstprüferin: Prof. Dr. Eva Wilk
Zweitprüfer: Matthias Fehling

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**

Department Medientechnik
Finkenau 35
20081 Hamburg

Zusammenfassung

In der vorliegenden Arbeit wird ein Produktionsverfahren für Audio Augmented Reality in einer urbanen Umgebung entwickelt. Mit dem Verfahren kann eine Schallquelle aus dem Innenraum eines Gebäudes mit Hilfe von Spatial Audio draußen vor dem Gebäude virtuell reproduziert werden. Dabei wird die Schallquelle anhand von Reflexionen in die umgebende Akustik eingefügt. Dieses Verfahren wurde entwickelt, um eine bestehende Lücke in der Literatur zu schließen. Trotz umfangreicher Forschung auf diesem Gebiet gibt es bisher keine Ansätze, die dieses Konzept behandeln. Der Schwerpunkt liegt dabei auf der Entwicklung von Prozessen zur Aufzeichnung und Wiedergabe der Schallquelle und der Implementierung von Technologien zur Erzeugung von Spatial Audio. Das Ziel der Forschung ist es, die virtuelle Schallquelle in hoher Immersion zu simulieren. Dazu wurden verschiedene Ansätze entwickelt und in der praktischen Anwendung getestet. Der Grad der Immersion wurde durch einen Hörversuch verifiziert. Die Untersuchung hat gezeigt, dass die Stabilität der Schallquelle von herausragender Bedeutung für die wahrgenommene Immersion ist und sich auf andere räumliche Attribute direkt auswirkt.

.

Abstract

In this thesis, a production process for audio augmented reality in an urban environment is developed. With this method, a sound source from the interior of a building can be virtually reproduced outside the building with the help of spatial audio. The sound source is inserted into the surrounding acoustics using reflections. This method was developed to close an existing gap in the literature. Despite extensive research in this field, there are no approaches that adequately address this concept. The focus is on the development of processes for recording and reproducing the sound source and the implementation of technologies for generating spatial audio. The aim of the research is to simulate the virtual sound source in high immersion. To this end, various approaches were developed and tested in practical applications. The degree of immersion was verified by a listening test. The investigation has shown that the stability of the sound source is of paramount importance for the perceived immersion and has a direct impact on other spatial attributes.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich bei der Anfertigung dieser Arbeit unterstützt haben. Mein besonderer Dank gilt Frau Prof. Dr. Eva Wilk, die diese Arbeit betreut hat, für die hilfreichen Anregungen und die konstruktive Kritik. Ich bedanke mich bei Matthias Fehling, der die Zweitkorrektur übernommen hat und die Tontechnik bereitgestellt hat. Mein besonderer Dank gilt Ben Supper von Supperware, der mir den Headtracker zur Verfügung gestellt hat. Weiterhin bedanke ich mich beim Team der Zinnschmelze Barmbek für die Unterstützung und bei den Mitarbeitern des Museums der Arbeit, die die Aufnahmen der Setzmaschine ermöglicht haben. Ich bedanke mich bei Felix Holzmüller vom IEM Graz für die Unterstützung bei der Simulationssoftware. Ein besonderer Dank gilt den Teilnehmern des Hörversuchs, ohne die diese Arbeit nicht hätte entstehen können. Außerdem bedanke ich mich bei meinen Eltern für die fortwährende Unterstützung während des Studiums.

Inhaltsverzeichnis

Zusammenfassung	ii
Danksagung	iii
Abkürzungsverzeichnis	VI
Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
1 Einleitung	1
2 Theorie	2
2.1 Kriterien	2
2.2 Begriffserklärung	2
2.3 Forschungsstand	3
2.4 Szenenbasierte Produktion in Ambisonics mit binauraler Dekodierung	5
3 Planung	6
3.1 Verfahren A: Software gestützte Raumsimulation	7
3.2 Verfahren B: Originalakustische Beschallungsaufnahme am Hörereignisort	9
3.3 Verfahren C: Aufnahme mit Soundfield-Mikrofon am Ursprungsort	12
3.4 Ressourcenplanung	13
4 Design des Hörversuchs	14
5 Realisierung	17
5.1 Aufnahme der Monotype im Museum der Arbeit	17
5.2 Beschallung und Aufnahme auf dem Campus Finkenau	18
5.3 Modellierung des simulierten Raums	20
5.4 Durchführung der Hörversuchs	21
6 Ergebnisse	22
6.1 Lokalisierung	23
6.2 Distanzwahrnehmung	26
6.3 Ausdehnung, Stabilität und Plausibilität	28
7 Auswertung	30
7.1 Lokalisierung	31

7.2	Distanzwahrnehmung.....	31
7.3	Ausdehnung, Stabilität und Plausibilität.....	32
7.4	Übersicht aller Testergebnisse	32
7.5	Gewichteter Durchschnitt	33
8	Fehleranalyse	34
9	Zusammenfassung	34
	Literaturverzeichnis	36
	Anhang	38
	Anhang A: Fragebogen Hörversuch.....	39
	Anhang B: Instruktion für den Hörversuch.....	42
	Anhang C: Rohdaten Hörversuch.....	44
	Anhang C1: Ergebnisse Verfahren A.....	44
	Anhang C2: Ergebnisse Verfahren B.....	45
	Eigenständigkeitserklärung	46

Abkürzungsverzeichnis

AAR	Audio Augmented Reality
BRIR	Binaural Room Impulse Response
DAW	Digital Audio Workstation
EQ	Equalizer
HRIR	Head-Related Impulse Response
IKL	Im-Kopf-Lokalisierung
IQA	Interquartilabstand
OSC	Open Sound Control
PC	Personal Computer
PCM	Pulse Code Modulation
Q	Quartil
HMD	Head Mounted Display
HRTF	Head-Related Transfer Function
SNR	Signal-to-Noise Ratio

Abbildungsverzeichnis

Abbildung 1: Signalfluss Verfahren A Quelle: Eigene Darstellung.....	10
Abbildung 2: Blockschaltbild vom Aufbau.....	10
Abbildung 3: Signalfluss Verfahren B Quelle: Eigene Darstellung.....	11
Abbildung 4: Signalfluss Verfahren C Quelle: Eigene Darstellung.....	12
Abbildung 5: Campus Finkenau mit Hörereignisort (gelb gestrichelt)	19
Abbildung 6: Aufbau der Beschallungsanlage	20
Abbildung 7: Hörereignisort mit Bezugspunkt (grün) und Schallquelle (orange)	21
Abbildung 8: entsprechende Darstellung im Room Encoder	21
Abbildung 9: Verteilung der Abweichung der wahrgenommenen Lokalisierung.....	24
Abbildung 10: Verfahren A: Prozentuale Verteilung der Abweichungen.....	25
Abbildung 11: Verfahren B: Prozentuale Verteilung der Abweichungen.....	25
Abbildung 12: Verfahren A: Verteilung der wahrgenommenen Distanz.....	27
Abbildung 13: Verfahren B: Verteilung der wahrgenommenen Distanz	28
Abbildung 14: Verfahren A: Häufigkeit der Bewertungen	29
Abbildung 15: Verfahren B: Häufigkeit der Bewertungen	30

Tabellenverzeichnis

Tabelle 1: Ressourcen der 3 Produktionsverfahren	13
Tabelle 2: Positionen der virtuellen Schallquelle/Lautsprecherpositionen	16
Tabelle 3: Maßstab für die Auswertung des semantischen Differentials	16
Tabelle 4: Maßstab für die Auswertung des gewichteten Durchschnitts	17
Tabelle 5: Abweichungen von der tatsächlichen Position in °	23
Tabelle 6: Streuungsmaße in °	23
Tabelle 7: Verfahren A: Distanzen in m	26
Tabelle 8: Verfahren A: Streuungsmaße in m	26
Tabelle 9: Verfahren B: Distanzen in m	27
Tabelle 10: Verfahren B: Streuungsmaße in m	27
Tabelle 11: Verfahren A: Häufigkeit der Bewertungen	28
Tabelle 12: Verfahren B: Häufigkeit der Bewertungen	29
Tabelle 13: Übersicht der Testergebnisse	32
Tabelle 14: Verfahren A: Testergebnisse und Gewichte	33
Tabelle 15: Verfahren B: Testergebnisse und Gewichte	33

1 Einleitung

In dieser Arbeit wird ein Produktionsverfahren für Audio Augmented Reality in einer städtischen Umgebung entwickelt, wobei das Ziel die Simulation einer virtuellen Schallquelle in hoher Immersion ist. Mit Hilfe des Produktionsverfahrens wird eine reale Schallquelle aus dem Inneren eines Gebäudes virtuell draußen vor demselben Gebäude abgebildet. Es geht speziell darum, eine Schallquelle aus dem Inneren eines Gebäudes aufzunehmen und mit Hilfe von Spatial Audio virtuell vor dem Gebäude abzubilden, sodass ein Hörer, der vor dem Gebäude steht, in das Gebäude „hineinhören“ kann, ohne es zu betreten. „Hineinhören“ soll in diesem Kontext nicht bedeuten, dass die Außenwände des Gebäudes akustisch transparent gemacht werden, sondern die Schallquelle soll vor dem Gebäude hörbar sein. Die Schallquelle wird aus dem Gebäude „herausgeholt“ und vor dem Gebäude platziert. Dieses Gebäude wird im Folgenden als Zielgebäude bezeichnet. Um den Ort, an dem die reale Schallquelle aufgenommen wird von dem Ort, an dem die virtuelle Schallquelle abgebildet wird, begrifflich abzugrenzen, werden die folgenden Bezeichnungen verwendet: Der Ursprungsort ist der Ort, konkret der Innenraum eines Gebäudes, in dem sich die Schallquelle originär befindet und aufgenommen wird. Der (virtuelle) Hörereignisort ist der Ort, an dem das Hörereignis stattfindet, konkret der Ort draußen vor dem Gebäude, an dem das Klangobjekt virtuell reproduziert wird. Das Klangobjekt wird mithilfe des Produktionsverfahrens vom Ursprungsort an den virtuellen Hörereignisort übertragen. Die reale Schallquelle wird in eine virtuelle Schallquelle umgewandelt. Ein denkbare Anwendungsszenario ist z.B. der touristische Sektor. Museen und Kultureinrichtungen können sich auch außerhalb der Öffnungszeiten nach außen hin präsentieren und das Interesse der Besucher für eine Ausstellung wecken, indem sie z.B. den Klang eines historischen Exponats erlebbar machen. Dies kann die Attraktivität des Ortes erhöhen und einen Beitrag zur Interaktivität der Stadt leisten.

Im Rahmen der Arbeit werden drei verschiedene experimentelle Ansätze zur Simulation einer virtuellen Schallquelle in einer urbanen Umgebung entwickelt und in der praktischen Umsetzung getestet. Das Hauptaugenmerk liegt dabei auf der Entwicklung von Prozessen zur Aufzeichnung und Wiedergabe und der Implementierung von Technologien zur Erzeugung von Spatial Audio. Logistische und organisatorische Aspekte werden am Rande betrachtet. In einem Hörversuch wird die Eignung der Verfahren, einen hohen Grad an Immersion beim Hörer zu erreichen, verifiziert. Konkret aufgenommen und virtuell abgebildet wird der Klang einer historischen Setzmaschine aus dem Museum der Arbeit.

2 Theorie

In diesem Abschnitt werden zunächst Kriterien festgelegt, die das Produktionsverfahren erfüllen soll. Die Schlüsselbegriffe werden erläutert und es wird der aktuelle Forschungsstand in diesem Gebiet beleuchtet. Schließlich wird die technische Grundlage der Produktionsverfahren vorgestellt.

2.1 Kriterien

Das Produktionsverfahren soll folgende Kriterien erfüllen: Es soll variabel auf verschiedene Gebäude in der Stadt anwendbar sein. Das Klangobjekt soll vor dem Gebäude abgebildet werden. Es soll sich im engeren Gesichtsfeld eines Hörers befinden, der frontal vor dem Gebäude steht. Die Wiedergabe soll von einem festen Standort aus über Kopfhörer erfolgen. Es sollen akustisch transparente Kopfhörer verwendet werden, um die Umgebungsgeräusche und die äußere Umwelt wahrnehmen zu können. Eine Isolation von der Realität (wie bei Virtual Reality) soll vermieden werden. Die Lokalisierung der Schallquelle soll außerhalb des Kopfes des Hörers sein und die Schallquelle soll aus einer bestimmten Richtung und Entfernung zum Hörer erscheinen. Die Schallquelle soll ortsstabil sein und ihre Position bei Kopfbewegungen nicht ändern, deswegen wird Headtracking implementiert. Es sollen 3 Freiheitsgrade in der Bewegung erreichbar sein. Durch die Einbeziehung der akustischen Eigenschaften der Außenwand des Zielgebäudes und der umgebenden Wände angrenzender Gebäude soll der realistische Eindruck erzeugt werden, die Schallquelle befände sich tatsächlich dort.

2.2 Begriffserklärung

Ein Verfahren zur Produktion von Audio Augmented Reality in einer urbanen Umgebung, dessen Ziel es ist, eine Schallquelle örtlich und zeitlich zu transformieren, umfasst alle Schritte und Methoden, um dieses Vorhaben umzusetzen. Ein Verfahren ist charakterisiert durch eine festgelegte Reihenfolge von Schritten, um ein bestimmtes Ziel zu erreichen. Die Verfahrensschritte sollen dabei wiederholbar sein und zu vergleichbaren Ergebnissen führen. Unter einer urbanen Umgebung ist eine städtische Umgebung im Freien mit charakteristischer Stadtarchitektur zu verstehen. Im Gegensatz zu ländlichen Gebieten ist die städtische Umgebung durch eine dichte Bebauung mit hohen Gebäuden und versiegelte Flächen gekennzeichnet. Die Anordnung der Gebäude und das Material der Fassaden und Straßen bietet somit ideale Voraussetzungen, um eine virtuelle Schallquelle akustisch durch Reflexionen in die Umgebung zu integrieren. Durch die große Zahl an kulturellen Einrichtungen in der Stadt ergeben sich vielfältige Anwendungsmöglichkeiten für ein solches Produktionsverfahren.

Wie lässt sich der Begriff Augmented Reality (eng.: „erweiterte Realität“) einordnen? Während das Konzept von Virtual Reality (eng.: „virtuelle Realität“) als Teil des allgemeinen Sprachgebrauchs weitgehend bekannt ist, gibt es oft Verwirrungen, was Augmented Reality bedeutet. Laut Azuma (1997, S. 2) ist Augmented Reality eine Variante von Virtual Reality. Virtual Reality-Technologien ermöglichen ein vollständiges Eintauchen in eine künstliche Umgebung. Die reale Welt wird nicht mehr

wahrgenommen. Im Gegensatz dazu bleibt bei Augmented Reality die physische Umwelt sichtbar und kann mit virtuellen Objekten erweitert werden. Das Ziel ist es beim Nutzer die Illusion hervorzurufen, dass die virtuellen Objekte Teil der Wirklichkeit sind (Azuma, 1997, S. 2). Nutzer von Augmented Reality können dabei weiterhin ihr peripheres Sehen und andere Sinne benutzen. Azuma (1997, S. 2) definiert drei charakteristische Merkmale von Augmented Reality Systemen, die unabhängig von der verwendeten Displaytechnologie sind: Die Kombination von Realität und Virtualität, Interaktivität in Echtzeit und die Verortung im 3-dimensionalen Raum. (Azuma, 1997, S. 2) Die Hauptkomponenten eines Augmented Reality Systems bestehen laut Billinghurst et al. (2015, S. 77) aus einer Display-Technologie, die reale und virtuelle Bilder verbinden kann, einem Computersystem das in der Lage ist interaktive Bilder zu erzeugen und in Echtzeit auf Eingaben reagiert und einer Tracking Technologie, die die Position des Nutzers erkennen und das virtuelle Bild in der Realität verankern kann (Billinghurst et al., 2015, S. 77).

