

BACHELORARBEIT

Vergleich von 3D-Audio-Systemen
für die Wiedergabe von bildbezogenen
Mehrkanal-Audioaufnahmen aus
Fußballstadien

vorgelegt am 07. Dezember 2023

Niklas Partsch XXXXXXXXXX

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

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**

Department Medientechnik
Finkenau 35
22081 Hamburg

Zusammenfassung

In dieser Arbeit werden die drei Wiedergabe-Systeme „Dolby Atmos“, „Ambisonics“ und „Direct Out“ miteinander anhand von Immersivität, Sound-Qualität und Lokalisation bei einer Übertragung von 3D-Audio aus dem Fußballstadion verglichen. Es wird eine Mehrkanal-Audioaufnahme für die Wiedergabe über Lautsprecher und Kopfhörer vorbereitet. Außerdem wird untersucht, ob die binaurale Wiedergabe über Kopfhörer genauso immersiv oder immersiver ist, als die Wiedergabe über Lautsprecher. Als letztes werden zwei verschiedene Audioperspektiven miteinander verglichen. Die Hörversuche haben ergeben, dass Dolby Atmos von den meisten Personen bevorzugt wird und bei den beschriebenen Aspekten am besten abschneidet. Die Unterschiede konnten mit einer statistischen Signifikanz belegt werden. Der Vergleich von Kopfhörern mit Lautsprechern konnte keine signifikanten Unterschiede aufweisen, vielmehr werden die unterschiedlichen Anwendungsmöglichkeiten deutlich. Bei der 3D-Audio-Übertragung für mehrere Personen eignet sich nur die Lautsprecher-Wiedergabe, aber Kopfhörer sind dennoch eine gute Alternative bei der Übertragung für einzelne Personen. Die mobile Audioperspektive, angepasst an das Kamerabild, wurde allgemein als eher störend oder verwirrend wahrgenommen. Die Proband*innen bevorzugten die statische Audioperspektive, bei der die Stadionatmosphäre unabhängig von der Kameraperspektive gleich bleibt.

Abstract

In this thesis three sound reproduction systems Dolby Atmos, Ambisonics and Direct Out are compared based on immersion, sound quality and localization using a 3D audio recording from a football stadium. A multi-channel audio recording is being prepared for the reproduction over loudspeakers and headphones. There will also be an examination, if the binaural reproduction over headphones will be as immersive as the reproduction over loudspeakers or better. At last, there will be a comparison between two different audio perspectives. The listening tests yield that Dolby Atmos is being preferred by most people and that it had the best performance of the described aspects. The differences were proven to be statistically significant. The comparison of headphones and loudspeakers could not result in significant differences, rather there are quite different possible applications. The 3D-audio broadcast for many people is only possible with loudspeakers, but headphones are nonetheless a viable alternative, if only one person is listening. The mobile audio perspective adjusted to the view of the camera was generally being perceived as annoying or confusing. The Subjects preferred the static audio perspective, in which the atmosphere of the stadium remains the same regardless of the camera perspective.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Bachelorarbeit unterstützt und motiviert haben.

Zuerst gebührt mein Dank Frau Prof. Dr. Eva Wilk, die meine Bachelorarbeit betreut und begutachtet hat. Für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Ich bedanke mich auch bei Matthias Fehling, der die Aufgabe des Zweitprüfers übernommen hat. Für die Hilfe bei der Umsetzung des Hörversuchs und hilfreiche Anregungen möchte ich mich ebenfalls herzlich bedanken.

Ein besonderer Dank gilt allen Teilnehmern und Teilnehmerinnen meines Hörversuchs, ohne die diese Arbeit nicht hätte entstehen können. Mein Dank gilt ihren interessanten Beiträgen und Antworten auf meine Fragen.

Ebenfalls möchte ich mich bei Chris Gebert von der Sportcast GmbH sowie Florian Jeschenko von der Deutschen Fußball Liga GmbH bedanken, die für mich das Audio- und Videomaterial und die Lizenzrechte zur Verfügung gestellt haben.

Außerdem möchte ich Thomas Klein für das Korrekturlesen meiner Bachelorarbeit danken.

Niklas Partsch

Hamburg, 07.12.2023

Inhaltsverzeichnis

1	Einleitung und Herangehensweise	1
1.1	Einleitung	1
1.2	Aufbau und Herangehensweise	3
1.2.1	Forschungsstand	3
1.2.2	Methodik und Vorbereitung	4
1.2.3	Hörtests	4
1.2.4	Auswertung und Diskussion	5
2	Forschungsstand	6
2.1	Immersives Audio	6
2.2	Dolby Atmos	7
2.3	Ambisonics	8
2.4	Objektbasiertes und kanalbasiertes 3D-Audio	11
2.5	Sweet-Spot	11
2.6	Binaurales Audio	13
2.7	Overall Listening Experience	14
3	Methodik und Vorbereitung	15
3.1	Methodik	15
3.1.1	Audio-Signal	15
3.1.2	Hörversuch	16
3.1.3	Bewertung der Sweet-Area	16
3.1.4	Auswertung der Hörversuche	16
3.2	Vorbereitung	17
3.2.1	Raum	17
3.2.2	Lautsprecher-Setup	18
3.2.3	Software und Plug-Ins	20
3.2.4	Setup des Wiedergabe-Systems Dolby Atmos	21
3.2.5	Setup des Wiedergabe-Systems Ambisonics	23
3.2.6	Setup des Wiedergabe-Systems Direct Out	25

3.2.7	Vorbereitung Hörversuch eins (HV 1): Vergleich der Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out . . .	25
3.2.8	Vorbereitung Hörversuch zwei (HV 2): Single-Listener Experience – LS vs. HP	26
3.2.9	Vorbereitung Hörversuch drei (HV 3): Vergleich von zwei Audioperspektiven – statisch und mobil	28
3.2.10	Bewertung der Sweet-Area	29
4	Durchführung und Ergebnisse	32
4.1	Durchführung des Hörversuchs	32
4.1.1	Proband*innen	32
4.1.2	Dauer des Versuchs	32
4.1.3	Ablauf	33
4.2	Ergebnisse	34
4.2.1	Hörversuch 1	35
4.2.2	Hörversuch 2	36
4.2.3	Hörversuch 3	37
4.2.4	Anmerkungen	37
4.2.5	Notizen	40
5	Auswertung und Diskussion	41
5.1	Hörversuch 1	41
5.2	Hörversuch 2	44
5.3	Hörversuch 3	46
5.4	Ausblick	48
6	Zusammenfassung	50
A	Fragebogen des Hörversuchs	55

Abbildungsverzeichnis

2.1	Sphärische Harmonien nach Franz Zotter	10
3.1	Skizze und Nummerierung des Lautsprecher-Setups	18
3.2	3D-Grafik der Lautsprecheranordnung	18
3.3	Frequenzabhängige Nachhallzeiten des Raums	18
3.4	Lautsprecher-Setup 5.1.4	19
3.5	Lautsprecher-Setup 11.1	19
3.6	Lautsprecher-Tabelle mit Winkeln und Abständen	21
3.7	ADM Authoring für Dolby Atmos	22
3.8	Dolby Atmos Setup Assistant	22
3.9	IEM AllRAD Decoder für Ambisonics	23
3.10	IEM MultiEncoder für Ambisonics	24
3.11	IEM EnergyVisualizer für Ambisonics	24
3.12	Beurteilung der Sweet-Area bei Dolby Atmos	31
3.13	Beurteilung der Sweet-Area bei Ambisonics	31
4.1	Aufbau des Hörversuchs	33
4.2	Ergebnisse des Hörversuch 1	35
4.3	Ergebnisse des Hörversuch 2	36
4.4	Ergebnisse des Hörversuch 3	38
5.1	Ergebnisse der einfaktoriellen Varianzanalyse von Hörversuch 1	42
5.2	Ergebnisse des t-Test von Hörversuch 2	45

Tabellenverzeichnis

1.1 Die 16 diskreten Audiospuren	2
3.1 Auswahl der verwendeten Lautsprecher	20
3.2 Winkelabmessungen und Abstände für die Beurteilung der Sweet-Area	30

Abkürzungsverzeichnis

ANOVA Analysis of Variance

C Center

CH Center-Height

DAW Digital Audio Workstation

HD High Definition

HLF Height-Left-Front

HLR Height-Left-Rear

HOA Higher Order Ambisonics

HP Kopfhörer

HRF Height-Right-Front

HRR Height-Right-Rear

HRTF Head-Related Transfer Function

HV 1 Hörversuch eins

HV 2 Hörversuch zwei

HV 3 Hörversuch drei

IEM Institute of Electronic Music and Acoustics

IKL Im-Kopf-Lokalisation

L Left

LFE Low-Frequency-Emitter

LS Left-Surround

LS Lautsprecher

MOM Monitoring Output Module

OLE Overall Listener Experience

QoE Quality of Experience

R Right

RS Right-Surround

T Top

UHD Ultra High Definition

Kapitel 1

Einleitung und Herangehensweise

1.1 Einleitung

In den letzten Jahren hat 3D-Audio immer mehr an Bedeutung gewonnen. 3D-Audio, das heißt die Wiedergabe von Sound mit horizontaler und vertikaler Information, wurde bereits in den 70er Jahren mit Ambisonics erfunden, wobei die Idee davon sogar schon viel früher durch Blumlein entstanden ist (Malham 2019). Erst mit der Erfindung von Dolby Atmos 2012 (*Dolby Atmos 2023a*) ist die Produktion und Wiedergabe von 3D-Audio vor allem in Kinos sehr viel verbreiteter geworden. Mittlerweile bieten sogar Streaming-Dienste wie Netflix Filme mit 3D-Audio an. Auch im Fernsehen überträgt der Sender Sky seit der Fußballsaison 2022/2023 einmal pro Woche das Topspiel am Samstagabend um 18:30 Uhr mit der Option, das Spiel in U-HD und mit Dolby Atmos zu sehen beziehungsweise zu hören. In dieser Arbeit werden verschiedene Möglichkeiten miteinander verglichen, eine Video-Übertragung von Sport-Events mit 3D-Audio zu hören. Beispielhaft wird dazu eine Produktion eines Fußballspiels betrachtet.

Für diese Arbeit liegt eine Mehrkanal-Audioaufnahme in Form von 16 diskreten Audiospuren vor (siehe Tabelle 1.1). Die Audiosignale wurden von Sportcast GmbH (*Sportcast 2023*) bereits aufgenommen und befinden sich in dem Zustand, wie sie der Sender Sky (*Sky 2023*) erhalten und mit Dolby Atmos übertragen würde. Die fertigen 3D-Audio-Mischungen bieten die Möglichkeit, dass Fans im Wohnzimmer oder Zuschauer*innen in einer Sports-Bar die Übertragung immersiv erleben können oder einzeln 3D-Audio über Lautsprecher (LS) oder Kopfhörer (HP) hören können. Die Wiedergabe über LS kann mit drei Methoden erzielt werden: Dolby Atmos, Ambisonics oder direktem Routing ohne Si-

Spur Nr.	Kanal	Abkürzung
1	Left	L
2	Right	R
3	Center	C
4	Low Frequency Emitter	LFE
5	Left Surround	LS
6	Right Surround	RS
7	Height Left Front	HLF
8	Height Right Front	HRF
9	Height Left Rear	HLR
10	Height Right Rear	HRR
11	Com + Exp Language 1	Com 1
12	Com + Exp Language 2	Com 2
13	Close Balls FX + AfV FX	FX
14	Highlevel (PA, EVS, Zuspüler, etc.)	PA
15	Fan Home	Fan H
16	Fan Away	Fan A

Tabelle 1.1: 16 diskrete Audio-Spuren der vorliegenden Mehrkanal-Audioaufnahme mit Bezeichnung und Abkürzung

gnalverarbeitung (im Folgenden als „Direct Out“ bezeichnet). Die drei verschiedenen Systeme werden auf ihre Immersivität, Sound-Qualität und Lokalisation hin untersucht und verglichen. Sie unterscheiden sich durch die Signalverarbeitung (objektbasiert, kanalbasiert und keine Signalverarbeitung) und besitzen einen unterschiedlich großen Bereich, in dem das Hörerlebnis optimal ist („Sweet-Spot“ oder „Sweet-Area“ genannt). Da sich in eine Sports-Bar oder im Wohnzimmer nicht alle Zuhörer*innen im optimalen Hörbereich des Audiosystems befinden können, ist es notwendig, die Größe der Sweet-Area zu analysieren. Dies erfolgt durch einen Vorversuch, der die Wahrnehmung der Audioszene am Randbereich überprüft. Außerhalb der Sweet-Area geht der Effekt des räumlichen Hörens nämlich verloren und es findet keine realitätsgetreue Lokalisation statt.

Zusätzlich wird neben der Wiedergabe über LS für mehrere Zuschauer*innen eine Kopfhörermischung angefertigt. Dabei werden die Audiosignale binaural kodiert, wodurch so der 3D-Sound entsteht. Bei der Kopfhörer-Wiedergabe kann die Person nicht außerhalb der Sweet-Area sitzen. Allerdings gibt es hier die Probleme, dass sich die Wiedergabe über HP nur für einzelne Zuhörer*innen eignet und es zur Im-Kopf-Lokalisation (IKL)¹ kommen kann. Die Wiedergabe über HP wird deshalb mit der Wiedergabe über LS verglichen und auf Immersivität, Lokalisation und Sound-Qualität hin untersucht.

Abschließend werden zwei verschiedene Audioperspektiven miteinander ver-

¹IKL ist der Effekt, dass Schallquellen nicht außerhalb in der Umgebung wahrgenommen werden, sondern im Kopf zu liegen scheinen (*Im-Kopf-Lokalisation* 2019)

glichen. Durch die Wiedergabe über eine 3D-Soundanlage kann die Ausrichtung der Audioszene beliebig um die Zuhörer*innen rotiert und so die Audioperspektive an das Bild der Kamera angepasst werden. Bei der statischen Audioperspektive befindet sich der/die Zuhörer*in, an einer festen virtuellen Position im Stadion, unabhängig davon, welche Perspektive die Kamera zeigt. Dies ist die herkömmliche Art ein Fußballspiel zu übertragen und wird bei allen Stereo-Produktion verwendet. Bei der zweiten Variante, der mobilen Audioperspektive, soll der/die Zuhörer*in die Stadionatmosphäre immer so erleben, wie das Kamerabild gerade das Stadion zeigt. In diesem Szenario befindet sich der/die Zuhörer*in direkt hinter der Kamera und das Audiosignal wird abhängig von dem Kamerabild präsentiert. Werden die Auswärts-Fans in der totalen Kameraperspektive beispielsweise auf der linken Seite gezeigt, müssen die Gesänge der Fans auditiv auf der linken Seite wiedergegeben werden. Wenn die Kamera gewechselt und auf das Tor gerichtet wird, sodass die Fans nun zentral im Kamerabild zu sehen sind, müssen die Gesänge der Fans auch zentral aus dem Center (C)-Kanal wiedergegeben werden. Über einen drei-dimensionalen Panner lässt sich die Atmosphäre im Stadion drehen und dem Kamerabild anpassen. Leitfragen zu diesem Vergleich sind, ob das Hörerlebnis durch die mobile Kameraperspektive immersiver wird, oder ob die auditiven Perspektivenwechsel irritierend oder sogar störend wahrgenommen werden.

1.2 Aufbau und Herangehensweise

1.2.1 Forschungsstand

Vorbereitend auf den Hörversuch wird der Forschungsstand zu relevanten Themen erarbeitet. Dies beginnt mit dem allgemeinen Thema „Immersive Sound“, wobei speziell binaural- und mehrkanal-Audio im Fokus liegt. Anschließend werden die beiden Methoden Dolby Atmos und Ambisonics thematisiert. Bei Dolby Atmos wird der Schritt von Surround (5.1/7.1) zu Atmos (5.1.4/7.1.4) vorgestellt und die Rolle der „unterstützenden“ Lautsprecher erarbeitet. Des Weiteren wird das Prinzip von objektbasiertem 3D-Audio in Dolby Atmos und die Verwendung von „Beds“ und „Objects“ erklärt. In einer Studie von Oramus und Neubauer vergleichen diese verschiedene Lautsprecher-Setups von Dolby (Oramus & Neubauer 2019). Bei dem Thema Ambisonics werden verschiedenen Studien betrachtet, welche unterschiedliche Formen von Higher Order Ambisonics (HOA) miteinander vergleichen. Thresh et al. untersuchen dabei die Lokalisationsgenauigkeit bei 1st-, 3rd- und 5th-Order Ambisonics (Thresh, Armstrong & Kearney 2017). Anschließend wird der Unterschied von objektbasiertem Audio

zu kanalbasiertem Audio betrachtet. Dadurch werden die beiden Wiedergabe-Systeme Dolby Atmos und Ambisonics erstmals gegenübergestellt und können so besser miteinander verglichen werden. Bei dem nächsten Thema wird der Sweet-Spot beziehungsweise die Sweet-Area untersucht. Frank und Zotter entwickeln in einer Studie eine Methode, diese bei verschiedenen Lautsprecheranordnungen genauer zu bestimmen. In einer weiteren Studie untersuchen Stitt et al. dann die off-center-Lokalisation. Anschließend wird das Thema „binaural Audio“ aufgearbeitet. Zum Schluss wird vorbereitend auf die Hörtests die Overall Listener Experience (OLE) vorgestellt, da diese bei der Entwicklung des Fragebogens für den Hörversuch verwendet wird.

1.2.2 Methodik und Vorbereitung

Nach dem Aufarbeiten des Forschungsstands der verschiedenen Themen wird die Methodik dieser Arbeit vorgestellt. Als Erstes wird erklärt, welches Signal verwendet wird. Dieses liegt in Form von 16 diskreten Audiospuren vor und muss für die Wiedergabe aufbereitet werden. Zur Erhebung der Daten wird ein Hörversuch durchgeführt, bei dem Proband*innen in einem vorbereiteten Raum die verschiedenen Wiedergabe-Formate hören und beurteilen. Außerdem wird die Sweet-Area der Wiedergabe-Systeme Dolby Atmos und Ambisonics untersucht. Anschließend an den Hörversuch werden die Daten analysiert und auf statistische Signifikanz geprüft.

Nach dem Kapitel der Methodik wird der Hörversuch vorbereitet. Hier wird der Raum, in dem der Hörversuch durchgeführt wird, sowie die Lautsprecheranordnung vorgestellt. Danach werden sämtliche Software und Plug-Ins genannt, welche für die einzelnen Wiedergabeformate benötigt werden. Ferner wird beschrieben, wie die einzelnen Wiedergabe-Systeme aufgebaut und eingerichtet werden. Danach werden die Hörversuche genauer erläutert und das Ziel des jeweiligen Hörversuchs dargelegt. Außerdem werden die Fragen formuliert, welche später den Fragebogen des Hörversuchs bilden. Abschließend wird nach Frank und Zotter die Sweet-Area von Dolby Atmos und Ambisonics ermittelt.

1.2.3 Hörtests

Es wird drei verschiedene Hörtests geben, welche voneinander unabhängig sind. Bei dem ersten Hörtest (Vergleich der Wiedergabe-Systeme) werden die drei Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out auf verschiedene Kriterien wie Immersivität, Soundqualität, Lokalisation und Präferenz hin mit-

einander verglichen. Es soll analysiert werden, ob sich die drei Systeme, auf die genannten Kriterien bezogen, unterscheiden und ein System von den Proband*innen bevorzugt wird. Mithilfe dieser Analyse wird ermittelt, welches Wiedergabe-System sich am besten für eine Sports-Bar oder das Wohnzimmer eignet. Der zweite Hörtest (LS vs. HP) untersucht, ob die Wiedergabe über HP vergleichbar ist mit der Wiedergabe über LS. Auch hier werden Aspekte wie Immersivität, Soundqualität, Lokalisation und OLE verglichen. Außerdem wird das Anwendungsgebiet der beiden Formate hinterfragt. Während sich die LS-Wiedergabe für mehrere Personen eignet, kann die HP-Wiedergabe für einzelne Personen genutzt werden. Zusätzlich ist die Audiowiedergabe über Kopfhörer kostengünstiger und benutzerfreundlicher für den/die Verbraucher*in. Schließlich ist es deutlich einfacher ein paar Kopfhörer zu kaufen, als sich ein 3D-Audio fähiges Heimkino einzurichten. Bei dem dritten Hörtest (Audioperspektive) werden zwei unterschiedliche Wiedergabe-Methoden miteinander verglichen. Neben der herkömmlichen statischen Audioperspektive wird eine zweite mobile Audioperspektive abgemischt, sodass der Zuschauer sich auditiv immer in dem Blickwinkel der Kamera befindet. Hier wird untersucht, welche Wirkung beide Perspektiven auf den Zuhörer haben und ob durch die mobile Audioperspektive ein immersiveres Erlebnis wahrgenommen wird.

1.2.4 Auswertung und Diskussion

Anschließend an die Durchführung der Hörversuche werden die erhobenen Daten präsentiert, ausgewertet und diskutiert.

Kapitel 2

Forschungsstand

In den folgenden Abschnitten wird der aktuelle Forschungsstand zu den verschiedenen Unterthemen sowie relevante Publikationen für diese Arbeit präsentiert. Dabei wird von einem grundlegenden Verständnis der Themen Hören, Klanglokalisation und allgemeinen Audioaspekten ausgegangen.

2.1 Immersives Audio

Unter Immersive Audio oder oft auch Spatial Audio oder Multichannel Audio versteht man in erster Linie einen den Hörer komplett umschließenden Sound. Dieser kann aus einzelnen, spezifischen Soundquellen, welche aus bestimmten Richtungen lokalisierbar sind, oder ambienten Sounds, welche Umgebungssounds wiedergeben, bestehen. Durch viele um den Hörer positionierte Lautsprecher kann der Sound aus allen Richtungen abgestrahlt werden, um damit ein möglichst realistisches Klangbild zu erzeugen. Dabei sollen vor allem auch Schallreflexionen simuliert werden, welche die Audio-Szene realistisch werden lassen. Die hörende Person hat so das Gefühl, bei der reproduzierten Audio-Szene an dem ursprünglichen Ort der Schallquelle zu sein (Roginska & Geluso 2017, Kapitel 1). Im Gegensatz zu dem Visuellen könne Sound komplett immersiv sein, da er aus allen Richtungen aufgenommen werden kann (Roginska & Geluso 2017). Die Wahrnehmung eines immersiven Sound-Erlebnis reicht von dem allgemeinen peripheren Hören bis hin zur Erkennung von verarbeiteten Informationen im Gehirn (Roginska & Geluso 2017, S. 33f). Ein Beispiel dafür ist die wahrgenommene Distanz von Schallquellen bei Pegel- oder Hall-Anpassungen. Ebenso kann Stereo in gewisser Weise als immersiv betrachtet werden, da es das Klangbild einer punktförmigen Schallquelle durch die Erweiterung auf die linke und rechte Seite ausdehnt. Deutlich immersiver ist dann der Surround Sound 5.1 oder 7.1, bei dem nicht nur links und rechts abgebildet wer-

den kann, sondern nun auch Sound von vorne und hinten abgespielt werden kann. Bei 3D-Audio wird durch das Hinzufügen von erhöhten Lautsprechern eine vertikale Abbildungsebene ermöglicht. Eine Abbildungsrichtung, welche oft vernachlässigt wird, ist unterhalb der hörenden Person, da dies in der Regel durch den Boden beschränkt wird (Grewe, Walther & Klapp 2019). Durch die Anordnung von Lautsprechern links und rechts, vorne und hinten und oben kann Sound auf einer Halbkugel über dem Hörer verteilt werden.

Auch bei Spatial-Audio bilden die drei Kanäle C, Left (L) und Right (R) die wichtigsten Lautsprecher, da durch sie ein Gefühl für die vordere Richtung vermittelt werden kann. Im Kino werden hier beispielsweise bedeutende Elemente wie die Dialoge wiedergegeben. Die Surround-Kanäle Left-Surround (LS) und Right-Surround (RS) sowie Height-Kanäle Height-Left-Front (HLF), Height-Right-Front (HRF) Height-Left-Rear (HLR) und Height-Right-Rear (HRR) können viel mehr für Atmosphärengeräusche oder Reflexionen des Raums genutzt werden. Diese lässt die Audio-Szene realistisch klingen.

Das Hinzufügen von Surround-Kanälen auf der horizontalen sowie vertikalen Achse bewirkt auch vor allem eine Steigerung in der wahrgenommenen Tiefe und einer erhöhten Verbindung von Sound-Objekten mit den visuellen Bildern (Kamekawa, Marui, Date & Enatsu 2011). Untersuchungen von Kim et al. zeigen, dass die Surround- und Height-Lautsprecher großen Einfluss auf die wahrgenommene Soundqualität nehmen. Diese haben die Proband*innen mit Aspekten wie Natürlichkeit, Geräumigkeit, Umfang, Immersivität und der Dimension des reproduzierten Soundfelds bewertet (Kim, Ko, Nagendra & Woszczyk 2013).

2.2 Dolby Atmos

Dolby Laboratories, Inc. ist ein Unternehmen, welches 1965 gegründet wurde und für mehrere Audio Standards wie Dolby Digital (Mehrkanal-Tonsystem, welches seit 1995 das Standardformat bei DVDs ist), Dolby Surround (Mehrkanal-Tonsystem für den Heimbereich) und die jüngste Entwicklung Dolby Atmos (objektbasiertes Surround-Sound System) bekannt ist (*Dolby Atmos 2023b*).

Ziel von Dolby Surround und Dolby Atmos ist es, das Erlebnis der Zuhörer*innen immersiver zu gestalten. Dolby Surround fügt im Vergleich zu Dolby Stereo von hinten kommenden Sound hinzu. Dabei werden die Lautsprecher L und R um einen C-Lautsprecher vorne und zwei Lautsprecher LS und RS hinten ergänzt. Die Surround-Kanäle sind in erster Linie dazu da, Reflexionen einer realen Schallquelle zu simulieren und so den/die Zuschauer*in immersiver in das Hörerlebnis eintauchen zu lassen.

In der Regel wird das Wiedergabe-Setup bei Dolby Atmos um vier erhöhte Lautsprecher erweitert. Diese ergänzen die Kanäle L und R mit HLF und HRF und die Kanäle LS und RS mit HLR und HRR, welche sich über den Lautsprechern auf Ohrhöhe befinden. Dadurch wird aus einem 5.1 bzw. 7.1 Surround ein 5.1.4 bzw. 7.1.4 Atmos. Mit dem Hinzufügen von erhöhten Lautsprechern erzeugt Dolby eine Halb-Sphäre oberhalb des/der Zuhörer*in, über die Sounds von oben wiedergegeben werden können. Außerdem ist es durch das objektbasierte Mischen möglich, einzelne Soundquellen im Raum zu verteilen.

Dolby Atmos ist ein objektbasiertes Format, welches mit sogenannten „Beds“ und „Objects“ arbeitet. Ein Bed ist ein kanalbasierter Premix, welcher Mehrkanal-Panning erlaubt und für Stereo- oder Surround-Panning benutzt wird (*Dolby Atmos 2023a*). Ein Bed kann unter anderem als 2.0-, 5.1-, 7.1- oder 7.1.2-Setup konfiguriert werden, wobei 7.1.2 die höchste Stufe ist. Beds werden in der Regel dafür genutzt, Sounds wie Atmosphäre, Ambiente und Geräusche abzuspielen, bei denen es sich nicht lohnt, ein eigenes Objekt zu erstellen. Objects erhalten im Dolby Atmos Renderer einen (mono) oder zwei (stereo) eigene Kanäle und können über einen eigenen Panner frei innerhalb des Setups bewegt werden. Anhand der oberen Grenze von 128 Kanälen innerhalb des Renderers ist somit die Anzahl an Objekten, die erstellt werden können, begrenzt. Stößt man an diese Obergrenze, lohnt es sich zu überlegen, verschiedene Objects von weniger hoher Relevanz mit in das Bed zu verlegen. Es können auch mehrere Beds erstellt werden.

In einer Vergleichsstudie von Oramus und Neubauer vergleichen diese die Wahrnehmung von zwei Zuhörer-Gruppen bei 5.1 Surround und Dolby Atmos (Oramus & Neubauer 2019). Dabei haben sie mit räumlicher Immersion, Lokalisation, Dynamik, Audio-Qualität und Format-Präferenz fünf Aspekte definiert, welche die Proband*innen bewertet haben. Sie sind zwar nicht zu einem signifikanten Unterschied der beiden Gruppen gekommen, beide Gruppen tendieren aber leicht zu Dolby Atmos. Auch in ihrem e-Brief kommen sie bei dem Vergleich von 5.1, 7.1 und Atmos bezüglich der Lokalisationsgenauigkeit zu einem ähnlichen Ergebnis (Oramus & Neubauer 2020). Während es zu keiner signifikant besseren Lokalisation bei Atmos kommt, konnten die Proband*innen mit einer höheren Sicherheit sagen, wo sich eine Soundquelle befindet.

2.3 Ambisonics

Ambisonics entstand bereits in den 70er Jahren durch britische Forscher um Michael Gerzon. Diese entwickelten ein Surround-Sound System, welches es ermöglichen würde, ein musikalisches Event aufzunehmen und dann in einem

Wohnzimmer möglichst genau mit den originalen akustischen Eigenschaften wiederzugeben (Elen 2001). Dabei haben sie ein spezielles Mikrofon-Array, das „Soundfield Microphone“ entwickelt oder benutzen eine besondere panpot Konsole, die die Lokalisation von Soundquellen kontrollieren kann. Dieses Equipment generiert ein 4-Kanal Signal (B-Format), welches alle Informationen des Soundfelds in links-rechts, vorne-hinten und oben-unten plus ein mono-Referenzsignal einordnet. Das Signal in B-Format wird durch einen Decoder dann für die einzelnen Lautsprecher umgerechnet und jeder einzelne Lautsprecher bekommt die Informationen abhängig von seiner Position im Raum. Dabei ist Ambisonics nicht an ein festes Lautsprecher-Setup gebunden, sondern die Informationen werden an beliebige Lautsprecherpositionen angepasst. Um den Sweet-Spot etwas zu vergrößern, werden dem Signal außerdem Richtungshinweise durch Pegel, Phase oder andere Arten gegeben, um das menschliche Hören nachzuahmen.

In einer Arbeit von Malham werden die frühen Jahre des Ambisonics aufgearbeitet und sie erwähnen, dass die Grundzüge für Ambisonics bereits mit dem Stereo Patent von Alan Blumlein in den 30er Jahren gesetzt worden wären (vgl. (Malham 2019)). Während Blumlein (Blumlein 1958) an den Ideen für eine Stereo-Wiedergabe arbeitete, wurden auch viele Ansätze angedeutet, die später zu Ambisonics geführt haben.

Es gibt mehrere Stufen des sogenannten Higher Order Ambisonics (HOA). Mit 1st-Order Ambisonics kann schon mit nur vier Kanälen und sechs Lautsprechern Sound auf der vollen Sphäre wiedergegeben werden. Auch das Stereo-Format gehört zu 1st-Order Ambisonics, es kann allerdings nur die horizontale Ebene vor dem/der Hörer*in abgebildet werden. 2nd-Order Ambisonics verwendet schon bis zu neun Kanäle und 3rd-Order Ambisonics bis zu 16. Es gibt mittlerweile auch Systeme, die mit 5th-Order oder sogar 7th-order Ambisonics arbeiten, allerdings werden dazu extrem viele Lautsprecher benötigt, sodass dies oft technisch schwierig umzusetzen ist. Die sphärischen Harmonien nach Zotter (Zotter 2023) sind in Abbildung 2.1 abgebildet.

Es gibt viele Studien, die untersuchen, wie sich die verschiedenen HOA-Formate in der Lokalisation oder Immersion unterscheiden. Thresh et al. haben zu dieser Fragestellung ein Lokalisation-Performance Test gemacht, bei dem sie 1st-, 3rd-, und 5th-Order Ambisonics auf realen und virtuellen Lautsprechern wiedergegeben haben (Thresh, Armstrong & Kearney 2017). Dabei kommen sie zu dem Ergebnis, dass 5th-Order Ambisonics eine leichte Verbesserung bei der Lokalisation gegenüber 3rd-Order Ambisonics zeigt, und sowohl 3rd- als auch 5th-Order Ambisonics eine starke Verbesserung gegenüber 1st-Order Ambisonics aufweisen. Außerdem konnten sie einen drastischen Unterschied zwi-

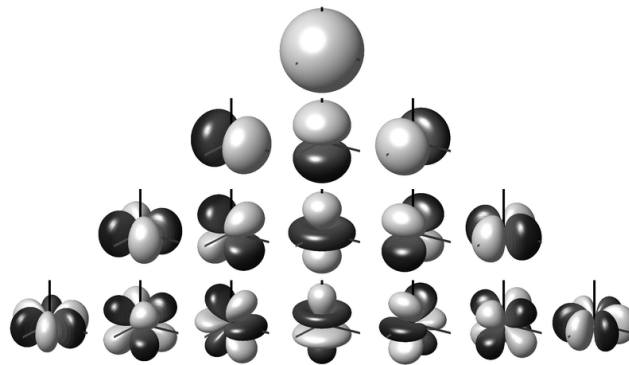


Abbildung 2.1: Eine Visualisierung der ersten ‚spherical harmonics‘ nach Franz Zotter

schen der Wiedergabe auf realen und virtuellen Lautsprechern feststellen. Bei der Wiedergabe über virtuelle Lautsprecher wurde bei HOA festgestellt, dass eine vorne-hinten und oben-unten Verwirrung entstand, falls keine individualisierte Head-Related Transfer Function (HRTF) benutzt wurde. Außerdem soll der sogenannte „cone of confusion“ – ein Bereich an beiden Ohren, bei dem die Schallquelle für dieselben interauralen Differenzen an beiden Ohren entstehen und so eine Lokalisation nur schwierig möglich ist – für Ausreißer im Datensatz gesorgt haben und eine klare Aussage war erst nach dem Ignorieren dieser möglich.

Im Vergleich zu anderen Surround-Formaten werden bei Ambisonics alle Richtungen gleichbedeutend behandelt. Es gibt keinen Hauptkanal (Center (C)), der wichtige Sounds von vorne wiedergibt und die Surround-Kanäle nur für Reflexionen, Ambiente oder Spezialeffekte zuständig sind. Außerdem ist bei Ambisonics das Audiosignal losgelöst vom Wiedergabe-Setup, das heißt Lautsprecher können frei im Raum platziert werden. Das Signal wird dann innerhalb des Decoders an die Lautsprecher-Anordnung angepasst. HOA ist auch abwärts kompatibel und kann auf LS-Anordnungen mit geringerer räumlicher Auflösung wiedergegeben werden. Es muss kein Downmix angefertigt werden.

Bis heute konnte sich Ambisonics jedoch noch nicht richtig als Wiedergabe-System bei Filmen durchsetzen und gerade objektbasierte 3D Wiedergabe-Verfahren finden hier vermehrt ihren Einsatz. Ambisonics und vor allem HOA wird dagegen häufig in der Spielindustrie und bei Anwendungen mit virtueller Realität benutzt (Olivieri, Peters & Sen 2019). Der Vorteil dabei ist, dass keine Lautsprecher aufgebaut werden müssen und die Informationen für die binaurale Wiedergabe kodiert werden. Auch in 360°-Videos, bei denen der Zuschauer die Kameraperspektive selbst verändern kann, kommt Ambisonics zum Einsatz (Thomas 2017).

2.4 Objektbasiertes und kanalbasiertes 3D-Audio

Objektbasiertes Audio wird in der Audio-Industrie und bei dem/der Konsument*in immer beliebter. Durch neue Werkzeuge, Codecs und Standards wird objektbasiertes Mischen schnell zu dem bevorzugten Ansatz für 3D-Audio in Kinos (2016 wurden neun von zehn Oscar-nominierte Filme mit einem objektbasiertem Format gemischt) (Roginska & Geluso 2017, S. 272)

In einer Studie von Oramus und Neubauer untersuchen diese, ob es bei objektbasiertem Audio zu einer genaueren Lokalisation von Schallquellen kommt, als bei kanalbasiertem Audio (Oramus & Neubauer 2020) (vgl. Kapitel 2.2). Seit der Erfindung von Dolby Atmos 2012 und dem damit verbundenen Aufstieg des objektbasierten Mischens wird häufig die Erwartung geweckt, dass durch diese Methode eine präzisere räumliche Verteilung von Schallquellen möglich ist. Mit einem Hörtest vergleichen diese die Wahrnehmung zweier Gruppen bei kanalbasiertem Mischen auf einem 5.1 und 7.1 Lautsprecher-Setup und objektbasiertem Mischen von Dolby Atmos. Zu beachten ist hier, dass bei Dolby Atmos mehr Lautsprecher benutzt wurden als es bei 5.1- oder 7.1-Surround der Fall war. Obwohl es keinen signifikanten Unterschied zwischen den beiden Gruppen gibt, tendieren beide für die Wiedergabe über Atmos.