Wie lässt sich nun Augmented Reality nun im Kontext von Audio definieren? Es ist schwierig eine allgemeingültige Definition für Audio Augmented Reality zu finden, aber es existieren zahlreiche Beschreibungen: Yang et al. (2022, S. 788) beschreiben Audio Augmented Reality als eine Technologie, die virtuelle auditive Inhalte in die physische Welt integriert. Die auditive Realität des Hörers wird ergänzt. Zur Nachahmung der realen Hörwahrnehmung in Bezug auf Richtung und Distanz werden virtuelle Klänge oft binaural im Raum verteilt und mit Nachhall versehen (Yang et al., 2022, S. 788). Das bedeutet, dass virtuelle Audiosignale so in die physische Umgebung eingebettet werden, dass sie aus bestimmten Richtungen und Entfernungen zu kommen scheinen, was ein immersives Hörerlebnis schaffen kann. Nach Jakka et al. (2004, S. 619) unterscheiden sich reale und virtuelle Audioumgebungen grundlegend dahingehend, dass virtuelle Klänge künstlich kreiert werden oder aus einer anderen Umgebung stammen. Audio Augmented Reality kombiniert diese Aspekte in einer Art, bei der reale und künstliche Klangszene gemischt werden, sodass virtuelle Klänge als eine Erweiterung der natürlichen Klänge empfunden werden (Jakka et al., 2004, S. 619). Die für die Umsetzung von Audio Augmented Reality benötigten Technologien sind nicht trivial. Die Verwendung einer vorberechneten virtuellen Szene in Virtual Reality Szenarien erleichtert das Rendern von Audioinhalten, wohingegen es schwieriger ist, virtuelle Klänge in die physische Umwelt zu integrieren und in Echtzeit an den Nutzer anzupassen (Yang et al., 2022, S. 788).

2.3 Forschungsstand

Ein wissenschaftlicher Artikel aus dem AES Journal zum Thema Audio Augmented Reality untersucht systematisch Veröffentlichungen zu Technologien, Anwendungen und Forschungsfragen. Dabei liegt der Fokus auf der Verwendung von Spatial Audio. Der Artikel identifiziert 5 Hauptkomponenten von Audio Augmented Reality Systemen: Anzeigetechnologie, Tracking, Interaktion, räumliche Klangsynthese und Modellierung der Raumakustik. Die Autoren des Artikels betonen, dass die Untersuchung Lücken aufweist, da einige der untersuchten Papers die verwendeten

Technologiekomponenten nicht spezifizierten, bzw. andere Schlagwörter benutzen, um AR-bezogene Forschung zu beschreiben. Außerdem liegt der Schwerpunkt auf Forschungsarbeiten und nicht auf kommerziellen Praktiken, sodass einige Arbeiten möglicherweise nicht mit einbezogen wurden.

Laut Yang et al. (2022, S. 794) sind rund 61 % der betrachteten AAR-Systeme allein auf die Wiedergabe von Audio fokussiert. In diesen Fällen werden virtuelle Klänge auf Endgeräten (z.B. Smartphones, Tablets und PCs) berechnet und dem Anwender über kabelgebundene/drahtlose Kopfhörer, In-Ears oder Knochenschall-Kopfhörer dargeboten (Yang et al., 2022, S. 794). Etwa 40 % der AAR-Systeme kombinieren Audio- und visuelle Augmentation unter Verwendung anderer Anzeigemethoden (z. B. HMDs und Handheld-Displays) (Yang et al., 2022, S. 797).

In 43% der betrachteten Arbeiten wird für das Tracking ein einziger Sensortyp benutzt, wobei das visuelle Tracking am häufigsten eingesetzt wird (Yang et al., 2022, S. 790). Nach Yang et al., (2022, S. 793) sind Trägheitssensoren der am zweithäufigste verwendete Sensortyp in dieser Statistik. Trägheitssensoren umfassen in der Regel Beschleunigungsmesser und Gyroskope, um die Position und Orientierung eines Objekts in Bezug auf einen bekannten Ausgangspunkt, die Ausrichtung und die Geschwindigkeit zu messen. Diese Sensoren dienen vor allem der Verfolgung der 3 Freiheitsgrade in kleinen Bereichen, in denen der Benutzer seine Position kaum verändert (Yang et al., 2022, S. 793). Allerdings werden Hybride Tracking Methoden, die die Stärken verschiedener Sensortypen kombinieren, favorisiert (Yang et al., 2022, S. 797).

Der Großteil der Systeme benutzt implizite Interaktion, daneben kommen auch 2D- und 3D-Schnittstellen sowie natürliche Körpereingaben zum Einsatz (Yang et al., 2022, S. 797). Implizite Interaktionen werden nicht aktiv durch den Nutzer initiiert, sondern das System reagiert auf die Aktionen des Nutzers in der Umgebung und passt die virtuellen Klänge entsprechend an (Yang et al., 2022, S. 794).

Die überwiegende Mehrheit der untersuchten AAR-Systeme verwendet generische HRTFs für die Synthese von Spatial Audio. Nur einige Arbeiten verwenden die BRIRs des Nutzers oder dessen individuelle HRTFs (Yang et al., 2022, S. 797). Die Kopfbezogene Übertagungsfunktion (engl.: „Head-related Transfer Function“) beschreibt die komplexe Filterwirkung, die Schallwellen durch Torso, Kopf und Ohren von ihrem Entstehungsort bis zum Gehörgang erfahren (Zhang et al., 2021, S. 1). Die Vorteile der Ambisonics-Technik, wie die flexible Anpassung an verschiedene Lautsprechersysteme führen dazu, dass Ambisonics bevorzugt für die Mischung von räumlichem Audio eingesetzt wird (Corrêa De Almeida et al., 2023, S. 266).

Etwa die Hälfte der in dem Artikel von Yang et al. (2022, S. 797) untersuchten AAR-Systeme modelliert keine Raumakustik, die andere Hälfte tendiert dazu, nur in Annäherung künstliche Nachhalleffekte zu erzeugen (Yang et al., 2022, S. 797). Die Autoren des Artikels schreiben, dass 27% aller betrachteten AAR-Anwendungen für eine Umgebung im Freien konzipiert sind, aber keine dieser Anwendungen die Raumakustik berücksichtigt (Yang et al., 2022, S. 795). Laut Yang et al. (2022, S. 796) sind passende

Reflexionsmodellierungen aber auch für Außenanwendungen bedeutend, um virtuelle Klänge nahtlos in eine reale Umgebung einzubetten.

Ein explizites Produktionsverfahren oder eine Anwendung zur Abbildung einer realen Schallquelle aus einem Innenraum in eine städtische Umgebung im Freien unter Einbeziehung der raumakustischen Eigenschaften der umgebenden Gebäude habe ich bei meiner Recherche der Fachliteratur nicht gefunden. Daher werden in dieser Arbeit verschiedene Ansätze ausprobiert, um ein solches Produktionsverfahren zu entwickeln. Allerdings gibt es ein Konferenzpapier aus dem Bereich Gesundheitsfürsorge, in dem der umgekehrte Ansatz verfolgt wird (Joshi et al., 2020). Anstatt ein Klangobjekt aus dem Inneren eines Gebäudes nach draußen zu transportieren, wird eine immersive Klanglandschaft aus der freien Natur in den Innenraum verlegt. Die Studie richtet sich an Menschen mit geriatrischen Depressionen, die in betreuten Einrichtungen leben und deren Möglichkeiten zur Teilnahme von Aktivitäten in der freien Natur eingeschränkt sind. Unter dem Einsatz von Kopfhörern mit Knochenleitung und Gestensteuerung wird eine interaktive Klangszene geschaffen, die ein immersives Naturerlebnis in Innenräumen ermöglichen soll. Ziel ist es, dass diese Menschen von den gesundheitsfördernden Eigenschaften der Naturgeräusche profitieren (Joshi et al., 2020, S. 1-2).

2.4 Szenenbasierte Produktion in Ambisonics mit binauraler Dekodierung

Als technische Grundlage für die Produktionsverfahren nutze ich die Ambisonics-Technik. Die Produktion der Audios findet in Ambisonics statt. Dabei werden sowohl direkt Ambisonics-Aufnahmen 1. Ordnung realisiert als auch Monoaufnahmen ambisonisch enkodiert. Für die Berechnung der Kopfhörersignale wird am Ende der Prozesskette ein Binaural Dekoder eingesetzt. Dieser verwendet meistens generische HRTFs, es gibt aber auch die Möglichkeit, individuell gemessenen HRTFs zu verwenden. (McKeag & McGrath, 1996, S. 6). Der Clou dieses revolutionären Ansatzes besteht in der Trennung des Syntheseschrittes (Aufnahme/ Processing) von der Wiedergabe (Binauraler Dekodierung). Das Verfahren bietet laut McKeag & McGrath (1996) eine Reihe von Vorteilen gegenüber anderen Techniken: So ist die Klangfelddrehung trivial und erfordert minimalen Rechenaufwand. Sie lässt sich durch Multiplikation mit Rotationsmatrizen in der ambisonischen Domäne realisieren, noch bevor die Signale mit HRIRs gefaltet werden. Dies ermöglicht die Integration von Head-Tracking in Echtzeit mit geringer Latenz. Die Technik ermöglicht außerdem eine höhere räumliche Auflösung als das Binaurale Rendering. Der HRTF-Datensatz ist kompakter und kann während der Laufzeit mit Kopfhörer-EQs kombiniert werden, da nicht für jede kleine Kopfbewegung neue HRTFs berechnet werden müssen. Stattdessen wird zwischen vorab berechneten HRTFs umgeschaltet (McKeag & McGrath, 1996, S. 5). Außerdem handelt es sich um ein offenes Verfahren. Es bietet die Möglichkeit, traditionelle kanalbasierte Aufnahmeverfahren in Ambisonics einzubetten. Die Dekodierung ist unabhängig von der Kanalanzahl und Ordnung für verschiedenste Lautsprecherkonfigurationen möglich und nicht allein auf die Wiedergabe für Kopfhörer beschränkt

(McKeag & McGrath, 1996, S. 6). Ich habe mich für den Einsatz dieser Technik entschieden, weil sie meine Anforderungen an eine Audio Augmented Reality Anwendung erfüllt:

- Wiedergabe für Kopfhörer
- Externalisierung (durch generische HRTFs)
- Interaktion in Echtzeit
- Klangfelddrehung durch die Integration eines Headtrackers
- Einbettung herkömmlicher Aufnahmetechniken
- Vergleichsweise wenig Rechenleistung nötig

Bei konventionellen binauralen Methoden wie der binauralen Aufnahme mit Originalkopf-Mikrofonen oder der Messung von HRIRS/BRIRS mit einem Kunstkopf und der anschließenden Faltung wird eine statische Klangszene generiert, die gewissermaßen mit dem Kopf des Hörers verankert ist. Der Kopf ist Mittelpunkt und Bezugspunkt dieses dreidimensionalen Klangraums und in seiner Bewegung fest mit ihm verbunden. In der auditiven Realität bewegen sich stationäre Schallquellen aber nicht parallel mit der Kopfbewegung. Damit der Eindruck einer ortsfesten Schallquelle entsteht, muss die gesamte Klangszene den Bewegungen des Kopfes entgegen gedreht werden (Armstrong & Kearney, 2022, S. 126). Aus diesem Grund wird Head-Tracking implementiert, um die Kopfbewegungen zu verfolgen. Die Invertierung der Klangszene geschieht dann mithilfe einer Software, die die Daten des Headtrackers verarbeitet. Die Einbindung des Headtrackers ermöglicht eine Änderung der Orientierung im Raum durch Rotation um die 3 kartesischen Achsen. Dadurch werden 3 Freiheitsgrade in der Bewegung erreicht. Um die virtuelle Schallquelle so realistisch wie möglich in die Umgebung einzufügen, wird die umliegende Raumakustik mit einbezogen. Die frühen Reflexionen der umgebenden Wände werden simuliert und das Klangspektrum der Reflexionen wird gefiltert. Diese Maßnahmen sollen zur Plausibilität von Höreindruck und visueller Umgebung beitragen.

3 Planung

Im Rahmen der Arbeit werden drei verschiedene Produktionsverfahren für Audio Augmented Reality entwickelt, getestet und in ihrer Wirkung miteinander verglichen. Die Produktionsverfahren werden in diesem Abschnitt beschrieben. Gemeinsamer Ausgangspunkt für die ersten beiden Verfahren ist eine Monoaufnahme der realen Schallquelle am Ursprungsort. Diese Aufnahme wird dann je nach Verfahren unterschiedlich verarbeitet, und soll dabei möglichst originalgetreu und ohne klangliche Verfärbungen reproduziert werden. Aus diesem Grund sollte das Ursprungssignal eine einwandfreie technische Qualität besitzen, also fehlerfrei sein und so angesteuert werden, dass es innerhalb des optimalen Dynamikbereichs liegt. Deshalb muss die Schallquelle einen ausreichend hohen Schallpegel besitzen. Die Aufnahme sollte so trocken wie möglich sein, deshalb sollte in Nahdistanz möglichst viel Direktschall und wenig Diffusschall aufgenommen werden. Eine akustische Behandlung des Ursprungsorts, die für die Produktionsverfahren nicht praktikabel ist, gilt es zu vermeiden. Um die

Schallquelle möglichst originalgetreu aufzuzeichnen, sollte das verwendete Mikrofon einen linearen Frequenzgang aufweisen. Der Frequenzgang der Kopfhörer sollte entzerrt werden, um die richtungsbestimmenden Frequenzbänder der HRTFs nicht zu verändern, die für die Lokalisierung entscheidend sind.

3.1 Verfahren A: Software gestützte Raumsimulation

Für die Simulation des virtuellen Raums wird die IEM Plug-in Suite vom Institute of Electronic Music and Acoustics in Graz verwendet. Die Software erfüllt genau meine Kriterien für die Simulation einer virtuellen Schallquelle:

- Der Room Encoder wurde als Plug-in für Raumsimulation entwickelt und ist speziell für die Kopfhörer Wiedergabe gedacht.
- Es gibt die Möglichkeit einen Headtracker zu integrieren, um das Klangfeld zu drehen.
- Die Plug-ins lassen sich in die Multikanal-DAW Reaper integrieren.
- Die Plug-ins sind alle aufeinander abgestimmt und lassen sich in Reihe schalten.
- Die Plug-ins bieten eine hohe räumliche Auflösung. (Die Enkodierung der Signale ist bis zur 7. Ordnung möglich.)
- Die Software ist sehr gut dokumentiert und es gibt Online-Support durch die Entwickler.
- Die Software ist kostenlos erhältlich.

Bei diesem Verfahren wird die Monoaufnahme ambisonisch enkodiert und die Raumakustik wird mit verschiedenen Plugins simuliert. Die raumakustischen Parameter (Raumdimensionen, Position von Schallquelle und Hörer, Reflexionseigenschaften der Wände, Klangspektrum der Reflexionen) können modelliert werden. Die Verarbeitung des Monosignals erfolgt durch eine Kette von Software-Geräten. Im Folgenden wird der Weg des Signals beschrieben, wobei ich auf die Modellierung des simulierten Raums genauer eingehen werde. Daniel Rudrich (o. D.) beschreibt in der Dokumentation zur IEM Plug-in Suite (Version 1.14.1) detailliert die Funktionen und Anwendungsmöglichkeiten der Software.

Toolbox

Die Toolbox steht am Anfang der Geräteketten und fungiert als Verstärker. Mit diesem Plugin wird der Signalpegel angehoben.

Coordinate Converter

Der Coordinate Converter konvertiert die VST-Parameter, die als Kugelkoordinaten vorliegen in kartesische Koordinaten, die vom Room Encoder verarbeitet werden können. Es findet keine Audioverarbeitung statt. Die berechneten Kartesischen Koordinaten werden auf die Position der Quelle im Room Encoder abgebildet. Mithilfe des Sphere Panners wird die Quelle in einem festgelegten Radius

um die Hörerposition rotiert. So wird bei verschiedenen Azimuten eine äquidistante Entfernung zur Hörerposition sichergestellt.