2.5 Sweet-Spot

Als Sweet-Spot oder auch Sweet-Area bezeichnet man den Bereich eines Wiedergabe-Systems, bei dem der Sound optimal wahrgenommen wird. Bei der Stereo-Wiedergabe wird der Sweet-Spot über das Stereodreieck ermittelt, indem der/die Hörer*in mit den Lautsprechern ein gleichseitiges Dreieck bildet. Befindet sich der Kopf von beiden Lautsprechern gleich weit entfernt, entsteht der optimale Stereo-Eindruck. Dabei hat jedoch auch die Akustik des Raums einen Einfluss auf den Sweet-Spot, da dieser durch unerwünschte Reflexionen verschoben werden kann. Bei der Wiedergabe mit mehr als zwei Lautsprechern, wie zum Beispiel Dolby Surround, ist es schon sehr viel schwieriger den Sweet-Spot zu ermitteln. Auch hier ist eine einheitliche Entfernung zu allen Lautsprechern notwendig. Falls die Entfernung zu allen Lautsprechern nicht gleich sein kann, muss dies über die Lautstärke-Pegel kompensiert werden. Wenn mehrere Personen 3D-Audio über das selbe Lautsprecher-Setup hören, wird es schwieriger, allen Personen das optimale Sounderlebnis zu ermöglichen.

In einer Forschungsarbeit von Frank und Zotter haben diese eine Methode entwickelt, die Sweet-Area bei verschiedenen Ordnungen von Ambisonics zu bestimmen (Frank & Zotter 2017). Sie behaupten, dass in der Praxis die Sweet-Area

sehr viel größer sei, als sie oft angenommen wird. Die Sweet-Area definieren sie als den Bereich, in dem die Wahrnehmung der wiedergegebenen Sounds verglichen mit der realen Audioszene plausibel ist. Um die Sweet-Area experimentell zu erfassen, wurden von der Mitte aus zu den jeweiligen Lautsprechern Linien gezogen, auf denen sich die Proband*innen bewegen. Während die Proband*innen immer den C-Kanal anschauen müssen, sollen sie dann auf den Linien markieren, ab wo sich die Wiedergabe nicht mehr plausibel anhört. Bei 1st-Order Ambisonics sind die Radien des plausiblen Playbacks nur die Hälfte der Entfernungen zu den Lautsprechern. Bei 3rd-Order Ambisonics sind die Radien mit zweidrittel der LS-Entfernungen größer. 5th-Order Ambisonics hat bei der Wiedergabe mit Hall eine noch größere Sweet-Area und bei dem trockenen Audiosignal ist die Sweet-Area sogar der komplette Hörbereich. Dementsprechend empfehlen Frank und Zotter für die Wiedergabe mit vielen Menschen HOA (Frank & Zotter 2017). Es ist allerdings auch festzuhalten, dass Frank und Zotter bei der Ermittlung der Sweet-Area von einem akustisch plausiblen Hörbereich ausgehen und nicht von der optimalen Wahrnehmung. Auch 3rd-Order Ambisonics hat schon einen relativ großen Bereich, in dem das Hörerlebnis noch realistisch wahrgenommen wird.

Stitt et al. haben in einer Forschungsarbeit die „off-center“ Lokalisation untersucht und bei 1st- und 3rd-Order Ambisonics auf zwei Lautsprecheranordnungen mit unterschiedlichen Radien getestet (Stitt, Bertet & Walstijn 2014). Während bei 1st-Order mit vier Lautsprechern die Lokalisation deutlich ungenauer ist als bei 3rd-Order Ambisonics, so gibt es bei beiden Formaten eine erhöhte Ungenauigkeit bei der off-center Lokalisation als bei der Lokalisation im Sweet-Spot. Der Effekt ist bei 3rd-Order Ambisonics mit acht Lautsprechern allerdings schon deutlich geringer. Daraus ist zu schließen, dass bei HOA die Sweet-Area deutlich größer ist, als bei 1st-Order Ambisonics und eine realitätsgetreue Lokalisation außerhalb der mittleren Hörposition deutlich besser möglich ist. Außerdem erkennen Stitt et al. bei dem Vergleich der zwei Setups mit unterschiedlichen Radien, dass die zeitliche Differenz bei der off-center Lokalisation deutlich unwichtiger ist als der Lautstärkeunterschied, der durch die unterschiedlichen Entfernungen zu den Lautsprechern entsteht.

Über die Jahre hinweg wurde oft versucht, den Sweet-Spot zu vergrößern. Bassères und Thevenot haben in einer Arbeit versucht, den Sweet-Spot bei 5.1 Monitoring zu vergrößern (Bassères & Thevenot 2008). Dabei haben sie die Ausrichtung der Lautsprecher verändert. Anstatt die Lautsprecher L, R, LS und RS in die Mitte auszurichten, haben sie die Lautsprecher etwas zu der gegenüberliegenden Seite gedreht. Damit sollten die Lautstärkeunterschiede beim Verlassen des Sweet-Spots nach links und rechts kompensiert werden. Es entsteht eine

Linie orthogonal zum Lot des C-Lautsprechers, auf dem sich der/die Hörer*in bewegen kann.

2.6 Binaurales Audio

Während das Wort „binaural“ im grundlegenden Sinne nur beschreibt, dass mit zwei Ohren gehört wird, wird unter dem Begriff genauer verstanden, dass alle „spatial cues“, die natürlich durch den Aufbau der Ohren oder Kopf- und Körper-Reflexionen ausgelöst werden, im Audiosignal beinhaltet sind (Roginska & Geluso 2017, Kapitel 4). Dies umfasst Zeit-, Frequenz- und Intensitäts-Hinweise, welche dafür sorgen, dass der Sound über Kopfhörer möglichst real und natürlich klingt. Ein Mittel die Verzerrungen durch den Kopf, Körper und die Ohren im Signal zu verarbeiten ist die HRTF. Allerdings ist bekannt, dass HRTFs von Person zu Person sehr unterschiedlich sind und eine verzerrte Wahrnehmung auftreten kann, falls die HRTF synthetisch erstellt oder von einer anderen Person aufgenommen wurde (Fisher & Freedman 1968).

Binaurales Audio wird in der Regel über HP und nur in Ausnahmefällen über LS wiedergegeben. Der Vorteil von Sound über HP ist ein sehr kontrolliertes Umfeld. Das Hörerlebnis ist unabhängig von der Sitzposition im Raum und es entsteht kein Crosstalk der beiden Signale (das rechte Ohr bekommt nur das Signal des rechten Kanals), was bei der LS-Soundreproduktion durch Reflexionen nicht garantiert werden kann und wiederum zu einer Verwirrung der räumlichen Wahrnehmung führen kann (Roginska & Geluso 2017, Kapitel 4). Allerdings kommt es bei der Wiedergabe von Sound über HP in der Regel zur IKL, bei der die Schallquelle nicht an einer Position im Raum, sondern im Kopf der hörenden Person wahrgenommen wird. Dies wird dadurch ausgelöst, dass das Gehirn wahrgenommene Sounds miteinander vergleicht und Sound über Kopfhörer mit gleichen oder ähnlichen Informationen versehen werden muss, wie eine reale Schallquelle. Realistischer Hall oder individualisierte HRTFs können den Sound natürlicher klingen lassen und so die IKL minimieren (Roginska & Geluso 2017, Kapitel 4). Ein weiterer großer Faktor bei der Wiedergabe über Kopfhörer ist die Isolation von Geräuschen um den Hörer herum. Das kann in manchen Situationen zwar als Vorteil genutzt werden, ist aber ein sehr großer Nachteil, falls ein Sport-Event mit mehreren Personen geschaut wird. Die Kommunikation untereinander wird dadurch verhindert und es geht das Gemeinschaftsgefühl verloren. Eine mögliche Lösung dieses Problems ist, offene Kopfhörer zu benutzen. Dabei wird der 3D-Sound weiterhin binaural über die HP wiedergegeben, allerdings können so die Teilnehmer noch miteinander kommunizieren.

In einer Arbeit von Schneiderwindt et al. untersuchen diese den Effekt ver-

schiedener offener Kopfhörer auf die Wahrnehmung einer realen Schallquelle (Schneiderwind, Neidhardt & Meyer 2021). Dabei schneiden die extra-auralen Kopfhörer am besten ab, mit der geringsten Verzerrung. Diese haben allerdings keinen guten Komfort und sind überhaupt nicht verbreitet.

2.7 Overall Listening Experience

Die OLE ist ein Mittel, um eine musikalische Wiedergabe zu beurteilen. OLE versucht, alle möglichen Einflüsse auf ein Hörereignis mit in Betracht zu ziehen. Beispielsweise wird die Stimmung, in der sich der/die Proband*in befindet, oder die Qualität der abgespielten Musik berücksichtigt. Verbunden mit der OLE ist die Quality of Experience (QoE), welche wie folgt definiert ist:

“Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.” (Brunnström et al. 2013)

QoE ist also ein Maß, um die Freude oder den Ärger einer Person über eine Applikation oder einen Service festzustellen. Sie resultiert aus der Erfüllung von Erwartungen des/der Nutzer*in und berücksichtigt die Persönlichkeit und Stimmung des Users.

Kapitel 3

Methodik und Vorbereitung

Sportevents sind Veranstaltungen, welche sich Tausende von Zuschauer*innen in Stadien vor Ort anschauen. Auch zu Hause können Zuschauer*innen die von den Fans erzeugte Atmosphäre durch Audio- und Videoübertragungen erleben. Es gibt viele unterschiedliche Möglichkeiten diese Events zu verfolgen. Von der Smartphone-Wiedergabe mit Stereo-Audio bis hin zu immersiven 3D-Klangerlebnissen können Zuschauer*innen an dem Event teilhaben. Für ein 3D-Audio Erlebnis ist es nötig, dass Produzenten von solchen Events bzw. Sender bei der Übertragung eine solche Wiedergabemöglichkeit bereitstellen. Seit der Fußball Bundesliga Saison 2022/23 überträgt der Sender Sky wöchentlich ein Topspiel mit der Möglichkeit, dieses über Dolby Atmos immersiv zu verfolgen. Für diese Arbeit liegen die diskreten Audiosignale aus dem Stadion vor, welche der Sender Sky erhalten und als Dolby Atmos umwandeln würde. In dieser Arbeit werden verschiedene Wiedergabe-Systeme verglichen, sowie die binaurale Wiedergabe untersucht. Es wird auch die Sweet-Area der Wiedergabe-Systeme ermittelt, da dies bei der Übertragung für mehrere Personen wichtig wird. Der Schwerpunkt dieser Untersuchung liegt dabei auf der Immersivität, der Sound-Qualität, der Lokalisation und der Sweet-Area.

3.1 Methodik

3.1.1 Audio-Signal

Für die Untersuchung liegen 16 diskrete Audiospuren vom Spiel Eintracht Frankfurt gegen Borussia Dortmund am 29. Oktober 2022 vor, welche von Sportcast GmbH bereitgestellt wurden. Das Audio-Signal muss also noch zur 3D-Audio-Wiedergabe aufbereitet werden. In Tabelle 1.1 sind die Spuren aufgelistet. Da das Spiel in voller Länge mit Vor- und Nachlauf bereitgestellt wurde und somit

mehrere Stunden Signal und über 600 GB Daten umfasst, wurde in Premiere ein Zusammenschnitt von ausgewählten Highlights angefertigt und die Qualität des Videosignals von UHD auf HD reduziert.

3.1.2 Hörversuch

Es gibt drei Hörversuche. Bei dem ersten Hörversuch werden die drei Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out miteinander verglichen. Dabei werden Aspekte wie die Immersivität, Realität, Lokalisation und Sound-Qualität beachtet. Im zweiten Hörversuch wird dann die Lautsprecher-Wiedergabe mit der Kopfhörer-Wiedergabe verglichen. Dabei wird ebenfalls die Immersivität, Realität und Lokalisation untersucht und es kommen weitere Aspekte wie OLE, IKL und das Gemeinschaftsgefühl dazu. Im dritten Versuch werden die statische und mobile Audioperspektive angepasst an die Kameraperspektive miteinander verglichen. Sowohl in Dolby Atmos als auch in Ambisonics kann die allgemeine Richtung der abgemischten Szene angepasst und automatisiert werden. Diese Anpassung der Audioperspektive an das Kamerabild wird mit der klassisch statischen Audioperspektive, bei der der Sound unabhängig von der Kamera gleich bleibt, verglichen. Bei der statischen Perspektive ist der auditive Blickwinkel aus Sicht der Hauptkamera, welche auf der Tribüne an der Seitenlinie positioniert ist.

3.1.3 Bewertung der Sweet-Area

Vorbereitend auf die Hörversuche wird die Sweet-Area der Wiedergabe-Systeme Dolby Atmos und Ambisonics nach Frank und Zotter untersucht. Besonders für den Fall, dass mehrere Personen zuhören und nicht alle im Sweet-Spot sitzen können, ist eine große Sweet-Area sehr vorteilhaft.

3.1.4 Auswertung der Hörversuche

In der Auswertung werden die Ergebnisse der einzelnen Hörversuche nacheinander untersucht. Bei allen drei Hörversuchen wird mit Ausnahme der Präferenz-Fragen mit der Likert-Skala gearbeitet. Diese ist streng genommen ordinal skaliert, das heißt die Skalenniveaus können nur in eine Reihenfolge gebracht und einem Rang zugeordnet werden, können allerdings zwischen den Stufen nicht in Abständen gemessen werden. In der Praxis wird bei der Likert-Skala allerdings ein metrisches Skalenniveau angenommen, vor allem wenn es sich bei den Stufen um die gängigen Bezeichnungen handelt (sehr, eher, weder noch, eher nicht und gar nicht) (*Likert-Skala | Definition, Anwendung und*

Auswertung 2023). In diesem Fall können den Stufen „gar nicht“ bis „sehr“ die Zahlen eins bis fünf zugeordnet werden. Somit lassen sich Mittelwerte und Standardabweichungen berechnen und statistische Analysen durchführen. Für jede Frage wird den Stufen der Likert-Skala die Zahlen eins bis fünf zugeordnet, wobei die Zahl fünf einer guten Bewertung entspricht. Nur bei der Frage sieben (HV 2) und den Fragen vier und fünf (HV 3) ist die Frage negativ konnotiert, so dass ein hoher Balken einem schlechten Ergebnis entspricht.

Um eine statistische Signifikanz zu prüfen, wird bei HV 1 eine einfaktorielle Varianzanalyse (Analysis of Variance (ANOVA)) durchgeführt. Die Varianzanalyse ist einfaktoriell, weil es sich um eine unabhängige Variable handelt. Da es bei dem ersten Hörversuch mehr als zwei Gruppen gibt (Dolby Atmos, Ambisonics und Direct out), kann ein einfacher t-Test nicht angewandt werden. Zu prüfen ist bei der ANOVA, ob (i) die unabhängige Variable ein kategoriales Skalenniveau besitzt, (ii) die abhängige Variable Kardinalskalenniveau besitzt, (iii) Varianzhomogenität vorhanden ist, (iv) Normalverteilung gilt und (v) die abhängige Variable möglichst keine Extremwerte enthält. Bei dem zweiten Hörversuch wird über den t-Test bei abhängigen Stichproben die statistische Signifikanz getestet. Hier müssen dieselben Voraussetzungen geprüft werden, da es aber nur zwei Gruppen gibt (HP und LS), kann der einfache t-Test verwendet werden. Beide Tests sind dazu da, Mittelwerte von verschiedenen Gruppen zu vergleichen und herauszufinden, ob ein signifikanter Unterschied vorliegt. Zur Auswertung wird die Datenanalyse in Excel ausgeführt. Da in HV 3 nur der auditive Perspektivenwechsel untersucht wird, kann hier kein Mittelwertvergleich durchgeführt werden. Die Ergebnisse werden dennoch aufgearbeitet und diskutiert.

3.2 Vorbereitung

3.2.1 Raum

Die Hörversuche werden im Tonlabor der HAW Hamburg Fakultät DMI Finkenau durchgeführt. Dort gibt es einen großen Aufnahmeraum, in dem ein 3D-Audio Dome installiert ist. Dieser hat 38 Lautsprecher auf sechs Ebenen und zwei Subwoofer (siehe Abbildung 3.1 und 3.2).

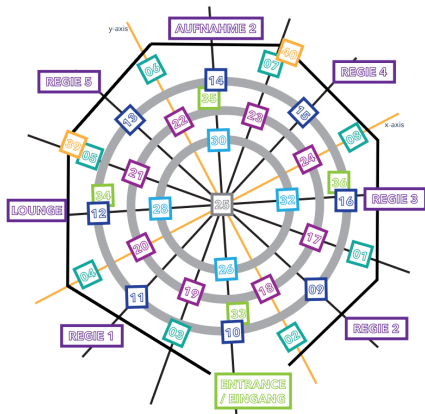


Abbildung 3.1: Die Lautsprecher des Hörversuchsraums mit Nummerierung

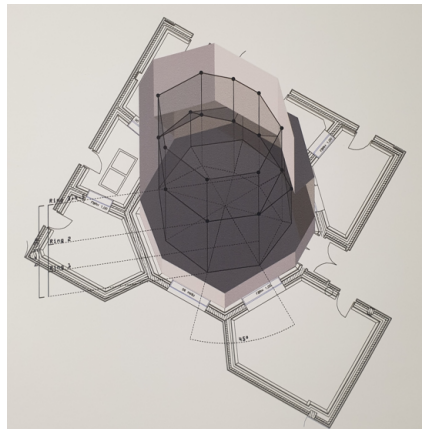


Abbildung 3.2: Eine 3D-Grafik der Lautsprecheranordnung im Hörversuchsraum

Der Aufnahmeraum hat ein Volumen von $V = 474m^3$ und hat die frequenzabhängigen Nachhallzeiten wie in Abbildung 3.3. Er eignet sich mit seiner Größe

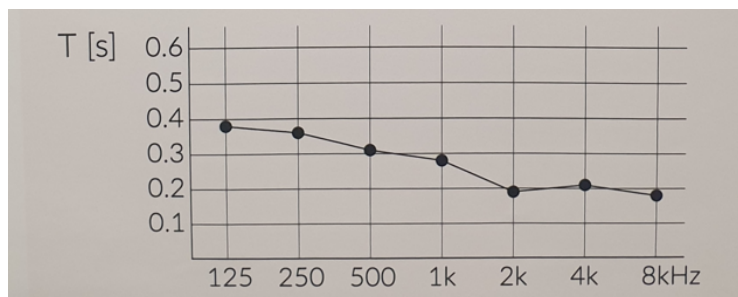


Abbildung 3.3: Frequenzabhängige Nachhallzeiten des Hörversuchsraums

und Akustik sehr gut für die Umsetzung dieser Arbeit. Während viele Wohnzimmer in der Regel kleiner sind und viele Sports-Bars größer, befindet sich der Raum mit $A = 69,5m^2$ Grundfläche im durchschnittlichen Bereich. Die Akustik im Raum ist sehr viel besser als die in den meisten Wohnzimmern oder in einer Sports-Bar, allerdings eignet sie sich dadurch gut für Hörtests, bei denen die akustische Verfälschung durch Reflexionen vermieden werden muss.