Directivity Shaper

Der Directivity Shaper ist ein Tool zur Verteilung der Signalenergie. Primär dient er zur Enkodierung und Formung der Richtwirkung der Schallquelle. Außerdem kann das Signal gefiltert und räumlich geschwenkt werden. Das Monosignal wird in vier unabhängigen Frequenzbändern gefiltert. Jedes Band wird mit spezifischen Gewichten versehen, um das Strahlungsmuster zu formen. Dabei werden die unteren Frequenzbänder weniger gebündelt und strahlen ungerichteter, zu den hohen Frequenzen hin nimmt die Richtwirkung zu. Bei der Basic-Gewichtung verteilt sich die Signalenergie auf eine schmale Hauptkeule, eine Gegenkeule mit invertierter Phase und verschiedene Seitenkeulen. Sie besitzt die beste Richtwirkung nach vorne, aber es wird auch Energie zu den Seiten und nach hinten abgestrahlt. Es ist entscheidend, dass die Hauptkeule der Richtcharakteristik schmal ist, um eine präzise Lokalisation und eine geringe wahrgenommene Quellbreite an der zentralen Hörposition zu gewährleisten [...] (Frank & Zotter, 2022, S. 7). Mittlere und hohe Frequenzbänder werden mit Maximum-Re-Energy (Max-r_E) und In-Phase gewichtet. Bei der Max-r_E-Gewichtung werden die Gegenkeulen und die Seitenkeulen zum Preis einer breiteren Hauptkeule gedämpft. Bei der In-Phase-Gewichtung verschwinden alle Invertierten und Seitenkeulen, aber die Hauptkeule wird noch breiter und ist weniger nach vorne gerichtet. Die Bänder können mit unterschiedlichen ambisonischer Ordnung enkodiert werden. Mit dem Sphere Panner wird allen Bändern eine gemeinsame Richtung zugewiesen. Das Resultat ist ein gerichtetes ambisonisches Signal mit Gewichtungen, das als Directivity-Signal bezeichnet wird. Das Directivity-Signal geht als Eingangssignal in den Room Encoder und wird dort zur virtuellen Schallquelle.

Room Encoder

Der Room Encoder ist das zentrale Werkzeug zur Raumsimulation und zur Synthese der virtuellen Schallquelle. Der Room Encoder simuliert einen dreidimensionalen Raum mit sechs Wänden („Shoobox-shaped Room“, eng.: „Schuhschachtelförmiger Raum“). Die Schallquelle wird zusammen mit der Hörerposition in einem virtuellen Raum synthetisiert. Schallquelle und Hörer lassen sich frei im Raum positionieren und bewegen. Der Direktschall und die frühen Reflexionen der umgebenden Wände werden simuliert. Der Schallpegel des Direktschalls ändert sich mit der Entfernung zur Schallquelle. Die Anzahl der Wände (und damit der Reflexionen) kann festgelegt werden. Die Dämpfung dieser Signale ist abhängig von der Distanz zwischen Schallquelle und Hörer und den Reflexionseigenschaften der Wände. Die Reflexionseigenschaften der Wände können angepasst werden, die Wandreflexionen können sogar komplett gedämpft werden, sodass die entsprechende Wand akustisch gar nicht mehr hörbar ist. Mit einem Shelving-Filter kann zudem das Klangspektrums der Reflexionen bearbeitet werden. Es handelt sich bei dem Room Encoder um ein leistungsstarkes Plugin zur Raumsimulation, es gibt jedoch Grenzen: So lassen sich nur quaderförmige Räume simulieren und die räumlichen

Dimensionen sind limitiert auf ein Volumen von 30 m x 30 m x 20 m (einstellbar auf 2 Nachkommastellen).

Scene Rotator

Der Scene Rotator ist ein Plugin zur Drehung der ambisonischen Klangszene. Er empfängt und verarbeitet Rotationsdaten vom Headtracker. Die Daten werden in Echtzeit als OSC-Messages versendet. Diese Daten und die drei Rotationsachsen können invertiert werden. Die Daten werden invertiert, um das Klangfeld entgegen den Kopfbewegungen zu rotieren. So soll der Eindruck einer ortsstabilen Schallquelle erzeugt werden. Der Scene Rotator ist somit ein entscheidendes Werkzeug zur Generierung einer authentischen virtuellen Schallquelle. Der Scene Rotator wird für alle Produktionsverfahren gleichermaßen benötigt. Er ist das letzte Glied in der Kette der ambisonischen Signalverarbeitung vor der Dekodierung.

Binaural Dekoder

Allen Verfahren gemein ist die Dekodierung zur Wiedergabe für Kopfhörer als finaler Verarbeitungsschritt. Hier erfolgt die Umwandlung der ambisonisch enkodierten Signale in ein binaurales Signal, das über Kopfhörer wiedergegeben werden kann. Dabei werden die Signale mit Hilfe von vorberechneten HRTFs direkt aus dem B-Format zu binauralen Signalen konvertiert, ohne sie vorher auf virtuelle Lautsprecher zu dekodieren. Die HRTFs werden so bearbeitet, dass der ursprüngliche Frequenzgang bei der Verwendung mit Ambisonics so gut wie möglich erhalten bleibt. Der Dekoder verwendet generische HRTFs, die mit dem Kunstkopf KU 100 der Firma Neumann gemessen wurden. Zusätzlich gibt es die Möglichkeit Frequenzgang Entzerrung für verschiedene Kopfhörer-Modelle anzuwenden, die in Messungen zusammen mit dem KU 100 ermittelt wurden (Rudrich, o. D.). Zur korrekten Wiedergabe der HRTFs ohne klangliche Verfärbungen wird für den Kopfhörer (Beyerdynamik DT 880 Pro), der beim Hörversuch eingesetzt wird, Frequenzgang Entzerrung angewendet. Abbildung 1 gibt einen Überblick über den Signalfluss des Produktionsverfahrens.

3.2 Verfahren B: Originalakustische Beschallungsaufnahme am Hörereignisort

Ausgangspunkt für das Verfahren ist die Monoaufnahme der Schallquelle des Ursprungsortes. Diese Aufnahme wird über einen gerichteten Lautsprecher am Hörereignisort wiedergegeben und mit einem Soundfield-Mikrofon erneut aufgenommen (siehe Abb. 2). Besser gesagt wird der Hörereignisort mit dem aufgenommenen Signal beschallt. Dabei wird der Lautsprecher in Richtung des Mikrofons ausgerichtet. Die Intention dieses Ansatzes ist es, die realen raumakustischen Eigenschaften am Hörereignisort aufzuzeichnen. Die Originalakustik des Ortes soll in der Aufnahme hörbar werden. Frühe Reflexionen, die durch die Interaktion des Schalls mit den angrenzenden Wänden entstehen und der Diffusschall tragen dazu bei, die Schallquelle aus dem Inneren des Gebäudes raumakustisch in die neue Umgebung zu integrieren und eine Verschmelzung zu ermöglichen.

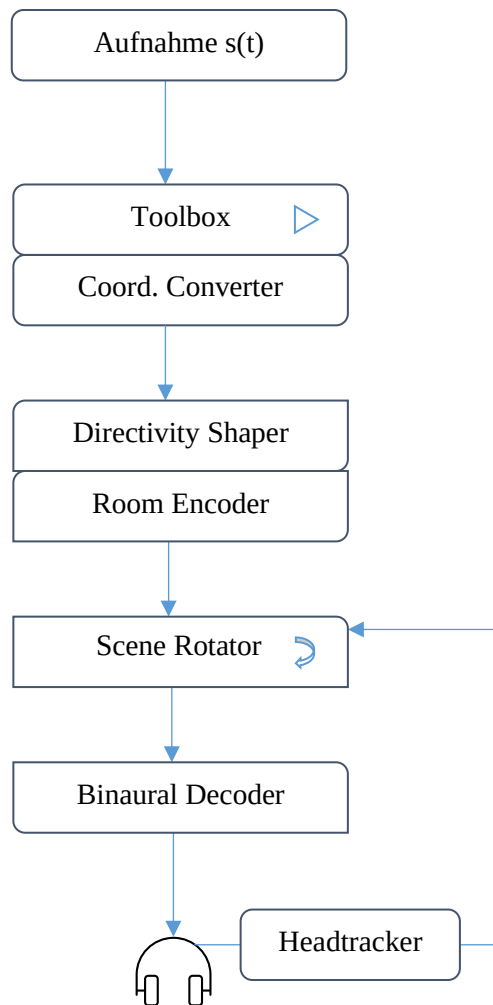


Abbildung 1: Signalfluss Verfahren A
Quelle: Eigene Darstellung

Um genau diese Signalanteile zu generieren, muss der Schalldruckpegel des Lautsprechers ausreichend hoch eingestellt werden. Aus eben diesem Grund sollte der Lautsprecher nahe den Wänden des Zielgebäudes aufgestellt werden. Die Lautsprecheraufstellung sollte mit Bedacht gewählt werden, da die Aufnahmen destruktiv sind und keine nachträgliche Änderung der Positionen möglich ist. Außerdem ist der Aufbau zeitintensiv und sollte daher nach Möglichkeit nur einmal gemacht werden.

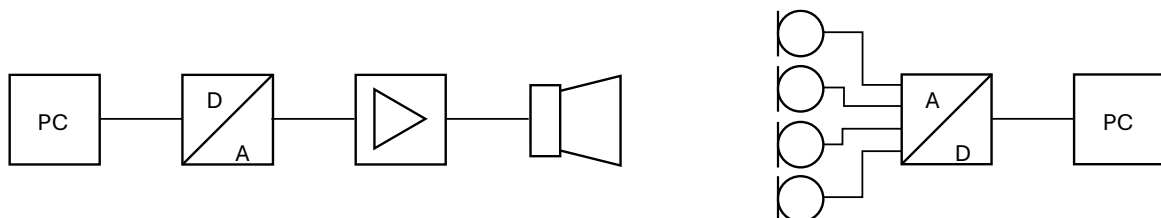


Abbildung 2: Blockschaltbild vom Aufbau
Quelle: Eigene Darstellung

Bei diesem Produktionsverfahren wird das Signal 4-mal gewandelt und durch eine Reihe von Geräten geführt. Dabei soll der Frequenzgang des Ursprungssignals möglichst gut erhalten bleiben. Deswegen ist darauf zu achten, dass die verwendeten Schallwandler lineare Frequenzgänge aufweisen. Es sollte

ein Lautsprecher mit einem linearen Frequenzgang verwendet werden, um die Schallquelle möglichst originalgetreu wiederzugeben. Da die Aufnahme im Freien in einem städtischen Umfeld stattfindet, ist damit zu rechnen, dass diffuse Umgebungsgeräusche und störende Frequenzen mit aufgezeichnet werden. Um dem vorzubeugen, sollte die Aufnahme zu einer Zeit gemacht werden, in der mit wenig Verkehr und Anliegern zu rechnen ist. Wenn dies nicht möglich ist, sollte die Aufnahme mit einem Gate zur Herabsenkung der Hintergrundgeräusche bearbeitet werden. Ein weiterer Nachteil dieses Verfahrens ist, dass sich mit jeder Wandlung das Grundrauschen addiert. Daher sollten die Geräte nach Möglichkeit einen hohen SNR besitzen. Die Höhe der räumlichen Auflösung ist abhängig vom verwendeten Mikrofontyp. Mit dem em64 Eigenmike des Herstellers mh acoustics z.B. lassen sich Aufnahmen bis zur 6. Ordnung realisieren! Der Signalfluss von Verfahren B ist in Abbildung 3 dargestellt.

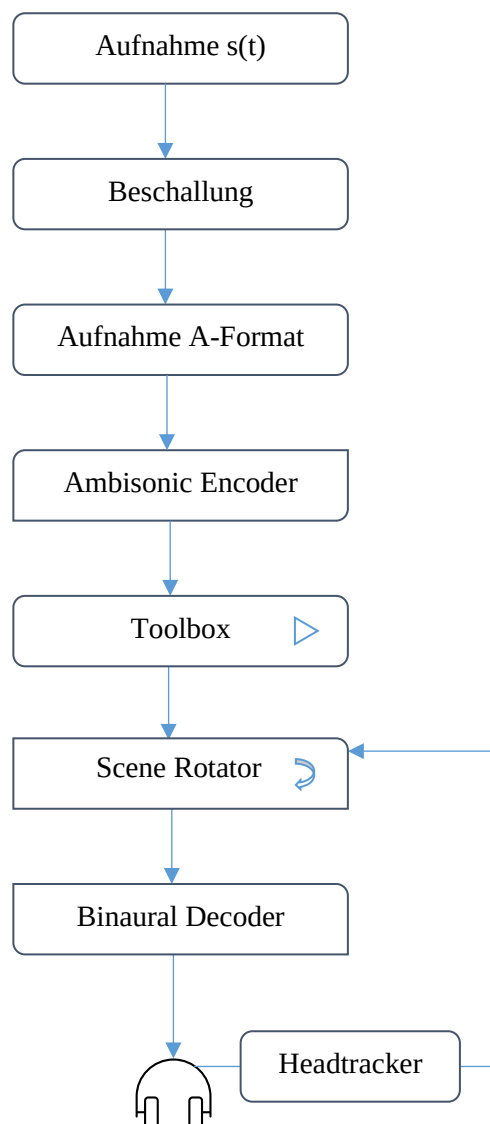


Abbildung 3: Signalfluss Verfahren B
Quelle: Eigene Darstellung

3.3 Verfahren C: Aufnahme mit Soundfield-Mikrofon am Ursprungsort

Die Schallquelle wird mit einem Soundfield-Mikrofon am Ursprungsort aus größerer Distanz aufgenommen. Die Verarbeitung der enkodierten Signale erfolgt analog zu Verfahren B. Die Idee hinter diesem Verfahren ist es, die Entfernungsinformation zwischen Schallquelle und Mikrofon bereits in der Aufnahme abzubilden. Die Aufnahmedistanz entspricht so der Distanz zwischen Hörer und Schallquelle am Hörereignisort. Dadurch ließe sich im Hinblick auf die Raumsimulation die Limitierung des Room Encoders in einer Dimension - der horizontalen Achse - umgehen. Als problematisch angesehen werden muss jedoch der Umstand, dass der Diffusschallanteil in der Aufnahme größer ist als der Direktschall, die Aufnahme ist weniger gerichtet. Das kann zu Schwierigkeiten bei der Lokalisierung führen. Außerdem ist die Raumakustik des Ursprungsortes in der Aufnahme hörbar, was zum „room divergence effect“ (eng.: „Raum Divergenz Effekt“) führen kann. Das bedeutet, dass das Höreindruck, der durch die Schallquelle hervorgerufen wird und der visuelle Raumeindruck nicht zusammenpassen, was der Immersion abträglich ist. Der Signalfluss von Verfahren C ist in Abbildung 4 dargestellt.

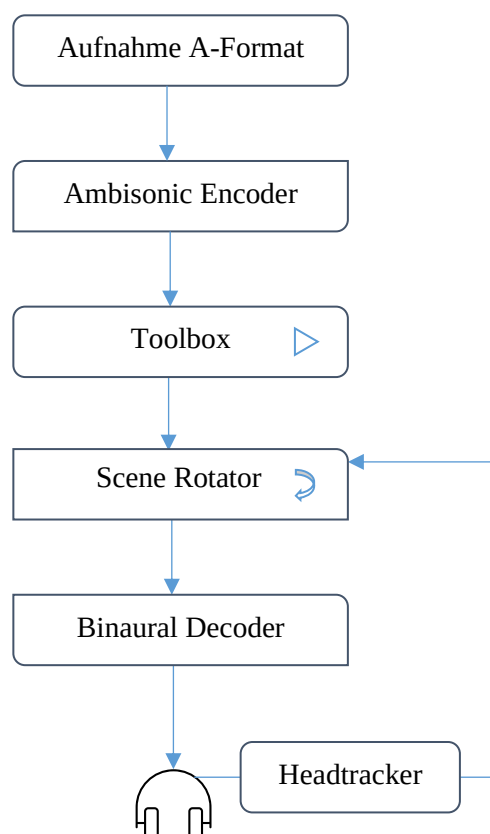


Abbildung 4: Signalfluss Verfahren C
Quelle: Eigene Darstellung

3.4 Ressourcenplanung

Tabelle 1: Ressourcen der 3 Produktionsverfahren

	Produktionsverfahren		
	Verfahren A: Software gestützte Raumsimulation	Verfahren B: Originalakustische Beschallungsaufnahme am Hörereignisort	Verfahren C: Soundfield- Aufnahme am Ursprungsort
Ausgangssignal	Monoaufnahme aus 1m Entfernung	Monoaufnahme aus 1m Entfernung	Soundfield-Aufnahme aus 8m Entfernung
Zeitlicher Aufwand	mittel (Aufnahme- Enkodierung- Verarbeitung- Dekodierung)	groß: mehrere Schritte nötig (Aufnahme- Wiedergabe Beschallung-Aufnahme- Enkodierung- Verarbeitung- Dekodierung)	klein (Aufnahme- Enkodierung- Verarbeitung- Dekodierung)
Finanzieller Aufwand	klein: wenige Geräte müssen gekauft werden	groß: Soundfield- Mikrofon und Monitorlautsprecher müssen zusätzlich gekauft werden	mittel: Soundfield- Mikrofon muss gekauft werden
Geräte	Mikrofon, Rekorder, PC, Interface, Headtracker Kopfhörer	Mikrofon, Soundfield- Mikrofon, Rekorder, PC, Interface, Headtracker, Kopfhörer Monitorlautsprecher	Soundfield-Mikrofon, PC, Interface, Headtracker, KH
Anzahl Wandlungen	2	4	2
Software	Reaper, Room Encoder, Directivity Shaper, Coordinate Converter, Scene Rotator, Toolbox, Binaural Dekoder, Headtracker-Software	Reaper, Scene Rotator, Toolbox, Binaural Dekoder, Headtracker - Software	Reaper, Scene Rotator, Toolbox, Binaural Dekoder, Headtracker - Software
Vorteile	Variable Anpassung der Raumakustik an örtliche Gegebenheiten, Raumakustische Parameter können variiert werden, nur 2 Wandlungen nötig	Realistische Abbildung der raumakustischen Bedingungen am Ereignisort, keine Limitierung der Dimensionen	Distanzinformation ist bereits in der Aufnahme, nur 2 Wandlungen nötig
Nachteile	Räumliche Dimensionen und andere Raumakustische Simulationsparameter sind limitiert	Aufnahme ist destruktiv, keine Änderung der räumlichen Parameter im Nachhinein möglich, 4 Wandlungen nötig, Summierung des Grundrauschens	Aufnahme ist destruktiv, keine Änderung der räumlichen Parameter möglich, kein trockenes Signal, Rauminformation steckt in der Aufnahme, hoher Diffusschallanteil

In Tabelle 1 sind die Ressourcen der 3 Produktionsverfahren gegenübergestellt. Vergleicht man die benötigten Ressourcen stellt sich Verfahren A als dasjenige mit dem geringsten finanziellen und logistischen Aufwand dar. Neben denen für alle Verfahren obligatorischen Geräte wird für die Aufnahme der Schallquelle lediglich ein Mikrofon benötigt. Der zeitliche Aufwand ist auf die Aufnahme und die anschließende Signalbearbeitung in der Simulationssoftware beschränkt. Durch die Variabilität der Software bietet das Verfahren große Vorteile gegenüber den anderen beiden Verfahren, allerdings gibt es auch Einschränkungen wie z.B. die Limitierung der räumlichen Dimensionen.

Im Gegensatz dazu werden für Verfahren B die meisten Ressourcen gebraucht: Der finanzielle Aufwand ist groß, denn es wird neben dem Mikrofon zusätzlich ein Lautsprecher und ein Soundfield-Mikrofon benötigt. Zeitlicher und logistischer Aufwand sind ebenfalls hoch, denn es müssen 2 Aufnahmen an verschiedenen Orten gemacht werden, bevor das Signal in die Software einfließt. Der große Vorteil besteht in der Aufzeichnung der realen raumakustischen Eigenschaften am Hörereignisort und es gibt keine Einschränkungen bezüglich der Architekturgeometrie oder der räumlichen Dimensionen. Das Verfahren lässt sich grundsätzlich an jedem städtischen Gebäude anwenden. Ein Nachteil ist u.a. die Summierung des Grundrauschens der Geräte durch die 4 Wandlungen.

Verfahren C besticht durch den geringsten zeitlichen Aufwand, neben der Aufnahme bedarf es nur wenigen Konfigurationen in der Software. Die Kosten sind überschaubar, es muss ein Soundfield-Mikrofon zur Aufnahme der Klangszene angeschafft werden. Die Intention, die räumliche Distanz zur Schallquelle bereits in der Aufnahme abzubilden, könnte sich jedoch auch als Nachteil erweisen, und zwar wenn die Akustik des Aufnahmeraums deutlich hervortritt. Wie bei Verfahren B sind die Aufzeichnungen permanent und erlauben keine nachträglichen Veränderungen der raumakustischen Parameter. Allen Verfahren gemein ist die abschließende Signalverarbeitung in der Software zur Drehung der Klangszene und zur Dekodierung der Signale.

4 Design des Hörversuchs

Die Produktionsverfahren sollen dahingehend überprüft werden, ob sie in der Lage sind, einen hohen Grad an Immersion beim Zuhörer zu erzeugen. Welches Verfahren ist dafür besser geeignet? Die dazu benötigten Daten werden in einem Hörversuch erhoben und anschließend statistisch ausgewertet. Um den Grad der Immersion zu untersuchen wird ein Vergleich mit den raumakustischen Eigenschaften einer realen Schallquelle angestrebt. Besonderes Augenmerk wird dabei auf die Wahrnehmung von Distanz und räumlicher Verortung gelegt. Laut Nicol (o. J., S. 102) ist die Entfernungswahrnehmung im speziellen Fall von Binauralem Audio von größerer Bedeutung. Gründe dafür sind zum einen die Fähigkeit der binauralen Wiedergabetechnik, verschiedene Distanzen wiederzugeben und auf der anderen Seite, Probleme mit der IKL und Externalisierung als ein häufiges Artefakt der binauralen Wiedergabe [...] (Nicol et al., o. J., S. 102). Die Suggestiv-Lokalisierung ist somit ein entscheidender Parameter für die Immersion. Mit ihr steht und fällt die Illusion, dass sich die Schallquelle außerhalb

des Kopfes befindet. Sie dient im Hörversuch als Prüfparameter, ob die Externalisierung funktioniert. Nicol et al. (o. J., S. 102) schreiben, dass der Standort einer Schallquelle bis zu einem gewissen Grad für einen Hörer direkt wahrnehmbar ist. Die Verortung oder räumliche Definition bezieht sich auf die Lokalisierbarkeit des Klangs und ist ein weiterer Gesichtspunkt, der beachtet werden sollte (Nicol et al., o. J., S. 102). Unter der Annahme, dass die Immersion von der Eigenschaft der virtuellen Schallquelle abhängig ist, wie eine reale Schallquelle wahrgenommen zu werden, wird eine Fokussierung auf die Bewertung der räumlichen Attribute der Schallquelle vorgenommen. Im Idealfall erzeugt die virtuelle Schallquelle beim Hörer denselben räumlichen Eindruck wie eine reale Schallquelle: Das bedeutet Sie wird außerhalb des Kopfes in einer bestimmten Entfernung zum Hörer und aus einer bestimmten Richtung kommend wahrgenommen. Sie befindet sich an einem festen Punkt im Raum, ist ortsstabil und ändert Ihre Koordinaten bei Kopfbewegung nicht. Die virtuelle Schallquelle hat eine begrenzte Ausdehnung, die sich nicht ändert und wird anhand von charakteristischer Klangfärbung und Reflexionen als Teil der umgebenden Raumakustik akzeptiert. Der visuelle Raumeindruck und das Hörereignis sind plausibel. Die Fähigkeit der Versuchsteilnehmer, die räumlichen Attribute einer virtuellen Schallquelle zu bestimmen, ist somit ein wichtiger Indikator für die Immersion. Der Grad der Immersion wird als hoch bewertet, wenn die virtuelle Schallquelle einer realen Schallquelle nahekommt. Daraus ergeben sich die folgenden Parameter, die für die Bewertung der Immersion relevant sind, wobei der Schwerpunkt auf die Lokalisierung und Distanzwahrnehmung gelegt wird:

- Lokalisierung
- Distanzwahrnehmung
- Ausdehnung
- Stabilität
- Suggestiv-Lokalisierung
- Plausibilität von Hörereignis und visuellem Raumeindruck

Unter Berücksichtigung dieser Gesichtspunkte wird ein Design für den Hörversuch entworfen. Es wird ein Fragebogen (siehe Anhang A) entwickelt, in dem die Wahrnehmung der einzelnen Parameter nacheinander abgefragt wird. Zur Überprüfung der Lokalisierung wird die Schallquelle in verschiedenen Winkeln um die Hörerposition herum positioniert. Es wird nur der Azimut betrachtet, die Elevation wird nicht untersucht. Die Positionen der Schallquelle wurden im Hinblick auf folgende Kriterien festgelegt: Die Schallquelle soll aus einer bestimmten Richtung und Entfernung in Bezug zum Hörer erscheinen. Das Klangobjekt soll vor dem Gebäude abgebildet werden. Es soll sich im engeren Gesichtsfeld eines Hörers befinden, der in Richtung des Zielgebäudes blickt. Außerdem soll die Schallquelle einmal direkt (2 m) vor der Wand aufgestellt werden, um frühere und stärkere Reflexionen zu generieren und einmal auf halber Strecke zwischen Wand (5 m vor der Wand) und Hörer aufgestellt werden, um ein höheres Verhältnis von Direktschallenergie zu Diffussschallenergie an der Hörerposition zu generieren. So

ergeben sich folgende Positionen für die virtuelle Schallquelle (bzw. die Aufstellung des Lautsprechers), die in Tabelle 2 zusammengefasst sind.

Tabelle 2: Positionen der virtuellen Schallquelle/Lautsprecherpositionen

	Winkel				
Distanz 5 m	45° 5 m	23° 5 m	0° 5 m	-23° 5 m	-45° 5 m
Distanz 8 m	45° 8 m	23° 8 m	0° 8 m	-23° 8 m	-45° 8 m

Im Hörversuch zeigen die Probanden mit einem Drumstick in die Richtung, in der Sie die Position der Schallquelle wahrgenommen haben. Die angezeigten Winkel werden vom Versuchsleiter notiert. Überprüft wird die Richtigkeit anhand der Übereinstimmung von wahrgenommener und tatsächlicher Position. Eine hohe Übereinstimmung wird mit einem hohen Grad an Immersion gleichgesetzt. Ähnlich verhält es sich mit der Distanzwahrnehmung, mit dem Unterschied, das hier 3 Antwortmöglichkeiten vorgegeben werden, um fehlerbehaftete Schätzungen zu vermeiden. Die Probanden müssen sich lediglich für eine Antwort entscheiden, die sich auf visuelle Anhaltspunkte in der Umgebung bezieht. Die Suggestiv-Lokalisierung als kritischer Parameter für die Immersion wird geschlossen abgefragt. Sie dient im Hörversuch zur Überprüfung, ob die Schallquelle externalisiert wahrgenommen wird und der Versuch auswertbar ist. Die Eindrücke zur Ausdehnung und Stabilität der Schallquelle werden mit passenden Attributen im semantischen Differential abgefragt, ebenso der Eindruck zur Plausibilität. Es wird eine 5-stufige Skala verwendet, wobei 5 einem positiven Attribut und 1 einem negativen Attribut, bezogen auf die Immersion entspricht. Summiert man die Häufigkeiten aller Bewertungen auf, erhält man eine Punktzahl für den Immersionsgrad. Der höchstmögliche Grad entspricht einer Punktzahl von 180. Dies ist der Fall, wenn alle Probanden jeden der 3 Parameter mit 5 bewerten. Bei 6 Versuchsteilnehmern und 2 abgegebenen Bewertungen pro Verfahren entspricht das 60 Punkten pro Parameter und damit einer Gesamtpunktzahl von 180 für alle 3 Parameter. Für die Auswertung wird folgender Maßstab zugrunde gelegt:

Tabelle 3: Maßstab für die Auswertung des semantischen Differentials

Bewertung	1	2	3	4	5
Immersionsgrad	Sehr gering	Gering	Mittel	Hoch	Sehr hoch
Punktzahl	36-71	72-107	108-143	144-179	180

Zur Bewertung des Immersionsgrads des gesamten Produktionsverfahrens wird folgender Maßstab angelegt:

Tabelle 4: Maßstab für die Auswertung des gewichteten Durchschnitts

Gewichteter Durchschnitt	5-26%	27-48%	49-70%	71-92%	93-100%
Immersionsgrad	Sehr gering	Gering	Mittel	Hoch	Sehr hoch

Die niedrigste Stufe beginnt bei 5% des gewichteten Durchschnitts aller Parameter, das entspricht der minimal erreichbaren Prozentzahl im Hörversuch. Dies ist der Fall, wenn bei der Auswertung des semantischen Differentials 36 Punkte ermittelt werden und die Richtigkeit von Lokalisierung und Distanzwahrnehmung 0% beträgt. Die ersten 4 Intervalle wurden in gleichgroße Stufen eingeteilt und so gewählt, das für einen mittleren Immersionsgrad ein gewichteter Durchschnitt von mindestens 49% erreicht werden muss. Die Größe des letzten Intervalls wurde kleiner gestaltet, sodass für den höchstmöglichen Grad an Immersion ein gewichteter Durchschnitt von mindestens 93% erreicht werden muss.

5 Realisierung

5.1 Aufnahme der Monotype im Museum der Arbeit

Im Vorfeld wurden mit dem Soundfield-Mikrofon einige Testaufnahmen gemacht, um die Ambisonics-Technik kennenzulernen und die gestalterischen Möglichkeiten in Bezug auf Raum und Distanz einschätzen zu können. Als Klangobjekt wurde die Monotype ausgewählt, eine Setzmaschine aus dem frühen 20. Jahrhundert. Das komplexe System, das auf druckluftbasierter Lochstreifensteuerung und präziser Lettern-Gießtechnik basiert, war seit 1900 über Jahrzehnte hinweg für den Satz anspruchsvoller Drucksachen verantwortlich. Auch Bücher für Technik und Wissenschaft wurden damit gefertigt. Die Qualität des Produkts ist dabei ebenso bemerkenswert wie der Geräuschpegel der laufenden Maschine. (Die Monotype, o. D.). Die Maschine erweist sich aufgrund des hohen Schallpegels und im Hinblick auf spätere Anwendungsmöglichkeiten der Produktionsverfahren im touristischen Sektor als geeignetes Klangobjekt. Im Museum der Arbeit in Barmbek finden regelmäßig Vorführungen der historischen Setzmaschine statt. Durch die freundliche Kooperation der Mitarbeiter vor Ort waren Aufnahmen problemlos möglich. Der Ausstellungsraum, indem die Monotype vorgeführt wird, ist der Ursprungsort. Da der Hörversuch im Sitzen durchgeführt werden soll, und die Schallquelle sich auf Höhe der Ohren befinden soll, wurden beide Mikrofone auf einem Stativ in einer Höhe von 1.3 m über dem Boden befestigt, was in etwa der Kopfhöhe eines sitzenden Menschen entspricht. Um Übersteuerungen zu vermeiden, wurden Dualaufnahmen durchgeführt, dabei werden zeitgleich zwei Aufnahmen vom Eingangssignal mit unterschiedlichen Eingangspegeln erstellt. Zusätzlich wurde ein Hochpassfilter mit

der Grenzfrequenz 100 Hz vorgeschaltet, um evtl. auftretende, tieffrequente Störgeräusche herauszufiltern. Es wurden zwei Aufnahmeszenarios mit unterschiedlichem Aufbau durchgeführt:

Beim 1. Szenario wurden die Aufnahmen aus 1 m Entfernung mit einem Kondensatormikrofon (sE Z3300a) und einem Linear-PCM-Recorder (Tascam DR-40x) durchgeführt. Für die Aufnahmen wurde Nierencharakteristik verwendet, um nach vorne gerichtet möglichst viel Direktschall aufzunehmen und gleichzeitig seitwärtige und rückwärtige Diffusschallanteile im Pegel abzusenken. Das Mikrofon wurde innerhalb des Hallradius aufgestellt, um möglichst viel Direktschall und wenig Diffusschall aufzunehmen.

Beim 2. Szenario wurde die Aufnahme aus 8 m Entfernung mit einem Soundfield-Mikrofon der 1. Ordnung (Soundfield SPS 2000) realisiert. Die Signale wurden mit einem Multitrack Field Recorder (Zoom F8n Pro) aufgezeichnet. Anschließend wurde das Aufnahmемaterial kontrolliert. Aufgrund des hohen Diffusschallanteils wurde die Aufnahme als ungeeignet für die weitere Verarbeitung bewertet und die Idee wurde verworfen.