3.2.2 Lautsprecher-Setup

Um die Hörversuche dieser Untersuchung vorzubereiten, werden als Erstes die verschiedenen Wiedergabe-Systeme eingerichtet. Dabei wird Dolby Atmos und Direct Out über das Lautsprecher-Setup 5.1.4 mit 5.1 Surround und vier weiteren Deckenlautsprechern wiedergegeben (Abbildung 3.4). Ambisonics ist nicht Lautsprechergebunden und wird per 11.1 wiedergegeben, bei dem zwei Ebe-

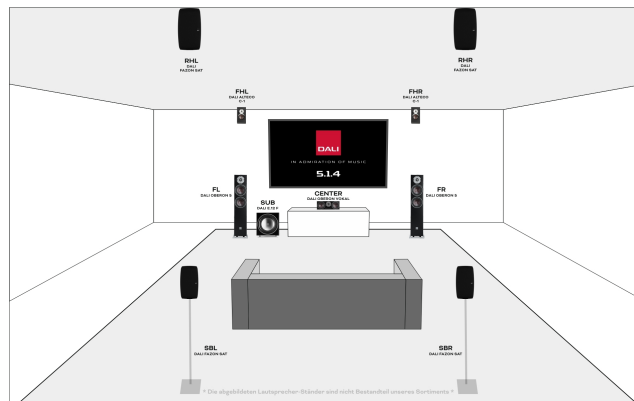


Abbildung 3.4: Lautsprecher-Setup 5.1.4 für Dolby Atmos und Direct Out

nen mit je fünf Lautsprechern übereinander nach dem System 5.1 aufgebaut werden (L, C, R, LS, RS). Dazu strahlt ein Top (T)-Lautsprecher von oben ab, dieser hängt direkt über dem Sweet-Spot in der Mitte (Abbildung 3.5).

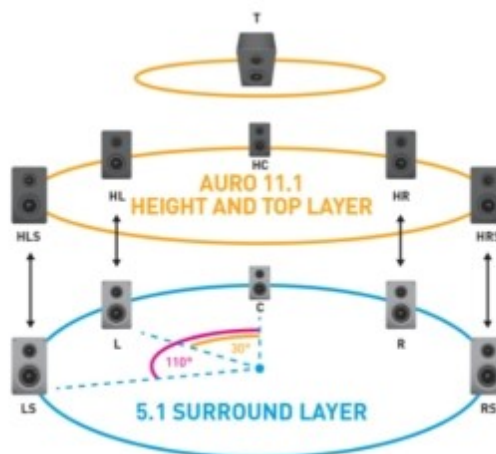


Abbildung 3.5: Lautsprecher-Setup 11.1 für Ambisonics

Von den 40 fest installierten Lautsprechern werden die Lautsprecher aus Tabelle 3.1 benutzt.

Die beiden Formate 5.1.4 und 11.1 haben zwar Subwoofer integriert, da allerdings die Low-Frequency-Emitter (LFE)-Spur des bereitgestellten Audio-Materials leer ist, werden die Subs nicht verwendet. Der Einfachheit halber werden sie im Folgenden dennoch mitgenannt.

Die Lautsprecher sind über Dante mit dem 3D-Audio Computer verbunden. über die Dante Virtual Soundcard werden die Dante-Ausgänge in Reaper mit dem DADman AD/DA converter verbunden. Dieser ist in der Signalkette allerdings nur als Monitoring-System in Verwendung. Angeschlossen an den DAD-

Abk.	Lautsprecher	Nummer
C*	Center	6
L*	Left	5
R*	Right	7
LS*	Left-Surround	3
RS*	Right-Surround	1
HLF*	Height-Left-Front	21
HRF*	Height-Right-Front	23
HLR*	Height-Left-Rear	19
HRR*	Height-Right-Rear	17
CH**	Center-Height	22
T**	Top	25

Tabelle 3.1: Auswahl der verwendeten elf Lautsprecher mit Kanal-Zuordnung
 * Lautsprecher wird bei der Wiedergabe von Dolby Atmos, Ambisonics und Direct Out verwendet

** Lautsprecher wird nur bei der Wiedergabe von Ambisonics verwendet

man ist das Monitoring Output Module (MOM), um gesamtheitlich die Lautstärke der Lautsprecher ändern zu können. Der DADman ist dann mit den Yamaha NIO Nuage Dante Interfaces verbunden. Hier wird das Signal dann A/D gewandelt und an die Lautsprecher gegeben. Über diese Signalkette werden die Spuren in Reaper mit den Lautsprechern verbunden.

3.2.3 Software und Plug-Ins

Das Wiedergabe-System Dolby Atmos wird über Nuendo 12 und das dort integrierte Audio Authoring laufen. Direct Out und Ambisonics werden über Reaper abgespielt. Reaper eignet sich sehr gut für die 3D-Audiowiedergabe, da es pro Kanal bis zu 64 Tracks erlaubt und somit bis hin zu 7th-Order Ambisonics wiedergeben kann. Außerdem hat Reaper eine 60-Tage Testversion, welche auch nach den 60 Tagen noch weiter evaluiert werden kann. Dies ermöglicht eine kostenfreie Nutzung des Systems. Die Dolby Atmos Plugin Suit ist allerdings nicht mit Reaper kompatibel. Dolby Atmos muss also über eine andere Digital Audio Workstation (DAW) wiedergegeben werden. Dabei bieten sich zwei Varianten an: Über MacOS kann die Dolby Atmos Plugin Suite mit Logic Pro ausgeführt werden und über Nuendo 12 kann die integrierte Audio-Authoring Funktion ausgeführt werden. In dieser Arbeit wird Dolby Atmos über eine 60-Tage Testversion von Nuendo 12 abgespielt.

Dolby Atmos wird in Nuendo über den ‚ADM Authoring Assistant‘ eingerichtet. Hier werden der Renderer mit dem jeweiligen Lautsprecher-Format (5.1.4, 5.1.7, etc.) und das Standard Bed (2.0, 5.1, 7.1.2, etc.) ausgewählt. Über das ADM Authoring können auch Objekte hinzugefügt werden und Spuren dem Bed zugeordnet werden. Mit dem Multi-Panner der einzelnen Spuren kann das Audio-

Signal der jeweiligen Spur beliebig drei-dimensional im Raum verschoben werden.

Ambisonics wird mit der IEM Plug-In Suite umgesetzt. Mit dieser können die einzelnen Spuren über den StereoEncoder beziehungsweise den Multi-Encoder im Raum verteilt werden. Über den AllRA Decoder wird das LS-Setup eingerichtet und die Kanalverteilung definiert. Über Azimuth- und Elevationswinkel kann die Position der Lautsprecher angepasst werden. Für die Lautsprecher im Raum liegt eine Tabelle vor, in der von der Hörposition ausgehend die Entfernungen der Lautsprecher im kartesischen Koordinatensystem angegeben sind, sowie die Azimuth- und Elevationswinkel (Abbildung 3.6). Diese können von der Ta-

Interface	CH	Levels	Y Axis	X Axis	Height	Azimuth	Elevation	DTSS*
NIO I - 1	1	E1	2.70	-2.94	-0.02	222.49	-0.24	3.99
NIO I - 2	2	E1	-0.33	-4.51	-0.05	175.77	-0.59	4.52
NIO I - 3	3	E1	-3.20	-2.79	-0.08	131.16	-1.07	4.25
NIO I - 4	4	E1	-4.31	0.22	-0.03	87.13	-0.42	4.31
NIO I - 5	5	E1	-2.85	2.94	-0.06	44.14	-0.87	4.10
NIO I - 6	6	E1	0.00	4.18	-0.08	359.97	-1.14	4.18
NIO I - 7	7	E1	2.73	2.80	-0.09	315.76	-1.26	3.91
NIO I - 8	8	E1	3.83	-0.05	-0.09	269.33	-1.31	3.83

Abbildung 3.6: Ein Ausschnitt der Lautsprecher-Tabelle mit den Winkelabmessungen und Abständen zum Mittelpunkt.

*DTSS = Distance to Sound Source

belle für den AllRA Decoder übernommen werden, allerdings müssen Winkel zwischen 180° und 360° umgerechnet werden auf -180° bis 0°. Nach dem AllRA Decoder folgt der IEM DistanceCompensator, welcher die Distanzen zu den Lautsprechern berücksichtigt und den Pegel anpasst. Die Entfernungen zu den jeweiligen Lautsprechern können ebenfalls der Tabelle entnommen werden. Optional kann vor dem AllRA Decoder auch der EnergyVisualizer von IEM hinzugefügt werden (Abbildung 3.11). Dieses Plug-In ist dazu da, die Richtung, aus der das Audiosignal kommt, zu visualisieren. Über den Fader „Peak level“ kann die Empfindlichkeit eingestellt werden, um die Richtung des Signals deutlicher zu machen.

3.2.4 Setup des Wiedergabe-Systems Dolby Atmos

Dolby Atmos wird, wie bereits erwähnt, über Nuendo wiedergegeben. In einem neuen Projekt wird über die Funktion „ADM Authoring for Dolby Atmos“ eingerichtet (Abbildung 3.7).

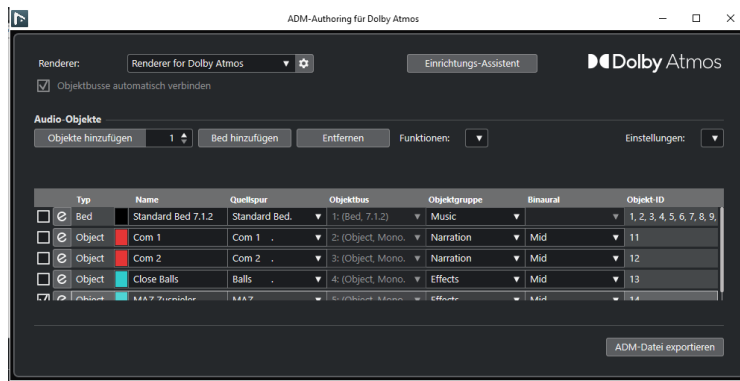


Abbildung 3.7: ADM Authoring für Dolby Atmos in Nuendo 12. Hierüber werden die allgemeinen Atmos-Einstellungen vorgenommen sowie Beds und Objects verwaltet.

Über den Einrichtungs-Assistenten kann das Wiedergabe-Format 5.1.4 ausgewählt und ein Bed-Kanal im Mixer angelegt werden (Abbildung 3.8). Als Bed wird

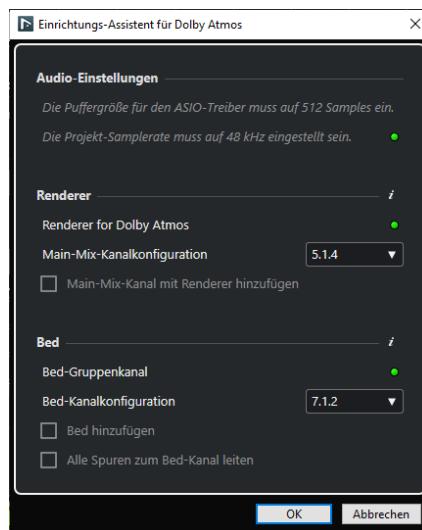


Abbildung 3.8: Über den Dolby Atmos Setup Assistant wird das Lautsprecherformat 5.1.4 ausgewählt und der Bed-Kanal dem Mixer hinzugefügt. Als Bed wird das Format 7.1.2 verwendet, welches auf das Setup 5.1.4 umgerechnet wird.

das Format 7.1.2 ausgewählt, da es das Format 5.1.2 nicht gibt, allerdings kann der Dolby Atmos Renderer das 7.1.2 Bed auf das 5.1.2 Wiedergabe-Format umrechnen. Die Spuren eins bis zehn werden im Bed wiedergegeben. Die Spuren elf und zwölf mit den deutschen und englischen Kommentatoren werden als Objekte hinzugefügt. Als Objekte können sie frei innerhalb des Setups bewegt werden.

Das Routing in Nuendo 12 ist vom Kanal zu den Lautsprechern folgendermaßen eingerichtet: Alle Spuren, welche in dem Bed angelagert sind, werden über

den von dem ADM Authoring Setup Assistenten angelegten Kanal „Standar-Bed“ geleitet. Dazu wird das Signal jeder Bed-Spur über den Bus zum Standard-Bed geroutet. Der Standard-Bed-Kanal wird über direktes Routing dem 5.1.4 Renderer hinzugefügt. Objektsuren werden direkt in die 5.1.4 Renderer-Spur geroutet. Ist dieses Routing nicht vorhanden, werden die Spuren der Objekte nicht im ADM-Monitoring erkannt und können nicht dem Renderer hinzugefügt werden. Objekte können nach dem eingestellten Routing über das ADM-Monitoring dem 5.1.4 Renderer hinzugefügt werden. Im Mixer werden die Spuren dann über Sends zum 5.1.4 Renderer geleitet. Im 5.1.4 Renderer werden dann die verschiedenen Outputs (Center, Left, Right, etc.) den entsprechenden Dante Outputs zugewiesen, welche der Nummerierung der Lautsprecher entspricht (1-64 auf 1-64). Über Dante wird das Signal dann (wie in Kapitel 3.2.2) an die Lautsprecher weitergegeben.

3.2.5 Setup des Wiedergabe-Systems Ambisonics

Ambisonics wird über die DAW Reaper wiedergegeben. Hier werden verschiedene „Parent“-Spuren angelegt. Die Parent-Spuren sind ähnlich wie eine „Mute“-Gruppe zu verstehen. Alle Einzelspuren können über die Gruppenspur mit einem Klick gemutet werden. Es gibt die beiden folgenden Gruppenspuren Ambisonics mit IEM En- und Decoder und direktes Routing der Einzelspuren zu den Lautsprechern ohne Signalverarbeitung (Direct Out). Der Ambisonics Gruppenspur mit IEM Plug-Ins wird der AllRAD Decoder mit dem entsprechenden 11.1 Lautsprecher-Setup hinzugefügt (Abbildung 3.9). Dieser dekodiert das erhalte-

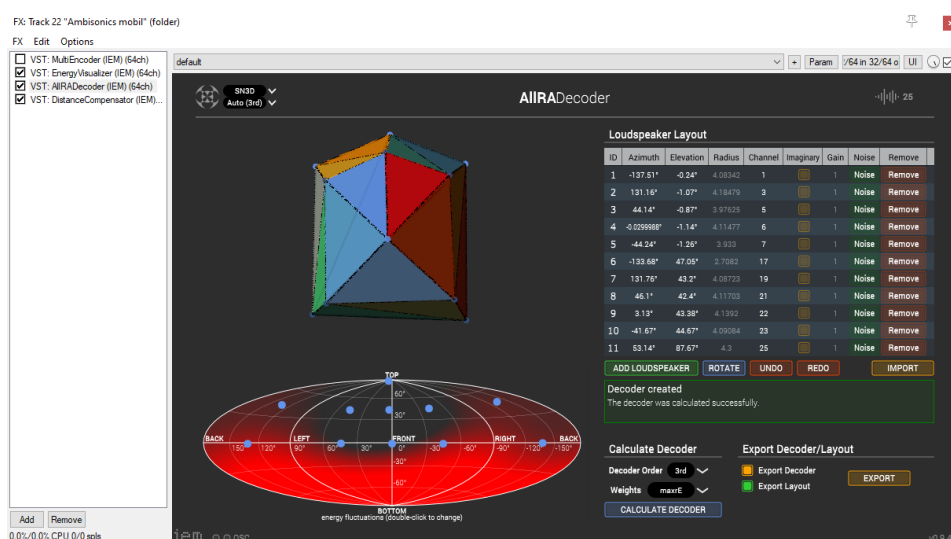


Abbildung 3.9: Der AllRAD Decoder ist dazu da, das Lautsprecher-Setup in Reaper einzurichten.

ne Ambisonics-Signal und gibt die entsprechenden Informationen an die jeweiligen Lautsprecher weiter. Auf den Einzelspuren mit den Audiosignalen liegt der MultiEncoder (Abbildung 3.10). Dieses Plug-In ist dazu da, das Signal an die ent-

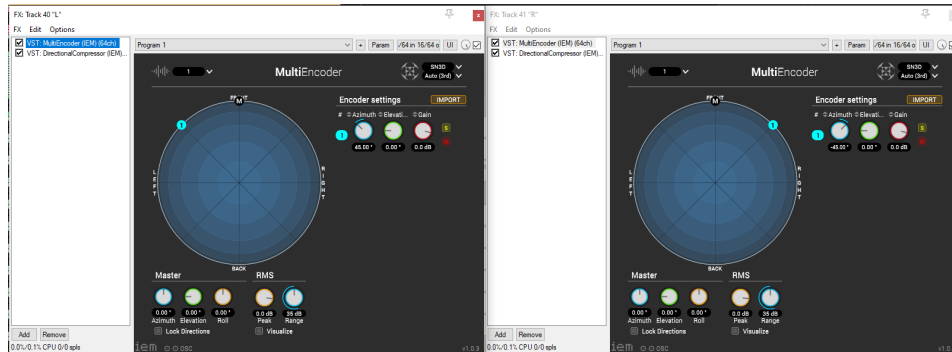


Abbildung 3.10: Der IEM MultiEncoder ist dazu da, eine einzelne Spur an einer bestimmten Position im Lautsprecher-Setup zu platzieren. Beispielhaft sind die Kanäle Left und Right abgebildet.

sprechende Position im Raum zu legen. Sowohl die Parent-Spur, als auch die Einzelspuren müssen im Routing mindestens mit 16 Tracks eingestellt werden, damit die Spur als 3rd-Order Ambisonics kodiert werden kann. Der MultiEncoder funktioniert so, dass deutlich lokalisiert werden kann, aus welcher Richtung ein Geräusch oder Signal kommt. Über den EnergyVisualizer wird diese Richtung ebenfalls deutlich sichtbar (Abbildung 3.11). Allerdings wird das Signal bei-

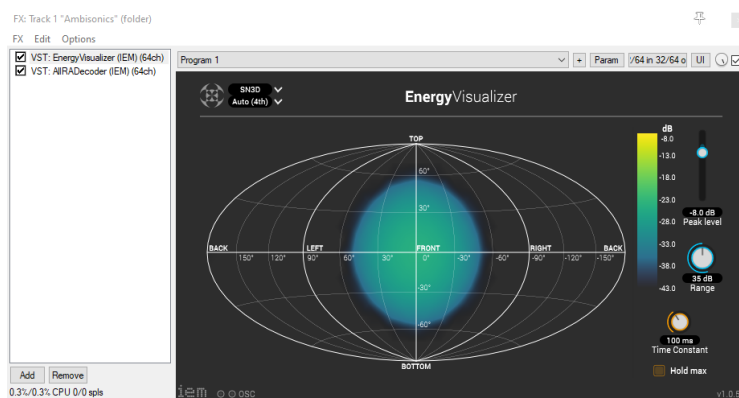


Abbildung 3.11: Der EnergyVisualizer zeigt an, aus welcher Richtung das Signal zu hören sein sollte.

spielsweise im Fall, dass es von vorne kommt, nicht nur aus dem C-Lautsprecher wiedergegeben, sondern es sind auch aus anderen Lautsprechern Signale zu hören. Dies kann zu einem ungewollten LS-übergreifenden „bleed“ führen. Werden dementsprechend korrelierte Signale an die Lautsprecher weitergegeben, kann es zu ungewollten Auslöschungen kommen.