5.2 Beschallung und Aufnahme auf dem Campus Finkenau

Die Intention der Produktionsverfahren ist es, den Klang der Setzmaschine vor dem Zielgebäude - also vor dem Museum der Arbeit - zu reproduzieren. Aus praktischen Gründen wird die Simulation jedoch auf dem Campus der Hochschule stattfinden und nicht vor dem Museum der Arbeit. Gründe für diese Entscheidung sind der geringere logistische Aufwand zur Durchführung der Beschallung/Aufnahme und die potenziell höhere Verfügbarkeit von möglichen Versuchsteilnehmern für den Hörversuch. Der Ort ist geeignet, weil die Geometrie und die räumlichen Dimensionen innerhalb der Grenzen des verwendeten Raummodells liegen. So ist gewährleistet, dass beide Verfahren im Hörversuch miteinander verglichen werden können. Der Hörereignisort befindet sich auf dem Innenhof des Campus. Der Platz ist an 3 Seiten von Gebäuden umgeben. Im Grundriss grenzt der Platz oben an die Mensa und links an das Gebäude B mit dem Forum Finkenau. Nach unten wird das Areal durch das Gebäude D begrenzt (siehe Abb. 5). Die Orientierung des Hörers erfolgt in Richtung der Mensa, die in der Simulation das Zielgebäude darstellt. Der Bezugspunkt zwischen Realität und Simulation ist die Sitzbank vor der Mensa, die auch die fixe Hörerposition darstellt.

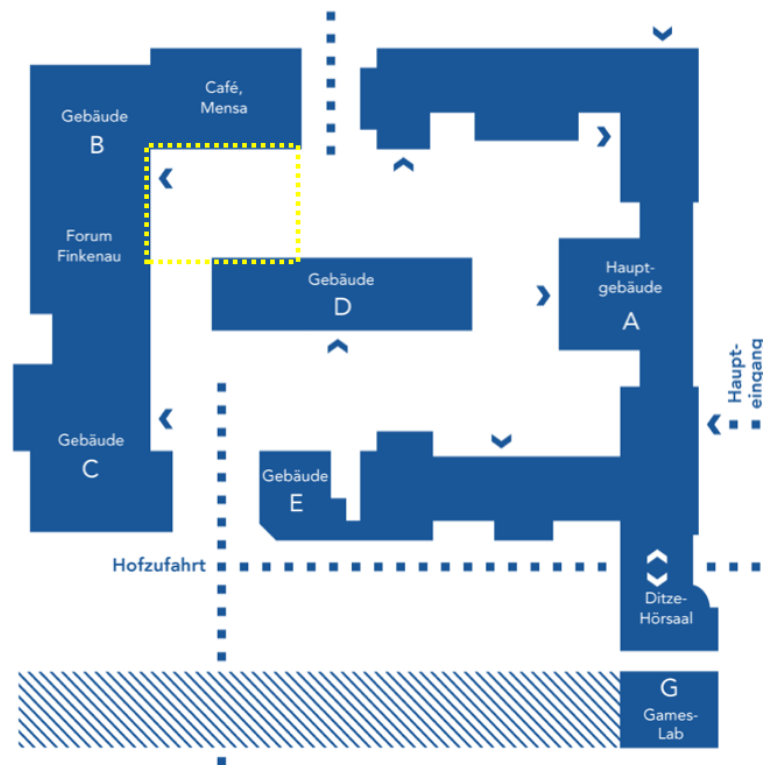


Abbildung 5: Campus Finkenau mit Hörereignisort (gelb gestrichelt)
 Quelle: <https://www.haw-hamburg.de/hochschule/design-medien-und-information/departments/medientechnik/unser-department/raum-lageplaene/#gallery-1>

Zunächst wurden Länge und Breite des rechteckigen Platzes vermessen. Die Koordinaten des Bezugspunkts (Hörerposition) und die 10 Lautsprecherpositionen wurden gemessen. Für die Beschallung wurde ein Studiomonitor (A6x von Adam) an einen PC mit Interface (RME Babyface) angeschlossen (siehe Abb. 6). Das Soundfield-Mikrofon (Soundfield SPS 2000) wurde an der Hörerposition aufgestellt. Die Positionen der Schallquelle wurden mit einem Multitrack Field Recorder (Zoom F8n Pro) aufgezeichnet. Dabei erfolgte die Orientierung des Monitorlautsprechers in Richtung des Mikrofons (parallel zum Vorgehen in der Simulation). Um beide Produktionsverfahren im Hörversuch miteinander vergleichen zu können, müssen die räumlichen Koordinaten der beiden Verfahren synchronisiert werden. So müssen die Standorte des Lautsprechers als Schallquelle mit den Positionen der virtuellen Schallquelle aus der Simulation übereinstimmen. Ebenfalls müssen die Position des Mikrofons und die virtuelle Hörerposition (Bezugspunkt) koinzident sein. Die Aufnahmen wurden abends gemacht, um den Einfluss von Umgebungsgeräuschen durch Publikums- und Straßenverkehr zu minimieren. Aufgrund des Wetters wurde ein Windschutz für das Soundfield-Mikrofon verwendet.



Abbildung 6: Aufbau der Beschallungsanlage
Quelle: Eigene Darstellung

5.3 Modellierung des simulierten Raums

Um den virtuellen Raum zu modellieren, wurden zuerst die Dimensionen des Hörereignisorts in das Modell übertragen. Als zweiter Schritt wurde die Anzahl der schallharten Wände bestimmt. Es musste im Vorfeld jedoch eine Vereinfachung vorgenommen werden, um die realen räumlichen Bedingungen an das Raummodell annähern zu können: Der Platz vor der Mensa ist zu zwei Seiten hin offen, betrachtet man den Raum im Grundriss „fehlt“ die rechte Seitenwand und es gibt eine 9,24 m breite Öffnung in der Rückwand (siehe Abb. 5). Bei einer Gesamtlänge der Rückwand von 30 m entspricht das etwa 33% der Fläche. In der Raumsimulation lassen sich durch Dämpfung allerdings nur komplette Wände entfernen, deswegen wird die Rückwand im Modell vereinfacht als geschlossen betrachtet. Der Platz ist nicht überdacht und zur rechten Seite hin nicht durch Wände begrenzt, deswegen wurden die Reflexionen von Decke und Seitenwand im Raummodell komplett bedämpft. Die verbleibenden angrenzende Wände und Flächen sind schallhart (Stein, Beton und Glas) und wurden daher nicht zusätzliche bedämpft.

Danach wurden die Positionen von Schallquelle und Hörer mit Hilfe von Polarkoordinaten in das Modell übertragen. Die virtuelle Hörerposition bildet dabei den korrespondierenden Bezugspunkt in der Simulation und befindet sich im Koordinatenursprung des Systems (Abb. 7 und 8). Mit dem Coordinate Converter wurden Winkel und Entfernung zwischen Schallquelle und Hörer festgelegt.

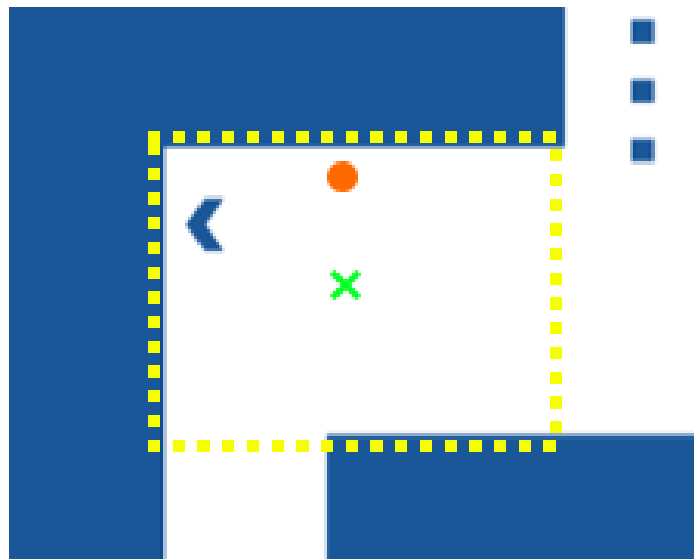


Abbildung 7: Hörereignisort mit Bezugspunkt (grün) und Schallquelle (orange)
Quelle: Eigene Darstellung

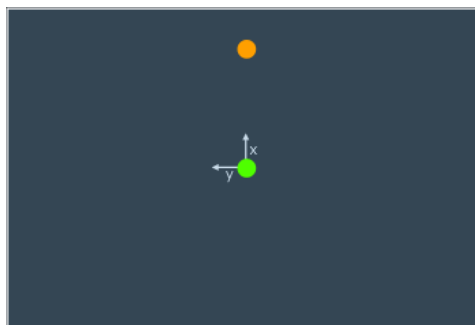


Abbildung 8: entsprechende Darstellung im Room Encoder
Quelle: Screenshot

Das Klangspektrums der Reflexionen wurde mit einem Filter angepasst. Die Schallwellen der hohen Frequenzen verlieren auf dem Weg durch Dissipation mehr Energie als die tiefen Frequenzen, dementsprechend wurden diese Frequenzen mit einem High-Shelve-Filter im Spektrum abgesenkt. Mit dem Directivity Shaper wurden alle Bänder gemeinsam in Richtung der Hörerposition im Room Encoder ausgerichtet. Die Richtung des Directivity-Signals korreliert mit dem Azimut zwischen Hörerposition und Schallquelle, sodass die meiste Signalenergie stets in Richtung des Hörers ausstrahlt. Für den Hörversuch wurde die virtuelle Schallquelle in der Raumsimulation (analog zu den Aufstellungen des Lautsprechers aus der Beschallungsaufnahme) im Abstand von 5 m und 8 m in verschiedenen Winkeln um die Hörerposition angeordnet.

5.4 Durchführung der Hörversuchs

Um die Vergleichbarkeit der 2 Produktionsverfahren im Hörversuch zu gewährleisten wurde identische Aufnahmetechnik für die Produktion verwendet. Die verwendeten Richtcharakteristiken sind gleich. So

wird sichergestellt, dass die beiden Verfahren miteinander verglichen werden und nicht die Mikrofone oder die Richtcharakteristiken. Die dargebotenen Testsignale für beide Verfahren haben denselben Ursprung. Sie stammen aus derselben Aufnahme und besitzen dieselbe technische Qualität (48 kHz/24 bit) und Länge. Die Pegel der Testsignale wurden nach der Signalverarbeitung einander angeglichen.

Zur Durchführung des Versuchs wurde ein Projekt in der DAW Reaper angelegt. Es wurden Audiosamples für jedes Verfahren angelegt, sodass 10 verschiedene Klangbeispiele pro Verfahren zur Verfügung stehen, die sich jeweils in Distanz und Winkel unterscheiden. Die Simulationsparameter für das virtuelle Raummodell wurden in der DAW automatisiert, um den Versuchsablauf effizient zu gestalten. Die Klangbeispiele wurden vor jedem Versuch randomisiert. Für den Hörversuch wurden 4 Klangbeispiele (2 pro Verfahren) ausgewählt und den Versuchsteilnehmern vorgespielt.

Der Hörversuch wurde auf dem Innenhof vor der Mensa durchgeführt. Der Aufbau bestand aus einem Laptop mit Interface (RME Ladyface) und einem Kopfhörer (Beyerdynamics DT 880 Pro), in dessen Bügel der Headtracker (Supperware Headtracker 1) befestigt wurde. Der Versuch wurde im Sitzen durchgeführt, um eine ortsstabile Hörerposition während des Versuchs zu gewährleisten. Die Versuchsteilnehmer wurden zunächst schriftlich und mündlich instruiert (siehe Anhang B). Vor Versuchsbeginn wurde der Kopfhörer auf richtigen Sitz kontrolliert und der Headtracker wurde kalibriert. Zu Beginn wurde den Versuchsteilnehmern ein Klangbeispiel aus jeder Kategorie vorgestellt. Dann wurde der 1. Stimulus dargeboten und die Probanden sollten nacheinander alle Fragen zu Ihrer Wahrnehmung beantworten. Auf Nachfrage wurde der entsprechende Stimulus wiederholt. Anschließend wurde das nächste Klangbeispiel präsentiert. An dem Hörversuch nahmen 6 Probanden teil.

6 Ergebnisse

In diesem Abschnitt werden die Ergebnisse aus dem Hörversuch anhand von Zahlenwerten und Diagrammen präsentiert. Die Suggestiv-Lokalisierung dient als Prüfparameter für die Externalisierung der virtuellen Schallquelle. Die Schallquelle wurden von allen Probanden externalisiert wahrgenommen. Die Ergebnisse aus dem Fragebogen wurden in Zahlenwerte überführt und in Excel-Tabellen (siehe Anhang C und D) übertragen. Die Daten wurden separat für jedes Verfahren aufbereitet. Es wurde die Häufigkeit der Abweichungen von der korrekten Lokalisierung und deren prozentuale Verteilung pro Verfahren dargestellt. Zusätzlich wurden die Abweichungen berechnet und für beide Verfahren in einem Boxplot visualisiert. Für die Distanzwahrnehmung wurde die Streuung um die Distanzen (5 m und 8 m) für jedes Verfahren dargestellt. Die Auswertung der Skalen des semantischen Differentials erfolgt anhand des in Tabelle 3 zugrunde gelegten Maßstabs.

6.1 Lokalisierung

Tabelle 5: Abweichungen von der tatsächlichen Position in °

Verfahren A	Verfahren B
-22,5	-90
-22,5	-22,5
-22,5	-22,5
0	-22,5
0	0
0	0
0	0
0	0
0	0
0	0
90	0
180	90

Tabelle 6: Streuungsmaße in °

Streuungsmaße	Verfahren A	Verfahren B
Min	-22,5	-90
Q1	-16,88	-16,88
Q2 (Median)	0	0
Q3	0	0
Max	180	90
IQA	16,88	16,88

In Tabelle 5 sind die Abweichungen in Grad von der tatsächlichen Position aufgeführt. Der Wert 0 entspricht einer Abweichung von 0° was bedeutet, dass richtig lokalisiert wurde. Die Vorzeichen beschreiben die Richtung. Ein negatives Vorzeichen steht für eine Abweichung nach rechts, ein positives Vorzeichen für eine Abweichung nach links. Die Streuungsmaße sind in Tabelle 6 dargestellt. Die Extremwerte sind nicht in die Berechnung der Quartile eingegangen. Aufgrund der Datenmenge wird die Verteilung der Abweichungen in einem Boxplot für beide Verfahren zusammengefasst dargestellt.