3.2.6 Setup des Wiedergabe-Systems Direct Out

Es wird ebenfalls ein direktes Routing der Einzelspuren zu den Lautsprechern vorgenommen, bei der sämtliche Plug-Ins umgangen werden (Direct Out). Dies bietet den Vorteil, bei der Wiedergabe in einer Sports-Bar die erhaltenen Spuren direkt auf die Lautsprecher durchzuschleifen und keine weitere Signalverarbeitung vornehmen zu müssen. Die Kanäle eins bis zehn werden intuitiv, so wie sie benannt sind, an die entsprechenden LS geroutet (Center-Center, Left-Left, etc.). Die Audio-Spuren der Kommentatoren werden ebenfalls zu dem Center-Lautsprecher geroutet. Angelehnt an die Audio-Wiedergabe im Kino, bei der der Dialog auch im C-Kanal lokalisiert ist, wird hier der Kommentar ebenfalls von vorne abgespielt. Die beiden Spuren „Fans-Heim“ und „Fans-Auswärts“ werden dem Stadion entsprechend zu den beiden Lautsprechern L und R geroutet (Fans-Auswärts – Links und Fans-Heim – Rechts). Dies kann sich von Stadion zu Stadion allerdings unterscheiden und muss dem Spiel entsprechend angepasst werden. Allerdings ist es fast ausnahmslos der Fall, dass die beide Fangemeinschaften im Stadion gegenüberliegend platziert sind und so eine Trennung in Left (L) und Right (R) immer vorgenommen werden kann. Die beiden restlichen Spuren „Close-Balls FX“ und „PA-Zuspieler“ werden ebenfalls zu dem C-Kanal geroutet. Der LFE-Kanal ist zwar leer, kann aber auch direkt zu dem entsprechenden Lautsprecher geroutet werden.

3.2.7 Vorbereitung Hörversuch eins (HV 1): Vergleich der Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out

Bei dem ersten Hörversuch werden die verschiedenen Wiedergabe-Formate Dolby Atmos, Ambisonics und Direct Out miteinander verglichen. Dabei werden Aspekte wie Immersivität, Lokalisation, Klangfarbe und Realität untersucht. Die Fragen eins bis fünf dienen dazu, die Immersivität der Wiedergabe-Systeme zu ermitteln. Die ersten drei Fragen werden zu den Adjektiven einschließend, umfassend und lebendig gestellt, da durch den Grad dieser Eigenschaften der Grad der Immersivität beurteilt wird. Frage vier und fünf zielen auf die Realität beziehungsweise Illusion der Realität ab. Mit Frage sechs wird der Frequenzbereich untersucht und mit Frage sieben die Lokalisation. Bei Frage sechs kann außerdem über ein leeres Feld der Bereich angegeben werden, welcher nicht gut wiedergegeben wird. Frage acht untersucht die Homogenität der Soundverteilung auf die Lautsprecher. Mit der letzten Frage muss der/die Proband*in dann angeben, welches Wiedergabe-System bevorzugt wird. Die Leitfragen des ersten Hörversuchs sind, ob sich die Wiedergabe-Systeme signifikant voneinander unterscheiden und welches der drei Systeme bevorzugt wird.

Im Folgenden werden die Fragen des HV 1 aufgelistet und mit den Antwortmöglichkeiten in den Klammern versehen:

1. Wie einschließend ist der Sound dieser Wiedergabe? (sehr einschließend, eher einschließend, weder noch, eher nicht einschließend, gar nicht einschließend)
2. Wie umfassend ist der Sound dieser Wiedergabe? (sehr umfassend, eher umfassend, weder noch, eher nicht umfassend, gar nicht umfassend)
3. Wie lebendig ist der Sound dieser Wiedergabe? (sehr lebendig, eher lebendig, weder noch, eher nicht lebendig, gar nicht lebendig)
4. Wie erfolgreich wird auditiv die Stadionatmosphäre vermittelt? (sehr erfolgreich, eher erfolgreich, weder noch, eher nicht erfolgreich, gar nicht erfolgreich)
5. Wie gut wird auditiv die Illusion der Realität wiedergegeben? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
6. Wie gut werden alle Frequenzbereiche wiedergegeben? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
7. Wie gut können einzelne Sounds lokalisiert werden? (Fangesänge, Ball-Sounds, ...) (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
8. Wie gut ist der Sound auf die verschiedenen Lautsprecher verteilt? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
9. Welches der Wiedergabe-Formate bevorzugen Sie?

3.2.8 Vorbereitung Hörversuch zwei (HV 2): Single-Listener Experience – LS vs. HP

Bei dem zweiten Hörversuch wird die Wiedergabe über HP mit der Wiedergabe über LS verglichen. Verwendet wird dazu Dolby Atmos 5.1.4 über LS und Dolby Atmos binaural über HP. Auch hier wird wie bei HV 1 die Immersivität, Lokalisation, Klangfarbe und Realität untersucht. Dazu kommen Aspekte wie OLE, IKL und das Gemeinschaftsgefühl bei der HP-Wiedergabe, falls ein Spiel zusammen geschaut wird. Leitfragen bei HV 2 sind, ob die Wiedergabe über HP genauso immersiv oder immersiver ist, als die Wiedergabe über LS und welche Wiedergabe von den Proband*innen bevorzugt wird. Es wird auch untersucht, was der/die Proband*in zu bezahlen bereit ist, um sich ein 3D-Sound bei einer

Sportübertragung anzuhören (Eintritt in einer Sports-Bar oder ein Heimkino mit 3D-Soundanlage).

Bei Frage eins und zwei wird die OLE beurteilt. Es wird damit allgemein untersucht, ob der Sound den Proband*innen gefallen hat und ob sie genervt oder erfreut über den Sound waren. Frage drei und vier beziehen sich auf die Immersivität und Illusion der Realität der Wiedergaben. Auch Fragen sechs und acht hinterfragen, wie realistisch die Wiedergabe wirkt. Frage fünf untersucht die Wiedergabe der Frequenzbereiche und Frage sieben die IKL. Da es auch bei binauralen Wiedergaben zur IKL kommen kann (Roginska & Geluso 2017, S. 106), wird mit dieser Frage untersucht, inwiefern eine binaurale Wiedergabe dieses Audiomaterials eine IKL erzeugt. Bei Frage elf muss eine Präferenz des Wiedergabe-Formats abgegeben werden. Fragen neun, zehn, zwölf, und vierzehn untersuchen die Wiedergabe von 3D-Audio zu Hause im Wohnzimmer und Frage dreizehn in einer Sports-Bar.

Im Folgenden werden die Fragen des HV 2 aufgelistet und mit den Antwortmöglichkeiten in den Klammern versehen:

1. Wie gut gefällt Ihnen der Sound? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
2. Wie erfreut oder genervt sind Sie von dem Sound während der Wiedergabe? (sehr erfreut, eher erfreut, weder noch, eher genervt, sehr genervt)
3. Wie immersiv ist die Audio-Wiedergabe? (einschließend, umfassend, lebendig, intensiv) – (sehr immersiv, eher immersiv, weder noch, eher nicht immersiv, gar nicht immersiv)
4. Wie gut wird auditiv die Illusion der Realität wiedergegeben? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
5. Wie gut werden alle Frequenzbereiche wiedergegeben? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
6. Wie groß oder mächtig wirkt die Wiedergabe auf Sie? (sehr groß, eher groß, weder noch, eher nicht groß, gar nicht groß)
7. Wie stark kommt es zur Im-Kopf-Lokalisation? (Hörereignisort nicht außerhalb vom Kopf in der Umgebung, sondern Orte von Schallquellen scheinen im Kopf zu liegen) – (NUR BEI KOPFHÖRERN) – (sehr stark, eher stark, weder noch, eher nicht stark, gar nicht stark)
8. Wie erfolgreich wird die Stadionatmosphäre vermittelt? (sehr erfolgreich, eher erfolgreich, weder noch, eher nicht erfolgreich, gar nicht erfolgreich)

9. Wie gerne würden Sie zu Hause eine Sportübertragung mit 3D-Audio hören? (sehr gerne, eher gerne, weder noch, eher nicht gerne, gar nicht gerne)
10. Wie gerne würden Sie zu Hause mit Freunden zusammen eine Sportübertragung per Kopfhörer hören? (sehr gerne, eher gerne, weder noch, eher nicht gerne, gar nicht gerne)
11. Welche Wiedergabe bevorzugen Sie? (HP oder LS)
12. Wie wichtig ist Ihnen eine gute Soundanlage zu Hause? (sehr wichtig, eher wichtig, weder noch, eher nicht wichtig, gar nicht wichtig)
13. Wie gerne würden Sie für ein besseres Hörerlebnis Geld ausgeben? (z.B. 5€ Eintritt bei einer Sportsbar mit 3D-Sound) – (sehr gerne, eher gerne, weder noch, eher nicht gerne, gar nicht gerne)
14. Wie wahrscheinlich ist es, dass Sie sich für ein besseres Hörerlebnis ein Heimkino mit 3D-Audio einrichten? (3D-Soundanlage mit acht oder mehr Lautsprechern) – (sehr wahrscheinlich, eher wahrscheinlich, weder noch, eher nicht wahrscheinlich, gar nicht wahrscheinlich)

3.2.9 Vorbereitung Hörversuch drei (HV 3): Vergleich von zwei Audioperspektiven – statisch und mobil

Bei HV 3 (Audioperspektive) werden zwei unterschiedliche Wiedergabe-Methoden miteinander verglichen. Neben der herkömmlichen statischen Audioperspektive wird eine zweite mobile Audioperspektive abgemischt. Bei der statischen Audioperspektive ist die Stadionatmosphäre gleichbleibend. Das heißt, dass beispielsweise die Auswärts-Fans immer links und die Heimfans immer rechts zu hören sind, unabhängig von der Kameraperspektive. Bei der mobilen Audioperspektive wird die Stadionatmosphäre so gedreht, dass die auditive Szene dem Kamerabild angepasst ist. Wenn die Auswärts-Fans sich beispielsweise in der Mitte des Kamerabildes befinden, so sind diese in der Mitte im C-Kanal zu hören und wenn sie sich hinter der Kamera befinden und nicht im Kamerabild sind, dann sind sie auch von hinten zu hören. Dies wird über Dolby Atmos mit dem 3D-Panner von Nuendo 12 realisiert, indem die gesamte Audio-Szene rotiert wird. Dies lässt sich per Automation speichern und kann über die ‚read‘-Funktion aktiviert werden. Die Leitfragen in HV 3 untersuchen, ob durch den auditiven Perspektivenwechsel die Immersivität bei dem/der Zuhörer*in gesteigert wird oder ob der zusätzliche auditive Perspektivenwechsel den/die Zuhörer*in verwirrt.

Nachdem der/die Proband*in beide Audioperspektiven gehört hat, wird mit der ersten Frage getestet, ob ein auditiver Perspektivenwechsel bemerkt wurde. Fragen zwei und vier dienen dazu, diesen Perspektivenwechsel zu bewerten. Bei der Frage drei wird eingeschätzt, ob durch den Perspektivenwechsel die Immersivität gesteigert wird. Bei Frage fünf soll der/die Proband*in abschätzen, ob ein auditiver Perspektivenwechsel bei einem Sportevent, welches 90 Minuten lang übertragen wird, auf die Dauer als störend wahrgenommen werden kann. Da bei dem Hörversuch nur ein Ausschnitt mit einer Dauer von einer Minute zu hören ist, soll damit getestet werden, ob die Wiedergabe 90 Minuten lang vorstellbar ist oder doch nur bei drei- bis fünf-minütigen Highlights benutzt werden sollte. Bei Frage sechs muss angekreuzt werden, welches Format bevorzugt wird.

Im Folgenden werden die Fragen des HV 3 aufgelistet und mit den Antwortmöglichkeiten in den Klammern versehen:

1. Wie stark merken Sie einen auditiven Perspektivwechsel? (sehr stark, eher stark, weder noch, eher nicht stark, gar nicht stark)
2. Wie gut finden Sie den Perspektivwechsel? (sehr gut, eher gut, weder noch, eher nicht gut, gar nicht gut)
3. Wie stark wird die Immersivität durch den Perspektivwechsel gesteigert? (sehr stark, eher stark, weder noch, eher nicht stark, gar nicht stark)
4. Wie verwirrend fanden Sie den Perspektivwechsel? (sehr verwirrend, eher verwirrend, weder noch, eher nicht verwirrend, gar nicht verwirrend)
5. Wie nervig kann das Hörerlebnis werden, wenn die auditive Perspektive 90 Minuten an das Kamerabild angepasst ist? (sehr nervig, eher nervig, weder noch, eher nicht nervig, gar nicht nervig)
6. Welche Audioperspektive bevorzugen Sie? (statisch, mobil)

3.2.10 Bewertung der Sweet-Area

Die Formate Dolby Atmos und Ambisonics werden nach der Arbeit von Frank und Zotter auf ihren Sweet-Spot hin untersucht (Frank & Zotter 2017). Die Untersuchung der Sweet-Area bei Direct Out ist nicht möglich, da hier der C-Lautsprecher eine Punktschallquelle bildet und so die Kriterien des Versuchs nicht mehr gegeben sind (dazu später mehr). Erst durch das Zusammenwirken von mehreren Kanälen wird bei Direct Out aus der Audio-Szene ein immersives Erlebnis.

Wie in Kapitel 2.5 beschrieben, wird die Sweet-Area der beiden Wiedergabe-Systeme ermittelt. Dabei wird ein neues Reaper-Projekt für Ambisonics und ein neues Nuendo-Projekt für Dolby Atmos mit demselben Routing erstellt. Für beide Formate wird ein Testsignal der EBU SQAM CD, welche „Sound Quality Assessment Material Recordings“ für subjektive Hörtests bereit stellen, verwendet (*Sound Quality Assessment Material recordings for subjective tests | EBU Technology & Innovation 2008*). Verwendet wird das Testsignal Nummer 54 „Male Speech German“ in Mono. Dieses Signal wird bei beiden Formaten in den C-Kanal gelegt. Bei Ambisonics geschieht dies über den MultiEncoder und in Nuendo 12 wird dies über den Multi-Panner realisiert. Bei beiden Formaten gibt es Informationen für die restlichen Kanäle, sodass das Signal nicht nur aus dem C-Lautsprecher zu hören ist (vgl. Abbildung 3.11). Es gibt acht Linien vom Mittelpunkt des Raumes in die acht Richtungen direkt zu den Lautsprechern auf Kopfhöhe und drei zusätzlichen Richtungen mit ungefähr 45° Abständen (Tabelle 3.2). Auf diesen Linien wird vom Mittelpunkt zu den jeweiligen Lautsprechern

		Ambisonics in m	Dolby Atmos in m	Abstand zum Lautsprecher in m	Ambisonics in %	Dolby Atmos in %
C	0°	1,75	1,00	4,12	43	24
L	45°	2,50	2,37	3,98	63	60
R	315°	1,75	1,60	3,93	44	41
LS	135°	0,81	2,58	4,19	19	62
RS	225°	2,24	2,48	4,08	55	61
Hinten	180°	1,20	2,96	4,53	27	65
Links	90°	1,10	2,72	4,19	26	65
Rechts	270°	1,72	2,96	3,93	44	75

Tabelle 3.2: Winkelabmessungen der acht Richtungen mit den Abständen der Lautsprecher zum Mittelpunkt, sowie der gemessenen Sweet-Area-Grenzen

beziehungsweise Positionen die Plausibilität der Audio-Szene geprüft. Das heißt, es wird evaluiert, ob (a) die Soundquelle nicht mehr zwischen den Lautsprechern L-C-R lokalisiert wird, oder (b) die Lokalisation nur noch bei einem einzelnen Lautsprecher statt findet. Frank und Zotter stellen in ihrer Arbeit ein drittes Kriterium, welches sich auf die Halligkeit bezieht, was in diesem Versuch allerdings ignoriert werden kann. Wichtig ist zu beachten, dass während die Person sich auf einer Linie zu einer LS-Position hin bewegt, immer der C-Lautsprecher angeschaut werden muss. Der Entfernung vom Mittelpunkt aus, ab der eines der beiden Kriterien nicht mehr zutrifft, wird gemessen und aufgeschrieben. Anhand der acht Entfernungen für die verschiedenen Richtungen kann die Sweet-Area visualisiert werden.