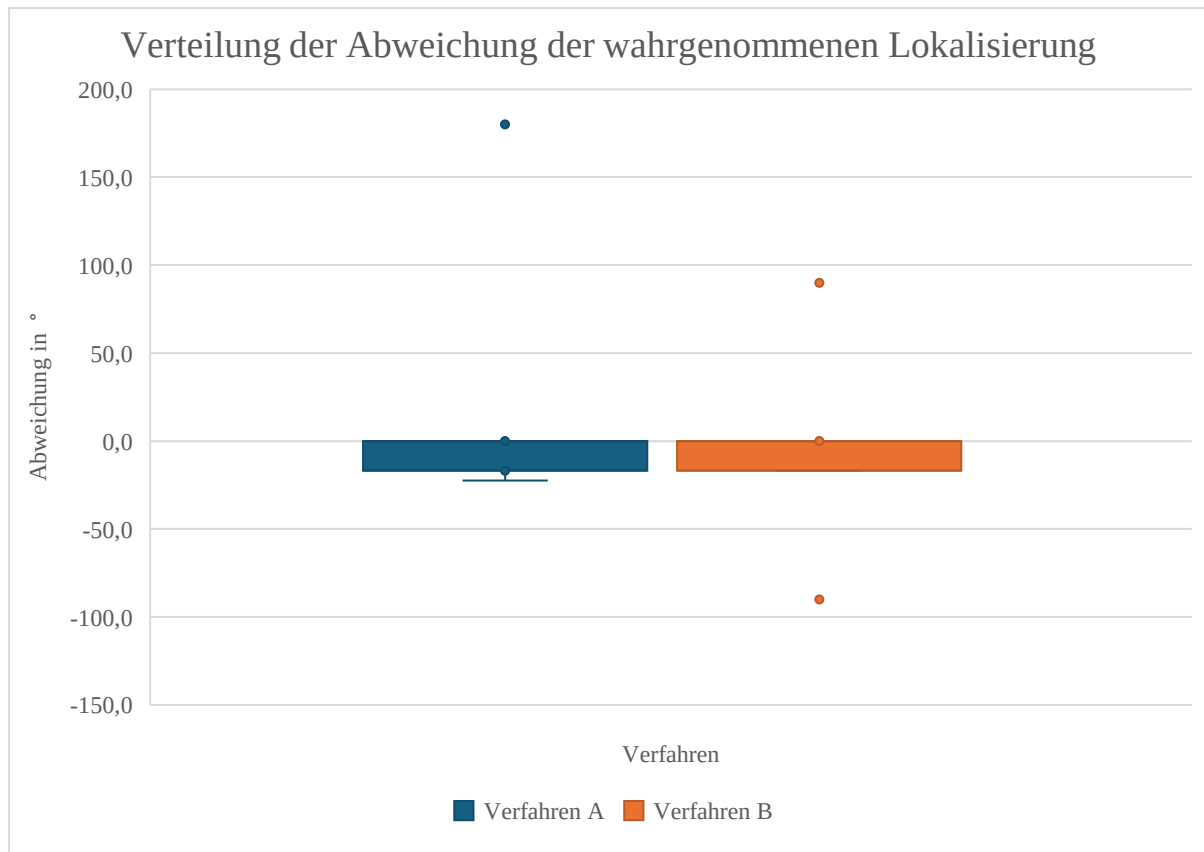


Abbildung 9: Verteilung der Abweichung der wahrgenommenen Lokalisierung

Bei der Betrachtung des Diagramms stechen zuerst die Ausreißerpunkte hervor. Bei Verfahren A sind diese stärker ausgeprägt. Die Punkte bei $\pm 90^\circ$ und 180° sind extreme Ausreißer, die sich deutlich oberhalb des 1,5-fachen IQA (Interquartilabstand) befinden. Der Minimalwert liegt bei $-22,5^\circ$, der Maximalwert bei 180° . Die mittleren 50% der berechneten Abweichungen liegen zwischen $-16,88^\circ$ und 0° , der IQA ist $16,88^\circ$ groß. Der Punkt $-22,5^\circ$ bildet die untere Grenze und ist in die Antenne eingefasst. Bei Verfahren B beträgt der IQA ebenfalls $16,88^\circ$. Der Minimalwert liegt bei $-16,88^\circ$, der Maximalwert bei 90° . Lässt man die Ausreißer bei der Betrachtung außen vor, weisen die Daten von Verfahren A die gleiche Streuung auf wie die Daten von Verfahren B. Der Median ist bei beiden Verfahren 0° .

Prozentuale Verteilung

Die Häufigkeit der Abweichungen und deren prozentuale Verteilung ist für jedes Verfahren separat in einem Kreisdiagramm visualisiert (Abb. 10 und 11). Es ist dargestellt, wie oft richtig lokalisiert wurde, was einer Abweichung von 0° entspricht und wie häufig kleinere und größere Abweichungen (bis zu 180°) von der korrekten Position auftraten. Es werden nur die absoluten Werte der Abweichungen betrachtet, die Richtung wird nicht unterschieden.

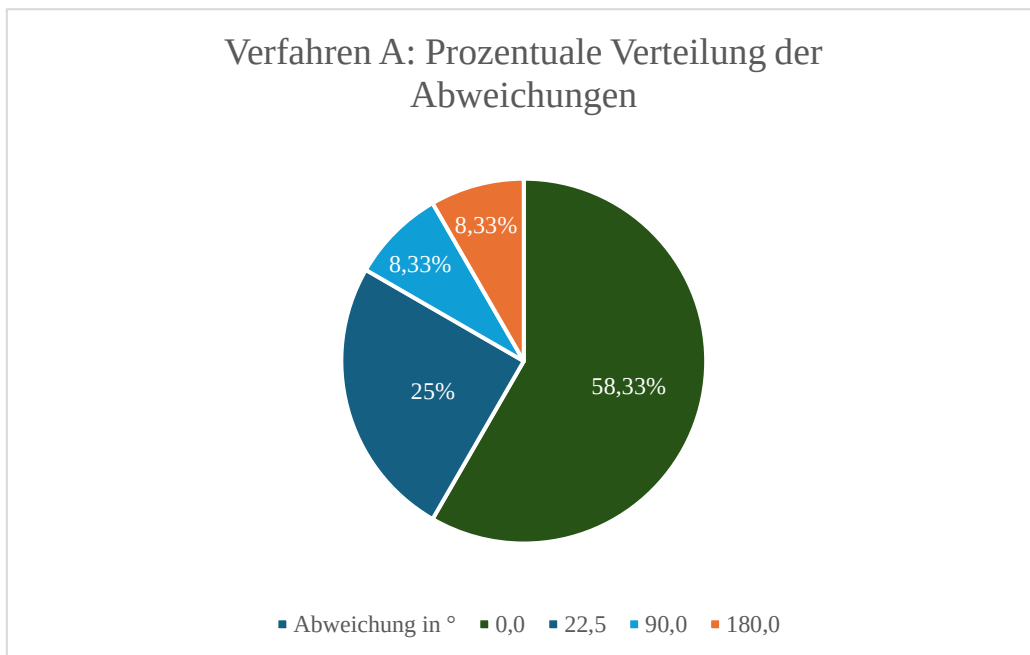


Abbildung 10: Verfahren A: Prozentuale Verteilung der Abweichungen

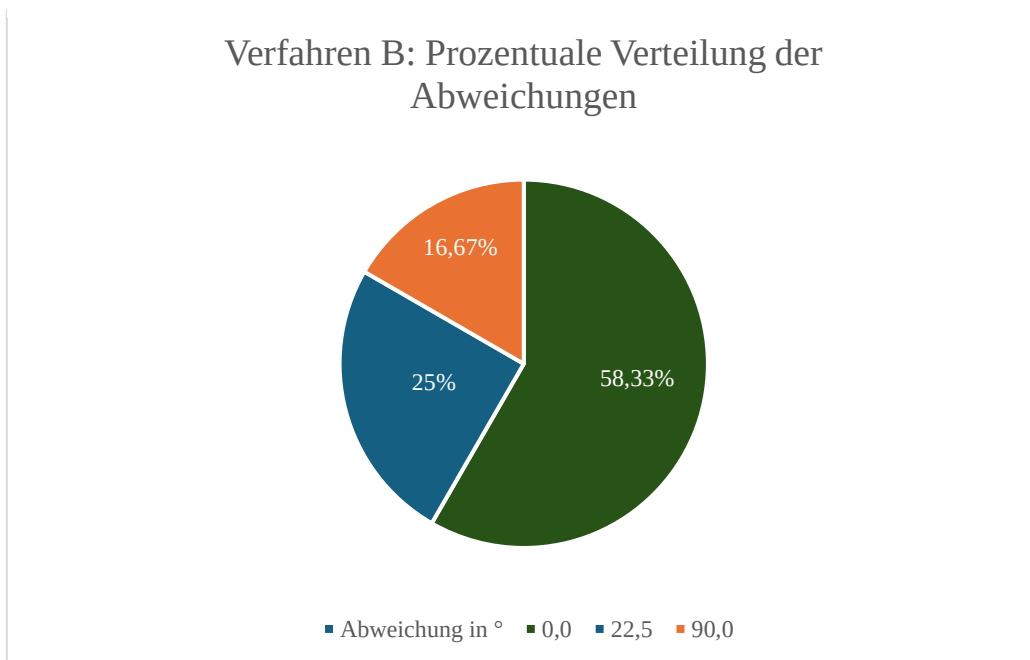


Abbildung 11: Verfahren B: Prozentuale Verteilung der Abweichungen

Rund 58,33% der Probanden hat die Positionen der Schallquellen in beiden Verfahren richtig lokalisiert. In 25 % der Fälle trat eine Abweichung von 22,5° von der korrekten Position auf. 8,33% der Lokalisierungen bei Verfahren A wichen um 90° und 8,33% um 180° ab. Bei Verfahren B trat bei 16,67% eine Abweichung von 90° auf. Ein Proband lokalisierte alle Positionen um $\pm 90^\circ$ gedreht. Derselbe Proband lokalisierte die 0° Position bei 180°, was einer front-back confusion entspricht.

6.2 Distanzwahrnehmung

Für die Distanzwahrnehmung steht eine größere Anzahl an Werten für die jeweilige Distanz zur Verfügung, daher wird die Streuung um beide Distanzen (5 m und 8 m) separat für jedes Verfahren in einem Boxplot visualisiert. In Tabelle 7 sind die wahrgenommenen Distanzen für Verfahren A aufgelistet. Tabelle 8 beinhaltet die Streuungsmaße aus denen das Boxplot erstellt wurde. Die Werte für Verfahren B sind analog dazu in Tabelle 9 und 10 aufgeführt.

Tabelle 7: Verfahren A: Distanzen in m

Verfahren A: Distanzwahrnehmung	
5 m	8 m
8	8
8	8
5	8
5	8
8	8
5	8

Tabelle 8: Verfahren A: Streuungsmaße in m

Streuungsmaße	5 m	8 m
Min	5	8
Q1	5	8
Q2 (Median)	6,5	8
Q3	8	8
Max	8	8
IQA	3	0

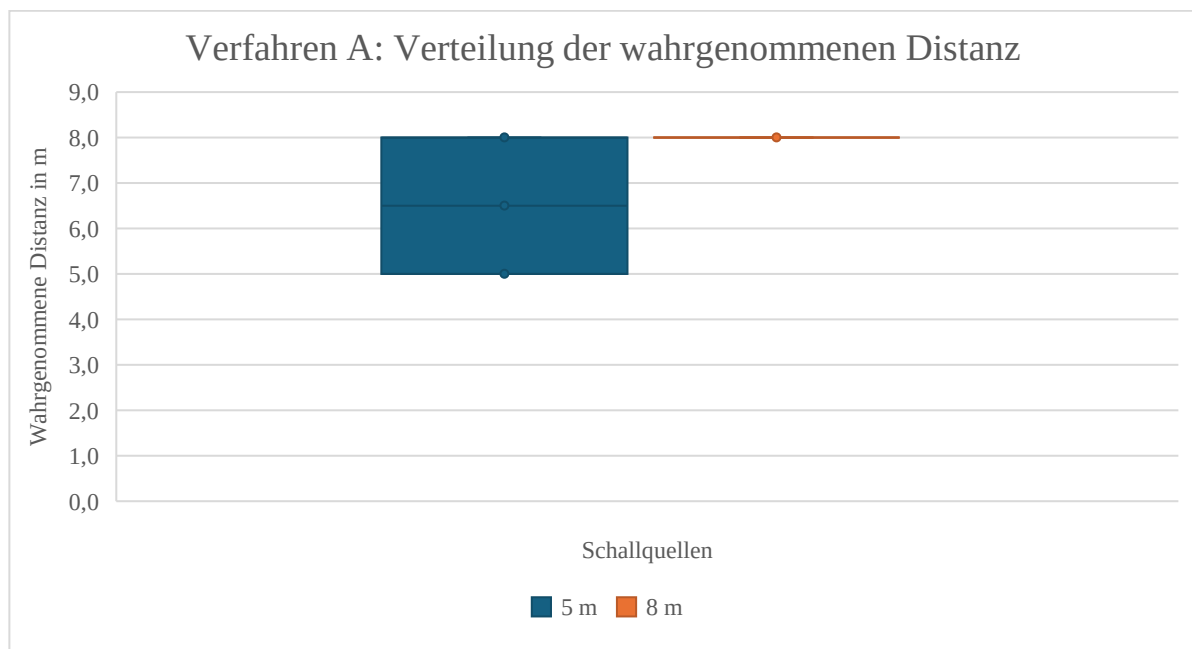


Abbildung 12: Verfahren A: Verteilung der wahrgenommenen Distanz

Die Werte für die Distanz von 5 m sind zwischen 5 m und 8 m verteilt. Das Minimum liegt bei 5 m und bildet gleichzeitig das 1. Quartil, das Maximum mit 8 m bildet gleichzeitig das 3. Quartil. In dem IQA befinden sich die mittleren 100% der Werte. Es gibt keine Ausreißer. Bei der Distanz von 8 m haben die Grenzwerte sowie alle 3 berechneten Quartile den Wert 8 m. Der IQA ist 0 m, weil alle Daten präzise bei 8 m liegen.

Tabelle 9: Verfahren B: Distanzen in m

Verfahren B: Distanzwahrnehmung	
5 m	8 m
8	5
1	5
5	5
1	8
5	8
5	8

Tabelle 10: Verfahren B: Streuungsmaße in m

Streuungsmaße	5 m	8 m
Min	1	5
Q1	2	5
Q2 (Median)	5	6,5
Q3	5	8
Max	8	8
IQA	3	3

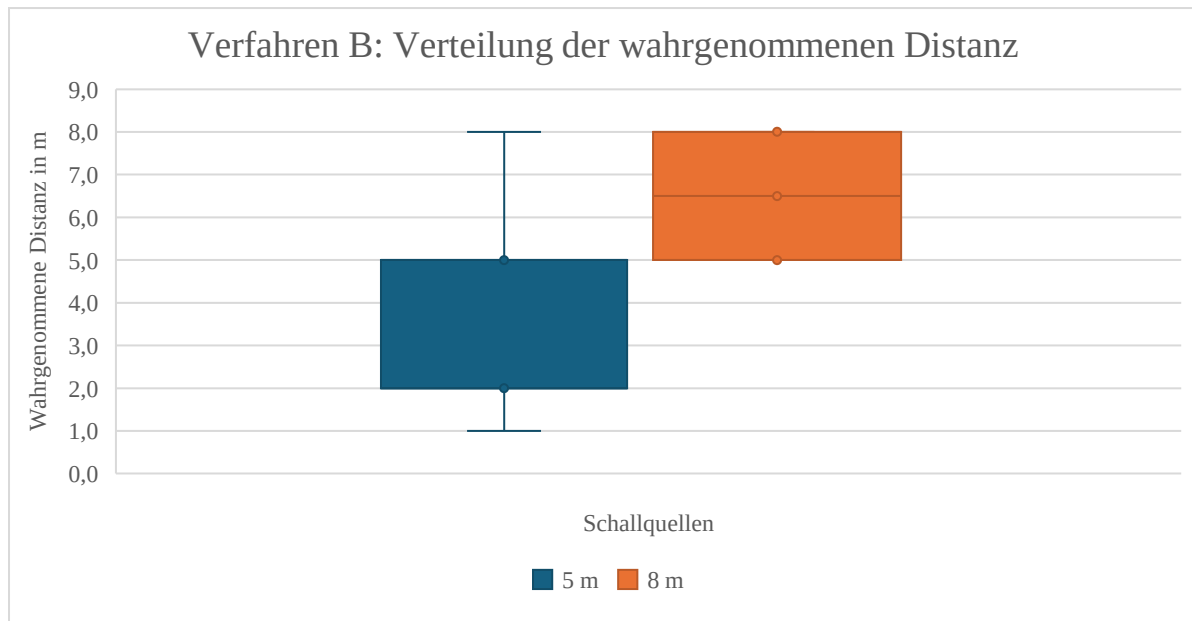


Abbildung 13: Verfahren B: Verteilung der wahrgenommenen Distanz

Die Werte für die Distanz von 5 m sind breiter gestreut, sie liegen zwischen 1 m und 8 m. Diese Werte sind in den Antennen eingefasst. Die mittleren 50% der Werte befinden sich zwischen 2 m und 5 m, der IQA ist 3 m. Es gibt keine Ausreißer. Die Werte für die Distanz von 8 m sind zwischen 5 m und 8 m verteilt. Das Minimum mit 5 m bildet gleichzeitig das 3. Quartil, das Maximum mit 8 m bildet das 1. Quartil. Der IQA hat den Wert 3 m, er umschließt 100% der Werte.

6.3 Ausdehnung, Stabilität und Plausibilität

Die Häufigkeit der abgegebenen Bewertungen zu Ausdehnung, Stabilität und Plausibilität für Verfahren A ist in Tabelle 11 zusammengefasst. Die Daten werden in einem Säulendiagramm visualisiert.

Tabelle 11: Verfahren A: Häufigkeit der Bewertungen

Verfahren A			
Bewertung	Ausdehnung	Stabilität	Plausibilität
1	1	2	2
2	5	5	5
3	1	0	1
4	3	4	2
5	2	1	2

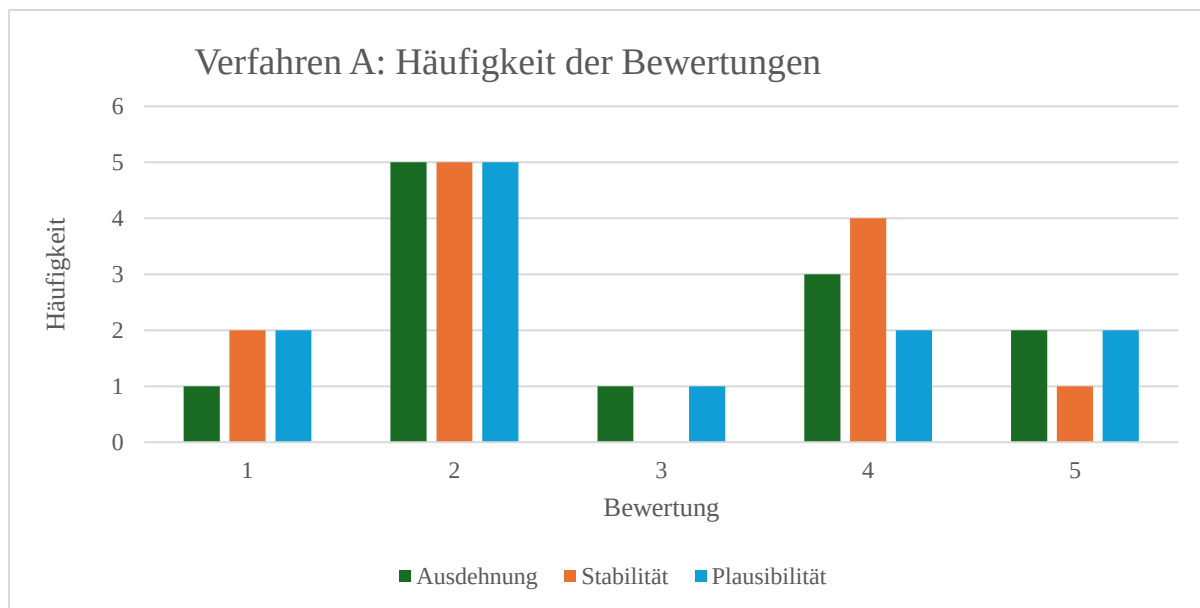


Abbildung 14: Verfahren A: Häufigkeit der Bewertungen

Zuerst fällt auf, dass alle 3 Parameter am häufigsten mit 2 (gering) bewertet wurden, am zweithäufigsten mit 4 (hoch). Die Mitte mit der Bewertung 3 (mittel) weist die geringste Häufung auf. Auffällig sind die abfallenden Zahlen zu den Seiten hin, die Bewertung 1 und 5 (sehr gering und sehr hoch) wurde jeweils 5-mal abgegeben.

Ausdehnung: Der Immersionsgrad der Ausdehnung wurde überwiegend als gering bewertet, wobei die Unterschiede in der Häufigkeit zwischen positiven und negativen Bewertungen gering ausfallen (6-mal Bewertung 1 und 2, 5-mal Bewertung 4 und 5).

Stabilität: Die Bewertungen der Stabilität zeigen eine leichte Tendenz zu niedrigen Bewertungen (7-mal Bewertung 1 und 2 gegenüber 5-mal Bewertung 4 und 5).

Plausibilität: Bei der Plausibilität ist die Tendenz zu niedrigen Bewertungen am stärksten ausgeprägt (7-mal Bewertung 1 und 2 gegenüber 4-mal Bewertung 4 und 5).

Die Häufigkeit der abgegebenen Bewertungen zu Ausdehnung, Stabilität und Plausibilität für Verfahren B ist in Tabelle 12 zusammengefasst.

Tabelle 12: Verfahren B: Häufigkeit der Bewertungen

Verfahren B			
Bewertung	Ausdehnung	Stabilität	Plausibilität
1	0	1	3
2	5	3	1
3	4	2	3
4	2	5	3
5	1	1	2

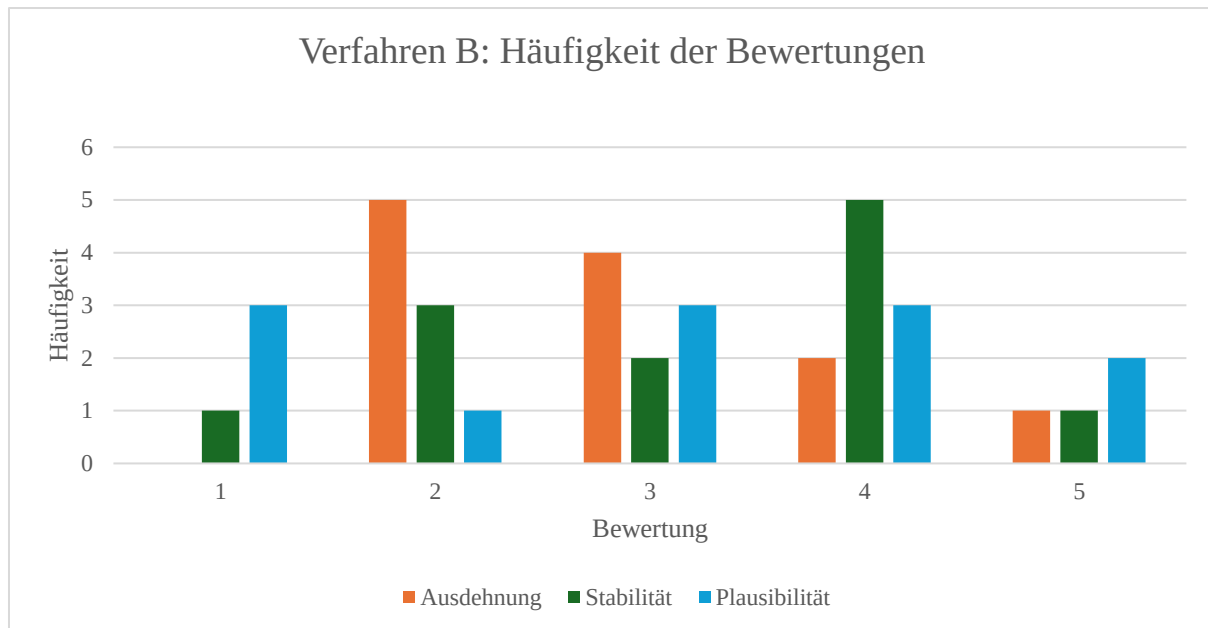


Abbildung 15: Verfahren B: Häufigkeit der Bewertungen

Die Stabilität wurde am häufigsten mit 4 (hoch) bewertet, die Ausdehnung am häufigsten mit 2 (gering). Die Plausibilität wurde jeweils 3-mal mit 1 (sehr gering), 3 (mittel) und 4 (hoch) bewertet. Insgesamt konzentrieren sich die abgegebenen Bewertungen um die Mitte herum. Betrachtet man alle 3 Parameter wurden die Bewertungen 2, 3 und 4 gleich häufig (9-mal) abgegeben. Auffällig sind auch hier die abfallenden Zahlen zu den Seiten hin, dort wurde jeweils 4-mal die Bewertung 1 und 5 abgegeben.

Ausdehnung: Die Verteilung der Ausdehnung zeigt eine klare Tendenz zu den niedrigen Bewertungen. Der Immersionsgrad wurde 5-mal als gering, 4-mal als mittel und 2-mal als hoch bewertet, die Bewertung 5 (sehr hoch) erhielt eine Stimme.

Stabilität: Die Stabilität erhielt überwiegend höhere Bewertungen, der Immersionsgrad wurde 6-mal überwiegend hoch bewertet (4 und 5), 2-mal mittel und 4-mal eher negativ (1 und 2). Die Daten zeigen eine klare Tendenz zu den höheren Bewertungen.

Plausibilität: Die Plausibilität ist überwiegend höher bewertet, wobei der Unterschied mit einem Punkt zwischen positiven und negativen Bewertungen gering ist. Insgesamt sind die Daten gleichmäßiger verteilt als die Daten von Ausdehnung und Stabilität.

7 Auswertung

Die verschiedenen Testarten werden zunächst getrennt ausgewertet. Die Daten von Ausdehnung, Stabilität und Plausibilität werden auf eine Prozentbasis skaliert, um die Ergebnisse der unterschiedlichen Testarten zu kategorisieren. Anschließend werden die Testergebnisse zusammengefasst und nach der Relevanz der Parameter gewichtet. Schließlich wird aus den Daten der gewichtete Durchschnitt für jedes Verfahren berechnet.

7.1 Lokalisierung

Betrachtet man die Verteilung der Abweichungen der wahrgenommenen Lokalisierung, weisen die Daten von beiden Verfahren im Boxplot die gleiche Streuung auf. Die mittleren 50% der berechneten Abweichungen liegen zwischen $-16,875^\circ$ und 0° . Der IQA ist $16,875^\circ$ groß, was auf eine gewisse Variabilität der Daten hinweist. Der Median beträgt bei beiden Verfahren 0° , was bedeutet, dass die zentrale Tendenz der Verteilung bei 0° liegt. Bei der grafischen Darstellung der prozentualen Verteilung der Abweichungen gibt es keinen Unterschied zwischen beiden Verfahren. Die überwiegende Mehrheit der Versuchsteilnehmer (58,33%) hat die Positionen der Schallquelle in beiden Verfahren richtig lokalisiert.

Die Extremwerte (Abweichungen von über 90° von der tatsächlichen Position) sind durch einen Probanden zu erklären, der in beiden Verfahren gleichermaßen fehlerhaft lokalisierte. Er lokalisierte alle Azimute um $\pm 90^\circ$ in die jeweilige Richtung verschoben. Beispielsweise lokalisierte er $22,5^\circ$ bei $112,5^\circ$ und $-22,5^\circ$ bei $-112,5^\circ$. Bei diesem Probanden trat auch eine Vorne-Hinten-Verwechslung auf, sodass er eine frontale Schallquelle bei 180° verortete. Dies deutet darauf hin, dass die bei der binauralen Dekodierung verwendeten generischen HRTFs für den Probanden nicht passend sind, weil seine Kopfanatomie zu stark von der des Kunstkopfes abweicht.

Die Stabilität der Schallquelle erweist sich im Hörversuch als einer der bedeutendsten Parameter für die Immersion. Eine gering wahrgenommene Stabilität wirkt sich vor allem negativ auf die Lokalisierung aus, da die Position der Schallquelle sich entgegen der Kopfdrehung verschiebt, wobei die Drehung des Kopfes mit dem Bewegungsradius der Schallquelle korreliert. Dadurch wird eine präzise Lokalisierung der Schallquelle erschwert. Dieses Phänomen wurde von einigen Versuchsteilnehmern stärker wahrgenommen als von anderen und bietet einen möglichen Anhaltspunkt für die Erklärung der identischen Ergebnisse bei der Lokalisierung. Die Stabilität der Schallquelle entsteht aus dem Zusammenspiel von Headtracker und Scene Rotator am Ende der Gerätekette und wirkt sich auf beide Produktionsverfahren gleichermaßen aus. Die Stabilität der Schallquelle hat ebenso Einfluss auf die Wahrnehmung der Ausdehnung. Werden alle 3 Parameter (Stabilität, Lokalisierung und Ausdehnung) als wenig immersiv wahrgenommen und bewertet, wirkt sich das insgesamt negativ auf die Bewertung der Plausibilität aus. Die Daten zeigen, dass die Distanzwahrnehmung nicht von der Stabilität der Schallquelle beeinflusst wird.

7.2 Distanzwahrnehmung

Verfahren A: 50% der Werte für die Distanz von 5 m wurden richtig bestimmt. Die Werte für die Distanz von 8 m wurden zu 100% richtig bestimmt. Der IQA von 0° weist auf eine hohe Präzision hin. Im Mittel wurden 75% aller Distanzen korrekt erkannt.

Verfahren B: Die Zahlen für die Distanz von 5 m sind breiter verteilt als die Zahlen bei Verfahren A, trotzdem wurden 50% der Distanzen richtig erkannt. Für die Distanz von 8 m wurden 50% der Werte richtig bestimmt. Bei Verfahren B wurden insgesamt 50% aller Distanzen korrekt bestimmt.

7.3 Ausdehnung, Stabilität und Plausibilität

Der Grad der Immersion bei Verfahren A wurde überwiegend als gering eingestuft (20-mal Bewertung 1 und 2 gegenüber 14-mal Bewertung 4 und 5). Summiert man die Häufigkeiten aller Bewertungen auf, ergibt sich eine Gesamtzahl von 102 Punkten für Verfahren A. Dies entspricht einem geringen Immersionsgrad. Dieser Wert wird in eine Prozentzahl umgerechnet, um ihn mit den anderen Testergebnissen zusammenzufassen zu können. Durch Skalierung ergibt sich eine Prozentzahl von 56,67% für Verfahren A.

Der Grad der Immersion bei Verfahren B wurde insgesamt überwiegend höher bewertet als bei Verfahren A (14-mal Bewertung 4 und 5 gegenüber 13-mal Bewertung 1 und 2). Summiert man die Häufigkeiten aller Bewertungen, ergibt sich eine Gesamtzahl von 109 Punkten, was einem mittleren Immersionsgrad entspricht. Durch Skalierung ergibt sich eine Prozentzahl von 60,56% für Verfahren B.

7.4 Übersicht aller Testergebnisse

In Tabelle 13 sind die Ergebnisse der verschiedenen Testarten für beide Produktionsverfahren in der Übersicht zusammengefasst.

Tabelle 13: Übersicht der Testergebnisse

	Verfahren A	Verfahren B
Korrekte Lokalisierung	58,33%	58,33%
Korrekte Distanzwahrnehmung	75,00%	50,00%
Bewertung von Ausdehnung, Stabilität und Plausibilität	56,67%	60,56%

Die Testergebnisse zeigen, dass 58,33% der Probanden die Schallquelle in beiden Verfahren richtig lokalisiert haben. Den größten Unterschied zwischen den Verfahren weisen die Ergebnisse der Distanzwahrnehmung auf. Bei Verfahren A beträgt die Übereinstimmung von wahrgenommener und tatsächlicher Distanz 75%. Bei Verfahren B wurden 50% der Entfernungen durch die Versuchsteilnehmer richtig eingeschätzt. Dies zeigt, dass Verfahren A eine höhere Genauigkeit bei der Distanzwahrnehmung bietet. Die Bewertung von Ausdehnung, Stabilität und Plausibilität zeigt eine leichte Präferenz für Verfahren B. Hier wurde der Grad der Immersion mit 60,56% etwas höher bewertet als bei Verfahren A mit 56,67%. Der Unterschied zwischen den Werten ist jedoch marginal, was auf eine ähnliche Wahrnehmung der Immersion in beiden Verfahren hinweist.

7.5 Gewichteter Durchschnitt

Abschließend werden alle Testarten zusammengefasst, indem der gewichtete Durchschnitt aus allen Testergebnissen für das jeweilige Produktionsverfahren berechnet wird. So kann eine Aussage über den Grad der Immersion der Produktionsverfahren getroffen werden. Aufgrund der besonderen Relevanz der Lokalisierung und der Distanzwahrnehmung werden die Ergebnisse dieser beiden Parameter mit einem Faktor von 1,5 gewichtet. Die Ergebnisse und Gewichte sowie die daraus berechneten Werte für Verfahren A sind in Tabelle 14 aufgeführt. Tabelle 15 enthält analog dazu die Werte für Verfahren B.

Tabelle 14: Verfahren A: Testergebnisse und Gewichte

	Verfahren A	Gewichtung	Gewichteter Wert
Korrekte Lokalisierung	58,33%	1,5	87,50%
Korrekte Distanzwahrnehmung	75,00%	1,5	112,50%
Bewertung von Ausdehnung, Stabilität und Plausibilität	56,67%	1	56,67%
Summe		4	256,66%
Gewichteter Durchschnitt 64,17%			

Tabelle 15: Verfahren B: Testergebnisse und Gewichte

	Verfahren B	Gewichtung	Gewichteter Wert
Korrekte Lokalisierung	58,33%	1,5	87,50%
Korrekte Distanzwahrnehmung	50,00%	1,5	75,00%
Bewertung von Ausdehnung, Stabilität und Plausibilität	60,56%	1	60,56%
Summe		4	223,05%
Gewichteter Durchschnitt 55,76%			

Der gewichtete Durchschnitt bildet das Gesamtergebnis aller untersuchten Parameter ab. Bei der Gegenüberstellung beider Verfahren zeichnet sich hier eine klare Präferenz für Verfahren A ab. Verfahren B erzielt mit 55,76% einen **mittleren Immersionsgrad**. Verfahren A weist mit 64,17% ebenso einen **mittleren Immersionsgrad** auf.