In den Abbildungen 3.12 und 3.13 ist deutlich zu sehen, dass die Sweet-Area

Sweet-Area Assessment bei Dolby Atmos

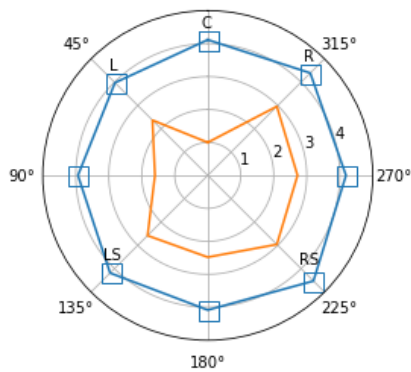


Abbildung 3.12: Die ermittelte Sweet-Area für Dolby Atmos

Sweet-Area Assessment bei Ambisonics

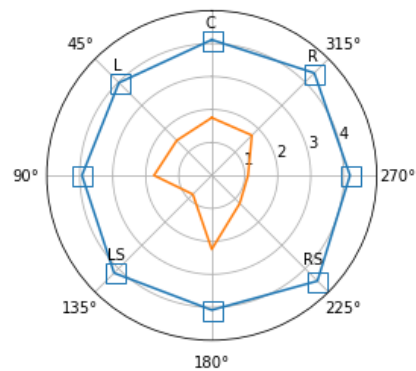


Abbildung 3.13: Die ermittelte Sweet-Area für Ambisonics

bei Dolby Atmos deutlich größer ist, als die bei Ambisonics. Vor allem nach hinten in Richtung 180° ist der Bereich der Sweet-Area, in der sich Personen hinsetzen können, deutlich größer. Gerade im Wohnzimmer wird dieser vergrößerte Bereich zu einem deutlichen Vorteil, da dort oft eine Couch oder andere Sitzgelegenheiten an der Wand und nicht mitten im Raum platziert sind. Der vergrößerte Bereich bei Dolby Atmos bietet auch in Sports-Bars einen deutlichen Vorteil, da hier mehr Stühle oder Sitzgelegenheiten in dem Bereich aufgestellt werden können und so das Hörerlebnis akkurat wahrgenommen werden kann.

Kapitel 4

Durchführung und Ergebnisse

4.1 Durchführung des Hörversuchs

4.1.1 Proband*innen

An dem Hörversuch haben elf Proband*innen teilgenommen, darunter sind zwei weiblich und neun männlich gelesene Personen in jungem Alter. Die Teilnehmer*innen haben kein eingeschränktes Hörvermögen und es werden sowohl Audio-affine Personen als auch Laien befragt. Alter und Geschlecht spielen keine Rolle in der weiteren Analyse und werden nicht weiter thematisiert. Aus Gesprächen wurde ebenfalls klar, dass vier Personen mäßig viel Fußball schauen und sieben Personen keine Fußballinteresse mitbringen. Nur zwei der elf Personen waren bereits mehrmals bei einem Fußballspiel live im Stadion. Dies wurde in Gesprächen mit den Proband*innen herausgefunden.

4.1.2 Dauer des Versuchs

Der Hörversuch dauert zwischen 30 und 40 Minuten. Die Proband*innen haben während der Hörtests immer wieder die Möglichkeit sich die Wiedergabe von den verschiedenen Systemen oder Formaten anzuhören, wodurch sich die Dauer von Proband*in zu Proband*in unterscheidet. Als Hörbeispiel für Dolby Atmos, Ambisonics, Direct Out, Lautsprecher, Kopfhörer und statische beziehungsweise mobile Audioperspektive wird eine Szene mit einem Tor verwendet. Diese dauert circa 60 Sekunden und es ist das Tor in realer Geschwindigkeit mit Torjubel der Spieler zu sehen, sowie drei Slow-Motion Wiederholungen des Tors und zwei Slow-Motion Wiederholungen des Torjubels. Die Szene wurde ausgewählt, da hier sehr laut der Jubel der Auswärts-Fans von Borussia Dortmund zu hören ist, lautes Pfeifen der Heimfans von Eintracht Frankfurt und ein Fan-Gesang der Auswärtsfans während den Wiederholungen des Tors. Re-

präsentativ bietet diese Szene auditiv das, was die meiste Zeit über das gesamte Spiel hinweg zu hören ist.

4.1.3 Ablauf

Zu Beginn des Hörversuchs lesen sich die Proband*innen die Einleitung auf dem Fragebogen durch. Dort wird ein Überblick über das Thema gegeben, sowie die drei Teile des Versuchs erläutert. Die Proband*innen können dann erste Rückfragen stellen, falls etwas unverständlich sein sollte. Vor den ersten Hörbeispielen wird das Lautsprecher-Setup erklärt. Der Fernseher steht zwischen dem C-Lautsprecher und dem R-Lautsprecher, um keinen LS zu verdecken (Abbildung 4.1). Deswegen müssen sich die Proband*innen leicht schräg zum Fernseher und frontal zum C-Lautsprecher positionieren. Außerdem wird erklärt, welche Lautsprecher den restlichen Kanälen L, LS, RS, HLF, HRF, HLR und HRR entsprechen und dass bei Ambisonics zwei Lautsprecher mit Center-Height (CH) und Top (T) hinzukommen.

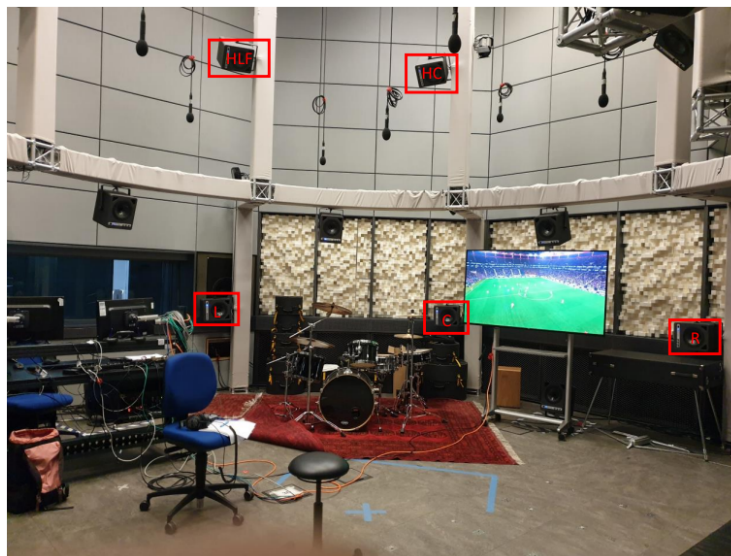


Abbildung 4.1: Aufbau des Hörversuchs; Fernseher ist zwischen den Kanälen C und R

Zu Beginn des ersten Teils des Hörversuchs (HV 1) werden alle drei Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out randomisiert abgespielt, wobei die Proband*innen noch nicht die Fragen beantworten sollen. Dies soll den Proband*innen einen ersten Überblick über den Sound der drei Systeme geben. Erst danach wird erneut das erste Hörbeispiel abgespielt, währenddessen die Fragen für das jeweilige Wiedergabe-System beantwortet werden. Es wird vor jedem Hörbeispiel klar angesprochen, welches der drei Wiedergabe-

Systeme gerade zu hören ist. So werden nacheinander die verschiedenen Systeme angehört und der erste Teil des Fragebogens ausgefüllt. Die drei Teile des Hörversuchs sind alle auf einem eigenen Blatt, um Missverständnisse zu vermeiden.

Bei HV 2 wird erneut vermittelt, dass es sich um einen Vergleich von LS mit HP handelt. Außerdem erfahren die Proband*innen, dass die Wiedergabe über LS Dolby Atmos 5.1.4 und über HP Dolby Atmos binaural zu hören sein wird. Erneut hören die Teilnehmer*innen eines der beiden Formate randomisiert ohne den Fragebogen auszufüllen. Im zweiten Durchgang der Wiedergabe werden dann die Fragen beantwortet. Zwischenzeitlich können auch immer wieder Fragen gestellt werden oder Anmerkungen geäußert werden, welche zusätzlich notiert werden und in Kapitel 4.2.5 präsentiert werden.

Für den Vergleich der statischen mit der mobilen Audioperspektive in HV 3 wird als Erstes das Konzept beider Perspektiven erklärt. Es wird erläutert, was genau unter der statischen und der mobilen Audioperspektive verstanden werden kann. Auch hier hören die Proband*innen beide Versionen randomisiert, bevor dann die Fragen des letzten Hörversuchs beantwortet werden.

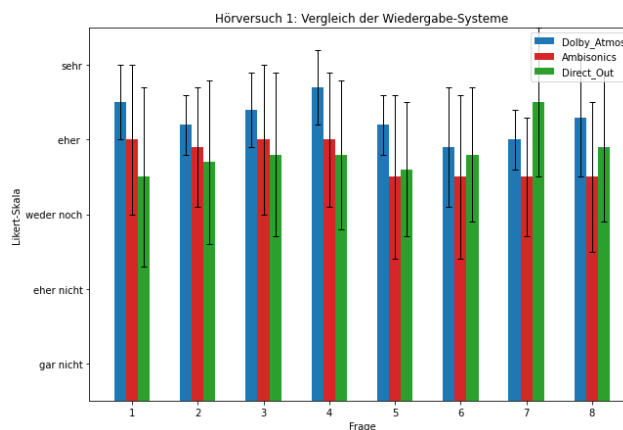
Abschließend hatten die Teilnehmer*innen dann die Gelegenheit, durch ein Gespräch Anmerkungen verbal zu äußern und von dem allgemeinen Eindruck zu erzählen.

4.2 Ergebnisse

Bei allen drei Hörversuchen wurde mit Ausnahme der Präferenz-Fragen mit der Likert-Skala gearbeitet. Wie bereits in Kapitel 3.1.4 erwähnt ist diese streng genommen ordinal skaliert, das heißt die Skalenniveaus können nur in eine Reihenfolge gebracht und einem Rang zugeordnet werden. Es können allerdings zwischen den Stufen nicht die Abständen gemessen werden. In der Praxis wird bei der Likert-Skala allerdings ein metrisches Skalenniveau angenommen, vor allem wenn es sich bei den Stufen um die gängigen Bezeichnungen handelt (sehr, eher, weder noch, eher nicht und gar nicht) (*Likert-Skala | Definition, Anwendung und Auswertung* 2023). In diesem Fall können den Stufen „gar nicht“ bis „sehr“ die Zahlen eins bis fünf zugeordnet werden. Somit lassen sich Mittelwerte oder Standardabweichungen berechnen und statistische Analysen durchführen. Für jede Frage wird den Skalenniveaus die Zahl eins bis fünf zugeordnet, wobei die Zahl fünf einer guten Bewertung entspricht. Nur bei Frage sieben (HV 2) und den Fragen vier und fünf (HV 3) ist die Frage negativ konnotiert, sodass ein hoher Balken einem schlechten Ergebnis entspricht.

4.2.1 Hörversuch 1

Bei Hörversuch eins, dem Vergleich von Dolby Atmos, Ambisonics und Direct Out, bevorzugten sechs Personen Dolby Atmos, nur zwei Personen Ambisonics und drei Personen Direct out. Dies spiegelt sich auch in der Auswertung der einzelnen Fragen wider (Abbildung 4.2).



Frage 1: Wie einschließend ist der Sound dieser Wiedergabe?

Frage 2: Wie umfassend ist der Sound dieser Wiedergabe?

Frage 3: Wie lebendig ist der Sound dieser Wiedergabe?

Frage 4: Wie erfolgreich wird die Stadionatmosphäre vermittelt?

Frage 5: Wie gut wird auditiv die Illusion der Realität wiedergegeben?

Frage 6: Wie gut werden alle Frequenzbereiche wiedergegeben?

Frage 7: Wie gut können einzelne Sounds lokalisiert werden? (Fangesänge, Kommentator, ...)

Frage 8: Wie gut ist der Sound auf die verschiedenen Lautsprecher verteilt?

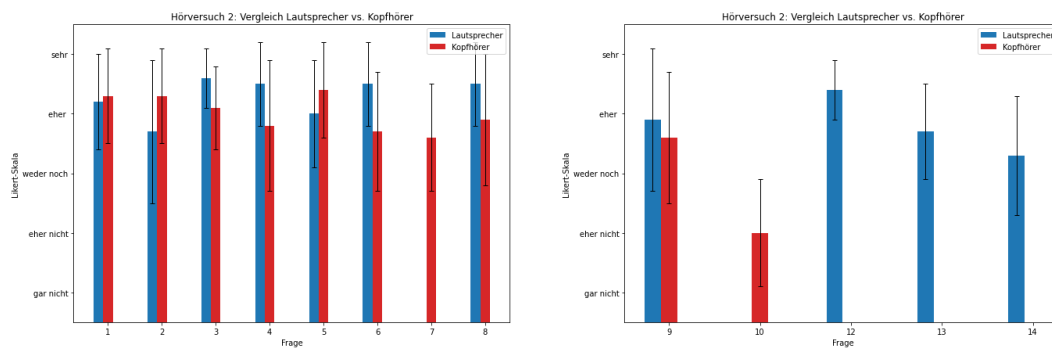
Frage 9: Welches der Wiedergabe-Formate bevorzugen Sie?

Abbildung 4.2: Die Ergebnisse des ersten Hörversuchs über die Proband*innen gemittelt. Die Stufen ‚gar nicht‘ bis ‚sehr‘ entsprechen den Zahlen eins bis fünf. Verglichen werden die Wiedergabe-Systeme Dolby Atmos (blau), Ambisonics (rot) und Direct Out (grün)

Dolby Atmos schneidet im Mittelwert bei allen Fragen außer Frage sieben am besten ab. Auch die Standardabweichungen sind bei Dolby Atmos vergleichsweise gering. Bei Frage sieben (Lokalisation von Sounds) wird Direct Out am besten bewertet. Bei dem Vergleich von Ambisonics mit Direct Out wird Ambisonics bei den Fragen eins bis vier besser bewertet (Immersion) und bei den Fragen fünf bis acht Direct Out (Frequenz und Lokalisation). Die Standardabweichungen sind bei Ambisonics und Direct Out deutlich größer, als bei Dolby Atmos.

4.2.2 Hörversuch 2

Bei dem Vergleich von HP mit LS wird die Wiedergabe über LS bevorzugt. Hier haben sieben von elf Proband*innen für die Wiedergabe über LS gestimmt und nur vier Personen für die Wiedergabe über HP. Aus Gesprächen mit den Teilnehmer*innen wurde aber auch klar, dass das Anwendungsgebiet beider Wiedergabe-Formate sehr unterschiedlich ist (dazu mehr in Kapitel 5). Die beiden Graphen sind unterteilt in die Fragen eins bis acht und neun bis vierzehn, ausschließlich Frage elf, welche die Frage der Präferenz ist. Die ersten acht Fragen beziehen sich auf den direkten Vergleich von LS und HP (Abbildung 4.3). Die Fragen neun bis vierzehn dienen dazu, den Ort und die Art der Anwendung zu erforschen. Die Fragen eins und zwei zur OLE wird bei der HP-Wiedergabe



- Frage 1: Wie gut gefällt Ihnen der Sound?
 Frage 2: Wie erfreut oder genervt sind Sie über den Sound während der Wiedergabe?
 Frage 3: Wie immersiv ist die Audio-Wiedergabe? (immersiv: umschließend, lebendig, umfassend, intensiv)
 Frage 4: Wie gut wird auditiv die Illusion der Realität wiedergegeben?
 Frage 5: Wie gut werden alle Frequenzbereiche wiedergegeben? (Bässe, Mitten, Höhen)
 Frage 6: Wie groß oder mächtig wirkt die Wiedergabe auf Sie?
 Frage 7: Wie stark kommt es zur Im-Kopf-Lokalisation? (Hörereignisort nicht außerhalb vom Kopf in der Umgebung, sondern Orte von Schallquellen scheinen im Kopf zu liegen)
 Frage 8: Wie erfolgreich wurde die Stadion-Atmosphäre vermittelt?
 Frage 9: Wie gerne würden Sie zu Hause eine Sportübertragung mit 3D-Audio hören?
 Frage 10: Wie gerne würden Sie zu Hause mit Freunden zusammen eine Sportübertragung per Kopfhörer hören?
 Frage 11: Welche Wiedergabe bevorzugen Sie?
 Frage 12: Wie wichtig ist Ihnen eine gute Soundanlage zu Hause?
 Frage 13: Wie gerne würden Sie für ein besseres Hörerlebnis Geld ausgeben? (z.B. 5€ Eintritt für eine Sportsbar mit 3D-Sound)
 Frage 14: Wie wahrscheinlich ist es, dass Sie sich für ein besseres Hörerlebnis ein Heimkino einrichten? (Acht oder mehr Lautsprecher 3D-Soundanlage)

Abbildung 4.3: Die Ergebnisse des zweiten Hörversuchs über die Proband*innen gemittelt. Die Stufen „gar nicht“ bis „sehr“ entsprechen den Zahlen eins bis fünf. Verglichen wird die Wiedergabe über Lautsprecher (blau) und Kopfhörer (rot)

besser bewertet. Bei den Fragen drei und vier zur Immersivität schneidet die LS-Wiedergabe besser ab. Ebenso bei Frage acht, welche sich auf die erfolgreiche Vermittlung der Stadionatmosphäre bezieht. Bei Frage fünf zu den Frequenzbe-

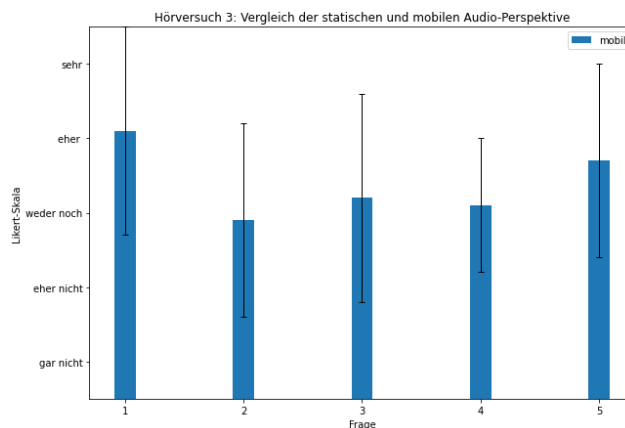
reichen ist die HP-Wiedergabe besser und bei Frage sechs zur Mächtigkeit sehr deutlich die LS-Wiedergabe. Über die Proband*innen gemittelt kommt es bei der HP-Wiedergabe eher zu einer IKL, allerdings tendieren hier acht der befragten Personen zu der Bewertung „eher“ bis „sehr“ und nur zwei Personen haben angegeben, dass es „eher nicht“ zur IKL kommt. Bei Frage neun zur Wiedergabe zu Hause schneiden beide Wiedergabe-Formate ungefähr gleich ab. Frage zehn zeigt, dass die Proband*innen zu Hause mit Freund*innen zusammen „eher nicht“ eine Sportübertragung mit 3D-Sound per HP schauen wollen. Frage zwölf und vierzehn zeigen, dass den Personen eine gute Soundanlage zu Hause wichtig ist, die Wahrscheinlichkeit, sich ein Heimkino mit 3D-Soundanlage einzurichten allerdings weniger hoch ist. Frage dreizehn zeigt, dass die Bereitschaft, Eintritt für eine Sports-Bar mit 3D-Audio zu bezahlen, etwas höher ist.