8 Fehleranalyse

Zukünftig sollte das Design des Hörversuchs so angepasst werden, dass weniger Richtungen getestet werden und dafür eine größere Anzahl an Klangbeispielen pro Azimut präsentiert werden. Dies kann die Anzahl der Werte für die Streuung der Lokalisierung erhöhen und die Aussagekraft der Boxplot-Diagramme verbessern. Statt einer 5-stufigen Skala bei der Verwendung des semantischen Differentials sollte man eine 4- oder 6-stufige Skala verwenden. Die Verwendung einer geradzahligen Skala zwingt die Teilnehmer dazu, sich für eine Ausprägung zu entscheiden und verhindert, dass der Mittelpunkt der Skala ausgewählt wird, was keine weiteren Informationen hinsichtlich einer Präferenz liefert. Bei einem Probanden traten Fehllokalisationen auf, die vermutlich auf den Einsatz generischer HRTFs bei der binauralen Dekodierung zurückzuführen sind. Bei einem erneuten Hörversuch sollte man individuelle physiologische Merkmale wie z.B. den Kopfumfang und den interauralen Abstand messen, um einen besser geeigneten, für den Probanden passenden HRTF-Datensatz aus einer Datenbank auszuwählen. So kann die räumliche Wahrnehmung verbessert und die Präzision von Lokalisierung und Distanzwahrnehmung erhöht werden, was wiederum den Grad der Immersion verbessert.

9 Zusammenfassung

In dieser Arbeit wurden 2 Produktionsverfahren entwickelt, die dazu in der Lage sind, eine reale Schallquelle aus einem Innenraum draußen vor einem Gebäude virtuell zu reproduzieren, mit dem Ziel einen hohen Grad an Immersion zu erreichen. Dazu wurden 2 unterschiedliche technische Ansätze getestet, miteinander verglichen und deren Grad an Immersion in einem Hörversuch verifiziert. Es wurden die Richtigkeit und Präzision von Lokalisierung und Distanzwahrnehmung analysiert. Zudem erfolgte eine Bewertung der Ausdehnung, Stabilität und Plausibilität durch die Versuchsteilnehmer.

Bei der Auswertung der Lokalisierung sind die Ergebnisse beider Verfahren identisch. Rund 60% der Probanden hat die Positionen der Schallquelle korrekt lokalisiert. Extreme Abweichungen bei der Lokalisierung deuten darauf hin, dass die bei der Binauralen Dekodierung verwendeten generischen HRTFs für einen Versuchsteilnehmer unpassend waren. Bei der Einschätzung der Distanzen zur Schallquelle weist Verfahren A eine höhere Präzision auf. Hier wurden im Mittel 75% aller Distanzen korrekt erkannt. Bei Verfahren B wurden 50% aller Distanzen richtig bestimmt. Die Auswertung von Ausdehnung, Stabilität und Plausibilität zeigt eine leichte Präferenz für Verfahren B. Hier ergibt sich eine Gesamtzahl von 109 Punkten, was einem mittleren Immersionsgrad entspricht. Für Verfahren A wurden 102 Punkte ermittelt. Dies entspricht einem geringen Immersionsgrad. Vergleicht man den gewichteten Durchschnitt aller untersuchten Parameter, weist Verfahren A insgesamt einen höheren Grad an Immersion auf als Verfahren B. Verfahren B erreicht eine Prozentzahl von 55,76%. Bei Verfahren A wurden 64,17% erzielt. Damit wurde in beide Produktionsverfahren ein **mittlerer Immersionsgrad** erreicht.

Im Rahmen des Hörversuchs konnten zudem einige bedeutsame Erkenntnisse gewonnen werden. Die Suggestiv-Lokalisierung eignet sich als Prüfparameter um festzustellen, ob der Versuch auswertbar ist. Die Stabilität ist von herausragender Bedeutung für das Gesamtsystem und wirkt sich vor allem auf die wahrgenommene Lokalisierung aus, hat aber auch Einfluss auf die empfundene Ausdehnung der Schallquelle. Die Wahrnehmung von Stabilität, Lokalisierung und Ausdehnung wirkt sich wiederum auf die Bewertung der Plausibilität aus und ist damit maßgeblich für den Immersionsgrad. Die Ergebnisse der Distanzwahrnehmung deuten darauf hin, dass keine Abhängigkeit von der Stabilität besteht.

Vergleicht man die benötigten Ressourcen für beide Verfahren, erweist sich Verfahren A als dasjenige mit dem geringsten finanziellen und zeitlichen Aufwand. Die Variabilität der Simulationssoftware ist ein großer Vorteil, allerdings hat das verwendete Raummodell Grenzen. Verfahren B benötigt dagegen die meisten Ressourcen, lässt sich dafür aber grundsätzlich an jedem städtischen Gebäude anwenden. Jedes der Produktionsverfahren hat seine Berechtigung für den spezifischen Anwendungsfall. Es obliegt dem Anwender, das geeignete Verfahren seinen Anforderungen entsprechend auszuwählen.

Literaturverzeichnis

- Armstrong, C., & Kearney, G. (2022). Ambisonics understood. In Paterson, J., & Lee, H. (Hrsg.). *3D Audio*. (S. 99-129). Routledge.
- Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355–385. <https://doi.org/10.1162/pres.1997.6.4.355>
- Billinghurst, M., Clark, A., & Lee, G. (2015). A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2–3), 73–272. <https://doi.org/10.1561/11000000049>
- Corrêa De Almeida, G., Costa De Souza, V., Da Silveira Júnior, L. G., & Veronez, M. R. (2023). Spatial Audio in Virtual Reality: A systematic review. *Symposium on Virtual and Augmented Reality*, 264–268. <https://doi.org/10.1145/3625008.3625042>
- Die Monotype (o. D.). Stiftung Historische Museen Hamburg. Abgerufen am 3. März 2025, von <https://www.shmh.de/veranstaltungen/die-monotype/619/20250113-1400/>
- Frank, M., & Zotter, F. (2022). Ambisonics. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (S. 1–20). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-60357-4_25-1
- Jakka, J., Tikander, M., Karjalainen, M., & Lokki, T. (2004). *Augmented Reality Audio for Mobile and Wearable Appliances*.
- Joshi, S., Stavrianakis, K., & Das, S. (2020). Substituting Restorative Benefits of Being Outdoors through Interactive Augmented Spatial Soundscapes. *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–4. <https://doi.org/10.1145/3373625.3418029>
- McKeag, A., & McGrath, D. S. (1996). Sound field format to binaural decoder with head tracking. *Audio engineering society convention 6r*. <https://www.aes.org/e-lib/browse.cfm?elib=7477>
- Nicol, R., Gros, L., Colomes, C., Warusfel, O., Noisternig, M., Bahu, H., Katz, B. F., & Simon, L. S. R. (o. J.). *A roadmap for assessing the Quality of Experience of 3D Audio binaural rendering*.
- Rudrich, D. (o. D.). *Plug-in Descriptions*. Plugins.iem. Abgerufen am 13. März 2025, von <https://plugins.iem.at/docs/pluginDescriptions/>

- Yang, J., Barde, A., & Billinghamurst, M. (2022). Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions. *Journal of the Audio Engineering Society*, 70(10), 788–809. <https://doi.org/10.17743/jaes.2022.0048>
- Zhang, M., Guan, T., Chen, L., Fu, T., Su, D., & Qu, T. (2021). *Individualized HRTF-based Binaural Renderer for Higher-Order Ambisonics*.

Anhang

Anhang A: Fragebogen Hörversuch

Lokalisierung:

Wo befindet sich die Schallquelle? Zeigen Sie mit dem Stick in die Richtung, in der Sie die Position der Schallquelle wahrgenommen haben.

Distanz:

In welcher Entfernung vermuten Sie die Schallquelle?

- a. Kurz vor dem Fenster der Mensa
- b. Auf halber Strecke zwischen meiner Position und der Mensa
- c. Kurz vor meiner Position

1. Klangbeispiel	
2. Klangbeispiel	
3. Klangbeispiel	
4. Klangbeispiel	

Ausdehnung:

Wie groß empfinden Sie die Ausdehnung der Schallquelle?

1. Klangbeispiel

punktförmig/ begrenzt

--	--	--	--	--

 unscharf/ verwaschen

2. Klangbeispiel

punktförmig/ begrenzt

--	--	--	--	--

 unscharf/ verwaschen

3. Klangbeispiel

punktförmig/ begrenzt

--	--	--	--	--

 unscharf/ verwaschen

4. Klangbeispiel

punktförmig/ begrenzt

--	--	--	--	--

 unscharf/ verwaschen

Stabilität:

Wie würden Sie die Stabilität der Schallquelle beurteilen?

Hat sich die Position der Schallquelle beim Hören verändert?

1. Klangbeispiel

stabil

--	--	--	--	--

 beweglich

2. Klangbeispiel

stabil

--	--	--	--	--

 beweglich

3. Klangbeispiel

stabil

--	--	--	--	--

 beweglich

4. Klangbeispiel

stabil

--	--	--	--	--

 beweglich

Suggestiv-Lokalisierung:

Hat sich die Schallquelle außerhalb Ihres Kopfes befunden? (ja/nein)

Wenn nicht, wo haben Sie die Schallquelle lokalisiert?

1. Klangbeispiel		
2. Klangbeispiel		
3. Klangbeispiel		
4. Klangbeispiel		

Plausibilität:

Wie glaubhaft erscheint Ihnen das Gehörte in dieser Umgebung?

Können Sie sich vorstellen, dass tatsächlich eine Maschine an der Stelle steht, wo Sie sie hören?

1. Klangbeispiel

glaubhaft

--	--	--	--	--

 unrealistisch

2. Klangbeispiel

glaubhaft

--	--	--	--	--

 unrealistisch

3. Klangbeispiel

glaubhaft

--	--	--	--	--

 unrealistisch

4. Klangbeispiel

glaubhaft

--	--	--	--	--

 unrealistisch

Immersion:

Immersion (lateinisch immergere = ein-, untertauchen) bezeichnet das vollständige Eintauchen in eine virtuelle Realität. Im Kontext von Audio Augmented Reality beschreibt es den kognitiven Zustand, bei dem ein virtuelles Klangobjekt als Teil der physischen Realität empfunden wird.

Welches Verfahren würden Sie spontan als immersiver bezeichnen?

1. Verfahren	
2. Verfahren	

Anhang B: Instruktion für den Hörversuch

Hamburg, Februar 2025

Mein Name ist Gunther Greim und ich studiere Medientechnik an der HAW. Im Rahmen meiner Bachelorarbeit zum Thema „Entwicklung eines Produktionsverfahrens für Audio Augmented Reality in einer urbanen Umgebung“ führe ich einen Hörversuch zur Untersuchung der immersiven Wirkung von virtuellen Schallquellen durch.

Testablauf:

1. **Einführung:** Sie erhalten eine kurze Einführung in den Versuch und die verwendeten Technologien.
2. **Geräuschedemonstration:** Um Sie mit der Art der verwendeten Klänge vertraut zu machen, hören Sie zu Beginn des Versuchs jeweils ein Geräusch aus jeder Kategorie.
3. **Hörproben:** Sie hören nacheinander vier verschiedene Klangbeispiele, die jeweils eine virtuelle Schallquelle darstellen. Auf Wunsch kann das jeweilige Klangbeispiel auch erneut gehört werden.
4. **Bewertung:** Nach jeder Hörprobe beantworten Sie Fragen zu Ihrer Wahrnehmung der Schallquelle.

Einführung:

Im folgenden Hörversuch werden zwei verschiedene Verfahren zur Simulation einer virtuellen Schallquelle angeboten. Bei der Schallquelle handelt es sich um eine industrielle Maschine. Ihre Aufgabe ist es, Ihre Wahrnehmung der Schallquelle anhand verschiedener Parameter zu bewerten. Sie können Ihren Kopf frei in alle Richtungen bewegen. Die Schallquelle ist virtuell, wird also nicht sichtbar, sondern nur zur hören sein. Es ist daher ausdrücklich erlaubt die Augen zu schließen, wenn es Ihnen bei der Beurteilung der Klänge hilft. Die Kopfhörer sind transparent. Das bedeutet, dass Sie die Umgebungsgeräusche zum Teil durchlassen. Dies ist natürlich und für den Versuch so beabsichtigt.

Hinweise zur Durchführung:

- Setzen Sie die Kopfhörer auf und stellen Sie sicher, dass Sie bequem sitzen.
- Sie können Ihren Kopf während der Wiedergabe frei in alle Richtungen drehen.
- Sie können Ihre Augen schließen.
- Das jeweilige Klangbeispiel kann auf Nachfrage erneut gehört werden.
- Beantworten Sie die Fragen ehrlich und spontan, es gibt keine richtigen oder falschen Antworten, wenn es um Ihre Wahrnehmung geht.
- Wenn Sie Fragen oder Unklarheiten haben, zögern Sie nicht, den Versuchsleiter anzusprechen.

Nach Abschluss des Versuchs haben Sie die Möglichkeit, Feedback zu geben und Fragen zu stellen. Ich danke Ihnen herzlich für Ihre Teilnahme und Ihre wertvollen Beiträge zu meiner Forschung!

Anhang C: Rohdaten Hörversuch

Anhang C1: Ergebnisse Verfahren A

Proband	Klangbeispiel	Winkel in °	Distanz in m	Lokalisierung in °	Richtigkeit
Benny	7	22,5	5	112,5	0
	1	0	8	180	0
Tobi	8	-22,5	5	-45	0
	2	22,5	8	22,5	1
Nico	4	45	8	45	1
	6	0	5	0	1
Niko	2	22,5	8	22,5	1
	8	-22,5	5	-22,5	1
Basti	2	22,5	8	22,5	1
	9	45	5	22,5	0
Niklas	6	0	5	-22,5	0
	4	45	8	45	1

Distanzwahrnehmung	Richtigkeit
8	0
8	1
8	0
8	1
8	1
5	1
8	1
5	1
8	1
8	0
5	1
8	1

Proband	Lokalisierung	Distanz	Ausdehnung	Stabilität	Plausibilität
Benny	0	0	2	2	2
	0	1	4	4	3
Tobi	0	0	2	2	1
	1	1	2	1	2
Nico	1	1	5	4	5
	1	1	4	4	4
Niko	1	1	4	4	5
	1	1	5	5	4
Basti	1	1	2	2	2
	0	0	1	1	1
Niklas	0	1	3	2	2
	1	1	2	2	2

Anhang C2: Ergebnisse Verfahren B

Proband	Klangbeispiel	Winkel in °	Distanz in m	Lokalisierung in °	Richtigkeit
Benny	12	22,5	8	112,5	0
	18	-22,5	5	-112,5	0
Tobi	14	45	8	22,5	0
	16	0	5	-22,5	0
Nico	11	0	8	0	1
	18	-22,5	5	-22,5	1
Niko	13	-22,5	8	-22,5	1
	20	-45	5	-45	1
Basti	12	22,5	8	0	0
	16	0	5	0	1
Niklas	13	-22,5	8	-22,5	1
	20	-45	5	-45	1

Distanzwahrnehmung	Richtigkeit
8	1
5	1
5	0
8	0
5	0
1	0
8	1
5	1
5	0
1	0
8	1
5	1

Proband	Lokalisierung	Distanz	Ausdehnung	Stabilität	Plausibilität
Benny	0	1	3	1	1
	0	1	3	4	3
Tobi	0	0	2	3	3
	0	0	3	4	1
Nico	1	0	2	2	3
	1	0	5	5	5
Niko	1	1	3	3	4
	1	1	4	4	5
Basti	0	0	2	2	2
	1	0	2	2	1
Niklas	1	1	2	4	4
	1	1	4	4	4

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel:

**Entwicklung eines Produktionsverfahrens für Audio Augmented Reality
in einer urbanen Umgebung**

selbständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Datum

Unterschrift