4.2.3 Hörversuch 3

Bei dem dritten Hörversuch, dem Vergleich der statischen mit der mobilen Audioperspektive, haben, mit Ausnahme von zwei Proband*innen, den Perspektivenwechsel bei der mobilen Variante „eher stark“ bis „sehr stark“ bemerkt. Bei Frage zwei ist zu sehen, dass die an die Kameraperspektive angepasste Audioperspektive von „eher nicht gut“ bis „eher gut“ durchschnittlich aufgenommen wurde (Abbildung 4.4). Lediglich ein*e Proband*in fand den Wechsel „sehr gut“ und zwei Proband*innen fanden den Wechsel „gar nicht gut“. Auch bei Frage drei, wie gut die Immersivität der Wiedergabe durch den auditiven Perspektivenwechsel gesteigert wird fanden drei Personen die Steigerung „sehr stark“ und eine Person „gar nicht stark“. Im Schnitt wird die Immersivität weder sehr stark noch gar nicht stark gesteigert. Fragen vier und fünf sind negativ konnotiert, dementsprechend ist ein hoher Balken schlecht behaftet. Bei Frage vier ist festzuhalten, dass die Proband*innen den Perspektivenwechsel durchschnittlich verwirrend fanden. Bei der Einschätzung, ob ein Spiel 90 Minuten lang mit auditiven Perspektivenwechseln nervig werden kann, zeigt sich, dass die Perspektivenwechsel als „eher nervig“ eingeschätzt werden. Dementsprechend bevorzugen mit 73% acht Personen die statische Audioperspektive und mit 23% drei Personen die mobile Audioperspektive.

4.2.4 Anmerkungen

Zum HV 1 hat ein*e Proband*in angemerkt, dass bei Direct Out einzelne Sounds wie Fanggruppierungen deutlich besser lokalisierbar sind. Dies hat eine weitere Person ebenso aufgefasst und den Sound von Direct out als „am klarsten“ beschrieben. Allerdings hat eine Person auch angemerkt, dass die Wiedergabe



Frage 1: Wie stark merken Sie einen auditiven Perspektivwechsel?

Frage 2: Wie gut finden Sie den Perspektivwechsel?

Frage 3: Wie stark wird die Immersivität durch den Perspektivwechsel gesteigert?

Frage 4: Wie verwirrend finden Sie den Perspektivwechsel?

Frage 5: Wie nervig kann das Hörerlebnis werden, wenn die auditive Perspektive über 90 Minuten an das Kamerabild angepasst wird? (Vermutung)

Frage 6: Welche Audio-Perspektive bevorzugen Sie?

Abbildung 4.4: Die Ergebnisse des dritten Hörversuchs über die Proband*innen gemittelt. Die Stufen „gar nicht“ bis „sehr“ entsprechen den Zahlen eins bis fünf. Verglichen wird die statische mit der mobilen Audio-Perspektive

über Direct Out lückenhaft ist, also dass einzelne Lautsprecher wahrnehmbar sind und kein homogenes Schallfeld um den/die Hörer*in herum entsteht. Auch auf der z-Ebene ist einer Person aufgefallen, dass zu wenig Sound von oben zu hören ist. Eine andere Person hat angemerkt, dass die 3D-Sound Reproduktion über Lautsprecher nicht mit der Atmosphäre im Stadion zu vergleichen ist, da über Lautsprecher die Tiefe im Sound fehlt und nicht gut zwischen nahen und fernen Sounds unterschieden werden kann. Im Stadion herrsche eine deutlich höhere Überflutung an Reizen, allerdings komme die 3D-Audio-Wiedergabe schon deutlich näher an das reale Erlebnis als eine Stereo-Wiedergabe. Zu Frage sieben wurde mehrmals angemerkt, dass gerade im unteren Bereich des Frequenzspektrums etwas fehlen würde. Dies hängt damit zusammen, dass der LFE-Kanal leer geblieben ist. Es werden somit keine Subs verwendet, was allerdings bei zukünftigen Produktionen noch hinzugefügt werden kann. Es ist aber dennoch bemerkenswert, dass bei manchen Wiedergaben-Systemen spezifisch angemerkt wurde, es fehle der untere Bassbereich und dass bei anderen Wiedergabe-Systemen nichts fehlt. Eine Person hat notiert, dass bei Ambisonics der Eindruck entsteht, man höre „von außen“ zu und bei Dolby Atmos sei man „mittendrin“.

Bei HV 2 haben mehrere Personen angemerkt, dass beide Wiedergaben über LS und HP sehr gut sind, es aber sehr auf die Situation ankommt, welches For-

mat dementsprechend bevorzugt wird. Auch durch Gespräche nach den Versuchen mit den Proband*innen wurde deutlich, dass die HP-Mischung gut geeignet ist, falls das Fußballspiel alleine geschaut wird. Wird das Spiel mit mehreren Personen geschaut, mit denen man sich auch unterhalten will, so kommt nur die LS-Wiedergabe in Frage.

Die Anmerkungen haben sich beim HV 3 am meisten gehäuft. Hier gibt es sowohl sehr positive, als auch sehr negative Anmerkungen zur mobilen Audioperspektive, aber auch sehr interessante Ideen, wie die mobile Perspektive eventuell angepasst werden kann. Eine Person merkt an, dass gerade bei schnelleren Schnitten die wechselnde Audioperspektive sehr störend ist. Anschließend an diesen Kommentar schreibt eine andere Person, dass durch die Perspektivenwechsel die einzelnen Kameraschnitte deutlich bemerkbarer werden. Dadurch verliere das Spiel den Gesamteindruck und wirkt nicht mehr wie ein fortlaufendes Spiel, sondern wie einzelne aneinandergereihte Video-Clips. Dadurch würde der Eindruck der Realität, welcher durch das 3D-Sounderlebnis erst erzeugt wird, wieder verloren gehen. Eine andere Person merkt an, dass sie rückblickend auf die statische Audioperspektive einen auditiven Perspektivenwechsel erwarten würde, dieser aber nicht vorkommt. Dementsprechend erfüllt erst die mobile Audioperspektive diesen erwarteten Wechsel. Eine weitere Person schreibt, dass die auditiven Perspektivenwechsel zu Beginn erst etwas verwirrend wirken, man sich aber schnell an diesen Wechsel gewöhnen würde. Eine Person merkt an, dass bei den Audioperspektiven geklärt werden muss, welches Ziel verfolgt werden soll. Eine statische Audioperspektive eigne sich viel mehr für das „normale“ Schauen eines Fußballspiels und die mobile Audioperspektive für beispielsweise einen Film über ein Fußballspiel. Sie sei also viel mehr ein filmtechnisches Stilmittel und steigert die Immersivität.

Über Gespräche mit den Proband*innen wurde auch angemerkt, dass durch eine mobile Audioperspektive die Verbundenheit mit der Kamera deutlich gesteigert wird und der/die Zuhörer*in das Gefühl hat, in der Rolle der Kamera zu stecken. Ein*e Proband*in hat auch angemerkt, dass es gut wäre, wenn Zuschauer*innen sich eine Position im Stadion aussuchen könnten und dann von dieser Position aus das Spiel mit einer statischen Perspektive verfolgen können. Beispielsweise wird ausgewählt, dass das Spiel aus der Perspektive der Gästefans auf der Auswärtstribüne verfolgt werden soll und der Sound der Atmosphäre dieser Position entspricht. Dies würde allerdings in der Postproduktion weitere Mischungen erfordern. Eine weitere Person hat angemerkt, dass die Schnitte der Audioperspektive weicher sein sollten und ein langsamer Übergang von eins bis zwei Sekunden das Audioerlebnis deutlich angenehmer machen würde.

4.2.5 Notizen

Des Weiteren wurden während des Hörversuchs folgende Notizen gemacht. Die beiden Begriffe „einschließend“ und „umfassend“ hätten direkt bei der Frage erläutert werden sollen. Viele Proband*innen haben nach dem Unterschied der beiden Wörter gefragt. Während ‚einschließend‘ bedeutet, wie ‚gefangen‘ man in einer virtuellen Umgebung ist, ist unter dem Begriff ‚umfassend‘ eher ‚nahezu alles enthaltend‘ zu verstehen. Ferner hätten die Fragebögen personalisierter sein können. Die Reihenfolge der Wiedergabe-Systeme bei HV 1 ist auf dem Fragebogen immer gleichbleibend sortiert, aber die Proband*innen haben die Systeme in randomisierter Reihenfolge gehört. Obwohl es nicht zu Verwechslungen kam, hätte dies eine Fehlerquelle sein können. Die Anpassung könnte allerdings auch bei der Auswertung zu Fehlern führen, wobei diese einfacher zu vermeiden und korrigieren sind. Auch die Fragen hätten an manchen Stellen geordneter sein können. Beispielsweise hätte bei HV 2 die Frage acht direkt nach Frage vier gestellt werden können, da diese das ähnliche Thema Immersivität und Realität befragen. Auch Frage elf hätte vor Frage zehn gestellt werden können, weil diese Frage nicht mehr den direkten Vergleich von HP mit LS untersucht. Durch das Sitzen am Rand zwischen dem L-Lautsprecher und dem LS-Lautsprecher während des Hörversuchs ist aufgefallen, dass die mobile Audioperspektive in HV 3 nur für den/die Hörer*in im Sweet-Spot geeignet ist. Sitzt eine Person außerhalb des Sweet-Spots und noch dazu sehr weit am Rand, ist die Lokalisation der Audioszene nicht mehr plausibel. Außerdem entsteht bei HV 1 Dolby Atmos ein deutlich größeres Richtungsgefühl als Ambisonics oder Direct Out. Während man bei Ambisonics am Rand sitzt und auditiv das Gefühl von vorn und hinten verliert, so erkennt man bei Dolby Atmos auch am Rand die vorderen Kanäle L-C-R und die hinteren Kanäle LS und RS.

Kapitel 5

Auswertung und Diskussion

Die drei Hörversuche werden, wie in Kapitel 3.1.4 beschrieben, ausgewertet. Dabei werden die Stufen der Likert-Skala als metrisch mit den Zahlenwerten eins bis fünf angenommen. Daraufhin lassen sich die Mittelwerte und Standardabweichungen der einzelnen Fragen bilden. Anhand der einfaktoriellen Varianzanalyse (ANOVA) und dem t-Test bei abhängigen Stichproben werden die Mittelwerte der verschiedenen Gruppen miteinander verglichen. Die Nullhypothese ist, dass es keinen signifikanten Unterschied zwischen den Mittelwerten gibt. Diese Nullhypothese soll widerlegt werden und ein signifikanter Unterschied bei den Mittelwerten bestätigt werden. Als α -Niveau wird der Wert $\alpha = 0,05$ gewählt. Bei den Hörversuchen wird die Prüfgröße mit dem kritischen Wert verglichen und der P-Wert betrachtet. Die Prüfgröße muss im Betrag über dem kritischen Wert liegen und der P-Wert kleiner als 0,05 sein, damit die Nullhypothese verworfen und die Alternativhypothese angenommen werden kann.

5.1 Hörversuch 1

Die Ergebnisse des ersten Hörversuchs (Kapitel 4.2.1) zeigen, dass Dolby Atmos am besten bewertet wird. Mit 55% bevorzugen über die Hälfte der Proband*innen das Wiedergabe-System Dolby Atmos, mit 27% nur drei Personen Direct Out und mit 18% nur zwei Personen Ambisonics. An den Mittelwerten der Fragen eins bis acht ist zu sehen, dass die Fragen durchschnittlich bei Dolby Atmos zwischen „eher“ und „sehr“ am besten bewertet werden. Die Mittelwerte der Fragen bei den Systemen Ambisonics und Direct Out werden durchschnittlich nur zwischen „weder noch“ und „eher“ eingeordnet. Dolby Atmos besitzt außerdem die geringsten Standardabweichungen, was bedeutet, dass alle Proband*innen die Fragen relativ ähnlich bewertet haben. Bei Ambisonics und Direct Out gibt es bei der Bewertung der Fragen größere Schwankungen

zwischen den Proband*innen. Bei den Fragen eins bis fünf zu der Immersivität sowie Realitätsillusion schneidet Dolby Atmos am besten ab. Auch bei der Frage über die Wiedergabequalität der Frequenzbänder wird Dolby Atmos am besten bewertet. Ambisonics und Direct Out werden allerdings ähnlich gut bewertet und nur bei Ambisonics ist anhand der Standardabweichung ein kleiner negativer Versatz zu sehen. Bei der achten Frage über die Lokalisationsgenauigkeit wird Direct Out am besten bewertet. Hier bietet das direkte Routing zu den Lautsprechern den Vorteil, dass einzelne Punktschallquellen wiedergegeben werden. Beispielsweise wird der Kanal 15 „Fans H“ nur über den R-Lautsprecher wiedergegeben und kann so genau aus dieser Richtung erkannt werden. Bei Dolby Atmos und Ambisonics hingegen wird die Information des Kanals „Fans H“ für die Richtung „R“ enkodiert und auf mehrere Lautsprecher verteilt. Wird die Spur im Solo wiedergegeben, so hört man auch aus angrenzenden Lautsprechern (beispielsweise C oder RS) ein Signal, welches jedoch deutlich leiser ist. Auf diese Weise entsteht zwar ein homogeneres Schallfeld, allerdings können einzelne Sounds schlechter aus einer bestimmten Richtung wahrgenommen werden. Bei Direct Out ist dagegen das Schallfeld nicht sehr homogen und einzelne Proband*innen haben „Lücken“ zwischen den Lautsprechern gehört. Dies ist ein klares Manko, welches sich auf die allgemeine Soundqualität auswirkt. Dennoch mögen manche Proband*innen die erhöhte Lokalisationsgenauigkeit, weswegen Direct Out auch von 27% bevorzugt wird.

Die statistische Analyse mittels einfaktorieller Varianzanalyse (ANOVA) zeigt, dass es einen statistisch signifikanten Unterschied zwischen den Mittelwerten der Fragen gibt (Abbildung 5.1).

ZUSAMMENFASSUNG						
Gruppen	Anzahl	Summe	Mittelwert	Varianz		
Dolby	7	29,5	4,21	0,04		
Ambi	7	25,9	3,70	0,06		
Dir Out	7	26,8	3,83	0,11		

ANOVA						
Streuungsursache	Quadratsummen (SS)	Freiheitsgrade (df)	Mittlere Quadratsumme (MS)	Prüfgröße (F)	P-Wert	kritischer F-Wert
zwischen den Gruppen	1,00	2	0,501	7,036	0,006	3,555
Innerhalb der Gruppen	1,28	18	0,071			
Gesamt	2,29	20				

Abbildung 5.1: Die Ergebnisse der statistischen Analyse mittels einfaktorieller Varianzanalyse (ANOVA). Wichtige Werte sind der Mittelwert, die Prüfgröße (F), der p-Wert und der kritische F-Wert

Es ist zu sehen, dass der kritische Wert $F = 3,555$ mit $7,036$ übertroffen wurde. Auch der p-Wert liegt mit $p = 0,006$ deutlich unter dem α -Niveau von

$\alpha = 0,05$, wodurch die Nullhypothese $H_0 =$ „Es gibt keinen signifikanten Unterschied zwischen den Mittelwerten der drei Gruppen“ verworfen werden kann. Es gilt somit die Alternativhypothese, dass ein signifikanter Unterschied festgestellt wird. Damit wird ein Unterschied zwischen den Mittelwerten bei den drei Gruppen Dolby Atmos, Ambisonics und Direct Out bestätigt. Dies heißt zwar erst mal nur, dass ein Unterschied existiert, dieser aber nicht positiv oder negativ bewertet werden kann. Allerdings ist im Zusammenhang mit den Mittelwerten bei den Fragen eins bis acht und dem Ergebnis der Frage neun, dass sechs der elf Proband*innen Dolby Atmos bevorzugen festzuhalten, dass Dolby Atmos als das statistisch signifikant bevorzugte Wiedergabe-System gilt.

Bei dem Vergleich der drei Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out ist festzuhalten, dass das vorliegende Audiomaterial zwar in Form von diskreten Signalen vorliegt und somit für verschiedene Systeme aufbereitet werden kann, allerdings erweckt es den Eindruck, speziell für das objektbasierte 3D-Audiosystem geeignet zu sein. Es gibt viele weitere Aufnahmetechniken, um beispielsweise Audiomaterial für Ambisonics bereitzustellen. Hier kann beispielsweise mit dem Ambeo VR Mic direkt ein Ambisonics B-Format aufgenommen werden. Ob die Wiedergabe dadurch besser wird oder die Immersivität gesteigert wird, gilt herauszufinden. Bei der Aufbereitung des Audiomaterials für die verschiedenen Wiedergabe-Systeme ist zu bemerken, dass sich der Aufwand bei Dolby Atmos und Ambisonics nicht viel unterscheidet. Beide Systeme erfordern einen Renderer beziehungsweise einen En- und Dekodierer. Dementsprechend ist die Umsetzung der beiden Formate für den Heimanwender im Wohnzimmer oder einen Besitzer einer Sports-Bar umständlicher umzusetzen als das System Direct Out. Hier werden alle Spuren, so wie sie angeliefert werden, direkt weiter auf die entsprechenden Lautsprecher geroutet und erfordern keine weitere Signalverarbeitung. Es gibt auch Endgeräte, welche die entsprechenden Codecs besitzen, Dolby Atmos oder Ambisonics zu lesen. Diese ermöglichen die Wiedergabe von 3D-Audio, ohne einen Laptop mit einer DAW in der Signalkette haben zu müssen und ist gerade für Sports-Bars sehr empfehlenswert.

Aufgrund des Ergebnisses des Hörversuchs und dem signifikanten Unterschied bei den Mittelwerten ist zu empfehlen, Dolby Atmos als Wiedergabe-System zu wählen. Bei Aspekten wie Immersivität, Illusion der Realität, Sound-Qualität und dem Frequenzbereich scheidet dieses Wiedergabe-System am besten ab. Auch die Sweet-Area ist, wie in Kapitel 3.2.10 gezeigt, am größten. Gerade nach hinten ist der Bereich, in dem eine plausible Lokalisation der Audioszene wahrgenommen werden kann sehr groß. Für die Anwendung im Wohnzim-

mer bedeutet das, dass die Zuhörer*innen sich relativ weit nach hinten setzen können. Eine Couch steht im Wohnzimmer in der Regel an der dem Fernseher oder der Leinwand gegenüberliegenden Wand, was zu der Sweet-Area des Wiedergabe-Systems passt. Auch in einer Sports-Bar ist der nach hinten ausgeweitete Bereich von Vorteil. Gerade bei Ambisonics ist die Sweet-Area deutlich kleiner und eignet sich somit nicht unbedingt für die Wiedergabe für mehrere Personen.

Während bei Dolby Atmos schnell eine obere Grenze von 12 LS erreicht ist (Wiedergabe-Setup 7.1.4), können bei Ambisonics deutlich mehr Lautsprecher zusätzlich aufgebaut werden. Das maximale Potential würde erst mit 64 Audio-Kanälen und 7th-Order Ambisonics ausgeschöpft werden. Dies ist in der Praxis allerdings sehr unwahrscheinlich und schwierig umzusetzen. Dennoch ist die Vermutung trivial, dass durch das Hinzufügen von Lautsprechern auch die Sound-Qualität bei Ambisonics steigt. Liegt es im Bereich des Möglichen für den/die Konsument*in, ein solches Audio-System mit mehr als 12 LS aufzubauen, kann über die Verwendung von Ambisonics durchaus nachgedacht werden.

5.2 Hörversuch 2

Der zweite Hörversuch zeigt, dass mit 64% sieben der elf Proband*innen die Wiedergabe über LS bevorzugen (vergleiche Kapitel 4.2.2). Mit 63% haben nur vier Personen die HP-Wiedergabe bevorzugt. Auch die Mittelwerte und Standardabweichungen sind bei den Fragen, welche LS und HP vergleichen sehr ähnlich. Während bei den ersten beiden Fragen über die OLE Kopfhörer (HP) bevorzugt werden, werden bei den nächsten beiden Fragen drei und vier LS bevorzugt. Dies zeigt, dass die OLE bei Kopfhörern zwar etwas besser ist, die Immersivität und Illusion der Realität allerdings bei der LS-Wiedergabe. Die Wiedergabe über LS wirkt außerdem größer beziehungsweise mächtiger, wie es bei Frage sechs gezeigt wird und es kann nicht zur IKL kommen. Bei Frage fünf zu den Frequenzbereichen werden HP bevorzugt, dies ist allerdings auch sehr von dem HP-Modell abhängig und dem damit verbundenen Frequenzgang. Im zweiten Teil von HV 2 werden weniger die beiden Wiedergabe-Formate LS und HP miteinander verglichen, sondern vielmehr das Anwendungsgebiet und die Präferenzen. An dem Ergebnis von Frage neun (vgl. Abbildung 4.3) ist zu sehen, dass beide Formate ungefähr ähnlich gerne für die Wiedergabe zu Hause verwendet werden würden. Allerdings zeigt Frage zehn, dass falls ein Sportevent mit Freunden zusammen geschaut wird, die Wiedergabe über HP „eher nicht“ vorstellbar ist. Hier geht zu sehr das Gemeinschaftsgefühl und die Möglichkeit, sich zu unterhalten, verloren. Die Fragen zwölf und vierzehn zeigen, dass den

Proband*innen eine sehr gute Soundanlage zu Hause zwar sehr wichtig ist, das Zulegen einer 3D-Audio Soundanlage (acht oder mehr LS) allerdings unwahrscheinlich. Bei den Proband*innen ist dafür eine eher hohe Bereitschaft vorhanden, eine Sports-Bar zu besuchen, in der ein Sportevent mit 3D-Audio wiedergegeben wird.

Die statistische Analyse mittels t-Test zeigt, dass es keinen statistisch signifikanten Unterschied zwischen den Mittelwerten der Fragen gibt (Abbildung 5.2). Im Gegensatz zu HV 1 ist zu sehen, dass der kritische Wert $t = 3,555$ mit $0,995$

Zweistichproben t-Test bei abhängigen Stichproben (Paarvergleichstest)		
	LS	HP
Mittelwert	4,29	4,07
Varianz	0,11	0,08
Beobachtungen	7	7
Pearson Korrelation	-0,749	
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	6	
t-Statistik	0,995	
P(T<=t) einseitig	0,179	
Kritischer t-Wert bei einseitigem t-Test	1,943	
P(T<=t) zweiseitig	0,358	
Kritischer t-Wert bei zweiseitigem t-Test	2,447	

Abbildung 5.2: Die Ergebnisse der statistischen Analyse mittels t-Test. Wichtige Werte sind der Mittelwert, die t-Statistik, P(T_i=t) einseitig und der kritische t-Wert bei einseitigem t-Test.

nicht übertroffen wurde. Auch der p-Wert (einseitig) liegt mit $p = 0,179$ etwas über dem α -Niveau von $\alpha = 0,05$. Somit kann die Nullhypothese $H_0 =$ „Es gibt keinen signifikanten Unterschied zwischen den Mittelwerten der beiden Gruppen“ nicht verworfen werden. Es gilt somit auch nicht die Alternativhypothese, dass ein signifikanter Unterschied festgestellt werden kann. Der nichtvorhandene signifikante Unterschied zwischen den Mittelwerten zeigt, dass die beiden Wiedergabe-Formate LS und HP sich nicht zwingend voneinander unterscheiden und eines der beiden besser beziehungsweise schlechter ist. Es ist vielmehr festzuhalten, dass beide Formate ihre Stärken haben und in manchen Bereichen bevorzugt und in anderen Bereichen abgelehnt werden.

Vielmehr kann festgehalten werden, dass die Anwendungsgebiete der beiden Formate sehr unterschiedlich sind. Die LS-Wiedergabe muss vor allem bei einem gemeinsamen Verfolgen eines Sportevents mit vielen Menschen verwenden

det werden. In dieser Situation ist es erforderlich, sich unterhalten zu können und allgemein miteinander interagieren zu können. Bei der HP-Wiedergabe ist jede Person zu sehr abgeschottet von der Umgebung. Dies kann aber auch zum Vorteil werden, wenn eine Person ein Sport-Event alleine verfolgen will, so kann es durchaus erwünscht sein, sich von der Umgebung abzukapseln. Events können sehr viel einfacher unterwegs mit 3D-Sound auf dem Smartphone mit Kopfhörern angehört werden. Auch zu Hause im Wohnzimmer ist die Wiedergabe über HP nicht undenkbar. Eine 3D-Audio Soundanlage ist kostspielig und möglicherweise möchte nicht jede*r sie im Wohnzimmer aufbauen. An der Stelle ist es vorteilhaft, auf die HP-Möglichkeit zurückzugreifen. Dies ist aber, wie bereits erwähnt, nur allein zu empfehlen.

Die Aufbereitung des Audiomaterials für die beiden Wiedergabe-Formate LS und HP sind sehr ähnlich. Dies ist auch unabhängig davon, ob als Wiedergabe-System Dolby Atmos oder Ambisonics verwendet wird. Dolby Atmos bietet über den Renderer die Möglichkeit einen binauralen Downmix anzufertigen. Das Endgerät, welches das Dolby Atmos Signal empfängt muss lediglich einen HP-Anschluss besitzen. Auch Ambisonics bietet über einen binauralen Dekoder die Möglichkeit an, eine HP-Mischung anzufertigen. Lediglich das Wiedergabe-System Direct Out ist nur für die Wiedergabe über LS geeignet. Würden die Spuren ohne Signalverarbeitung direkt auf die beiden HP-Kanäle L und R geroutet werden, würde ein Stereo-Signal entstehen. Erst durch das Aufbereiten mit einer HRTF entsteht ein binauraler Mix.

5.3 Hörversuch 3

Bei dem dritten Hörversuch ist deutlich zu sehen, dass mit 73% acht der elf Proband*innen die klare Mehrheit eine statische Audioperspektive bevorzugt. Während die wechselnde Audioperspektive bei der mobilen Variante von den Proband*innen bemerkt wurde, ist das Empfinden sehr durchwachsen. An der hohen Standardabweichung (vgl. Abbildung 4.4) ist zu sehen, dass ein Teil der Proband*innen die wechselnde Audioperspektive als „eher gut“, „eher Immersivitäts steigernd“ und „eher nicht nervig“ wahrnehmen. Ein anderer Teil der Proband*innen sieht das anders. Hier wird der auditive Perspektivenwechsel als „eher nicht gut“, „eher nicht immersivitäts-steigernd“ und „eher nervig“ wahrgenommen. Im Durchschnitt ergibt sich ein neutraler Mittelwert bei „weder noch“. Bei Frage fünf ist zu sehen, dass der auditive Perspektivenwechsel, sollte er 90 Minuten lang vorkommen, als ‚eher nervig‘ eingeschätzt wird. Dies ist etwas stärker der Fall, als bei einem kürzeren Audio-Ausschnitt. Es ist also möglich, beispielsweise für ein Highlight-Video, welches etwa drei bis fünf Minuten dau-

ern kann, die mobile Audioperspektive zu verwenden. Für ein Sport-Event, welches 90 Minuten andauern kann, ist zu empfehlen, auf die statische Perspektive zurückzugreifen. Da die mobile Audioperspektive nur von 27% bevorzugt wird ist es auch empfehlenswert, Sport-Events mit einer weiterhin statischen Audioperspektive zu übertragen und die mobile Variante nur als Option zur Verfügung zu stellen.

Anhand der Anmerkungen der Proband*innen (vgl. Kapitel 4.2.4) und den Notizen während des Hörversuchs (vgl. Kapitel 4.2.5) wird aber auch klar, dass die mobile Audioperspektive deutlich andere Ansätze verfolgen kann. In diesem Versuch wurde das Ziel verfolgt, die mobile Audioperspektive an die Kameraperspektive anzupassen und die Stadionatmosphäre aus dem Hör- und Blickwinkel der Kamera zu verfolgen, um so das Hörerlebnis noch immersiver zu gestalten. Dies konnten auch einige Proband*innen erfolgreich bestätigen, da sie sich während der Wiedergabe an die Kameraposition im Stadion versetzt gefühlt haben.

Ein anderer Ansatz, die Stadionatmosphäre immersiver zu gestalten, wäre es, die Atmosphäre von einer bestimmten Position im Stadion statisch zu verfolgen. Es kann beispielsweise ein Fußballspiel von der Position der Haupttribüne, der Auswärts-Fans oder der Heim-Fans aus verfolgt werden. Je nach Position müsste ein unterschiedlicher Mix angefertigt werden. Für den Fall, dass Zuschauer*innen das Spiel aus der Position der Heim-Fans verfolgen wollen, müssen die Kanäle der Heim-Fans deutlich lauter und über den 3D-Panner anders im Raum verteilt werden. Diese Perspektive ändert sich über das Spiel hinweg allerdings nicht, da sich die Position eines/einer Zuschauer*in im Stadion auch nicht verändern würde. Die verschiedenen Kameraperspektiven bieten in dem Sinne, wie auch Wiederholungen, nur einen visuellen Mehrwert, den ein/e Zuschauer*in im Stadion so nicht erfahren würde. Realisierbar wäre dieses Prinzip über ein Auswahlmenü, ähnlich wie es bei der Sprache oder den Untertiteln eingesetzt wird. Ein weiterer Vorteil davon ist, dass Fans einer bestimmten Mannschaft so die Übertragung aus dem Stadion „verbundener“ mit den eigenen Fans verfolgen könnten. Dazu wird einfach nur die Option mit der jeweiligen Position im Stadion ausgewählt.

Die mobile Audioperspektive kann an sich auch noch weiter optimiert werden. Die Blenden der verschiedenen Audioperspektiven können bei dem Wechsel eines Kamerabilds deutlich weicher geschehen, indem ein ein- bis zwei-sekündiger Übergang verwendet wird. Dies sorgt dafür, dass die Schnitte deutlich weicher und weniger auffällig werden. Wiederholungen von Szenen bieten auch verschiedene Ansätze, das Hörerlebnis immersiver zu gestalten. Während bei diesem Versuch das Audiomaterial während den Slow-Motion Wiederholungen

kontinuierlich mit der Atmosphäre im Stadion weiterläuft, ist es auch möglich, das Audiomaterial der Szene wiederzuverwenden. Damit ist der/die Zuhörer*in deutlich mehr mit der wiederholten Szene verbunden. Die Audioperspektive an die Kameraperspektive anzupassen, ergibt mit diesem Ansatz auch mehr Sinn, da jetzt bei den Wiederholungen der Kamerabilder die entsprechenden Audio-szenen zu hören sind. Ein Problem ist allerdings, dass die Wiederholung einer Szene aufgrund des Slow-Motion-Effekts länger dauert, als die eigentliche Szene. Ein zeitliches verlängern des Audiomaterials verursacht, dass der Sound sich tiefer anhört. Dieser Effekt ist sehr stark hörbar und kann sich negativ auf das Hörerlebnis auswirken.

Auch das Aufnahme-Verfahren kann an die Wiedergabe einer mobilen Audioperspektive angepasst werden. Gerade für Live-Spiele ist es erforderlich keine weitere Signalverarbeitung zu benötigen. Würde beispielsweise auf jeder Kamera ein 3D-Mikrofon platziert werden, kann automatisch bei dem Schnitt des Kamerabildes auf das jeweilige Mikrofon zugegriffen werden. Das Mikrofon befindet sich dann auch in der richtigen Position, da es fest auf der Kamera befestigt ist. Ein großes Problem dieses Ansatzes ist allerdings, dass Sport-Übertragungen nicht nur mit einem Mikrofon beziehungsweise dem Schnitt von Mikrofon zu Mikrofon aufgenommen werden, sondern dass ein Mix mit mehreren Mikrofonen angelegt wird, was die Produktion dieses Ansatzes wieder erschweren würde.

Anhand dieser Szenarien wird deutlich, dass auf diesem Gebiet noch viel untersucht werden kann. Es gilt noch herauszufinden, welches Format und welche Idee der mobilen Audioperspektive dann das beste immersive Erlebnis bietet.

5.4 Ausblick

Alle drei Hörversuche haben gewinnbringende Erkenntnisse geliefert. Anschließend an HV 1 kann die Produktion einer Stadionaufnahme weiterentwickelt und speziell für Ambisonics produziert werden. Hier gibt es verschiedene Ansätze, wie beispielsweise 3D-Mikrofone zu benutzen. Ob dies, verglichen mit Dolby Atmos, ein besseres oder immerisveres Klangerlebnis liefert, gilt es zu testen. Dolby Atmos ist insofern eingeschränkter als Ambisonics, als dass die maximale Anzahl an Lautsprechern mit 7.1.4 erreicht ist. Bei Ambisonics können weitaus mehr Lautsprecher aufgestellt werden. Ein großer Vorteil von Dolby Atmos ist allerdings die größere Sweet-Area. Ob und wie diese bei Ambisonics weiter vergrößert werden kann muss untersucht werden. In der Literatur gibt es mehrere Ansätze diesen Bereich des optimalen Hörens bei Stereo oder Surround zu vergrößern.

HV 2 hat gezeigt, dass die Wiedergabe über Lautsprecher oder Kopfhörer sehr anwendungsabhängig ist. Beide Varianten erzielen ein gutes 3D-Sound-Erlebnis. Produzenten oder TV-Sender können versuchen, einen Weg zu finden, wie sie beide Optionen anbieten, sodass der/die Konsument*in zwischen den beiden Optionen auswählen kann.

Bei HV 3 gibt es sehr viele Ansätze die mobile Audioperspektive anzupassen. Je nachdem, welches Ziel durch die Perspektivenwechsel erzielt werden soll, können verschiedene Audio-Mischungen angelegt werden. Allgemein ist dennoch zu empfehlen, weiterhin eine statische Audioperspektive zu benutzen.

Für Heimanwender oder Bar-Inhaber ist es eine Überlegung wert, eine 3D-Soundanlage einzurichten. Die Immersivität und das Sound-Erlebnis wird deutlich gesteigert und ist für Bar-Inhaber eine Möglichkeit über Eintritt zusätzlich Geld einzunehmen. Da das Einrichten eines Heimkinos zu Hause eine sehr kostspielige Angelegenheit ist, sollten 3D-Audio-Inhalte zusätzlich binaural zur Verfügung gestellt werden. Dies ermöglicht es Personen, die ein Spiel alleine schauen, die Stadionatmosphäre immersiv zu erleben.

Kapitel 6

Zusammenfassung

In dieser Arbeit werden 3D-Audio-Systeme für die Wiedergabe von bildbezogenen Mehrkanal-Audioaufnahmen aus Fußballstadien verglichen. Das vorliegende Audiomaterial, in Form von 16 diskrete Audioaufnahmen, wird für die Wiedergabe über Dolby Atmos, Ambisonics und ein direktes Routing ohne Signalverarbeitung (Direct Out) vorbereitet. Die beiden Systeme Dolby Atmos und Direct Out werden über das Lautsprecher-Setup 5.1.4 wiedergegeben und Ambisonics über ein 11.1 Setup, bei dem auf Ohrlevel fünf Lautsprecher (LS) und auf einer erhöhten Ebene fünf LS aufgebaut sind, plus ein Top-LS über dem Sweet-Spot in der Mitte. Es wird außerdem in Dolby Atmos ein binauraler Downmix für die Wiedergabe über Kopfhörer (HP) eingerichtet. Über Hörversuche wird dann ermittelt, ob sich die Wiedergabe-Systeme anhand von Immersivität, Sound-Qualität, Lokalisation und Illusion der Realität unterscheiden. Die Antwortmöglichkeiten zu den Fragen werden durch die Likert-Skala definiert. Es gibt fünf Stufen von „gar nicht“ bis „sehr“. Die Likert-Skala ist streng genommen ordinal skaliert, wird aber in der Praxis sowie in dieser Arbeit oft als metrisch angenommen. Die Leitfrage bei Hörversuch eins (HV 1) ist, ob sich eins der drei Systeme von den anderen signifikant unterscheidet und bevorzugt wird. In einem weiteren Versuch wird die Wiedergabe über LS mit der über HP verglichen. Die Wiedergabe über HP ist deutlich einfacher und zugänglicher, als die Wiedergabe über LS. Leitfrage bei Hörversuch zwei (HV 2) ist also, ob die Immersivität, Sound-Qualität, Lokalisation und Overall Listener Experience (OLE) bei der HP-Wiedergabe an die der LS-Wiedergabe herankommt oder sogar übertrifft. Auch hier wird ermittelt, ob es einen signifikanten Unterschied bei den Mittelwerten gibt. In einem letzten Versuch werden zwei Audioperspektiven miteinander verglichen. Neben der herkömmlichen statischen Audioperspektive, bei der die Atmosphäre des Stadions immer statisch gleich bleibt, wird eine mobile Audioperspektive abgemischt, bei der die Atmosphäre des Stadions immer angepasst

an den Blickwinkel der Kamera ist. Leitfrage bei Hörversuch drei (HV 3) ist auch hier, ob die Immersivität dadurch deutlich gesteigert werden kann oder ob diese Perspektive für den/die Zuhörer*in verwirrend wirkt. Die Versuche werden mit einer Audioaufnahme aus einem Stadion mit einem Fußballspiel gemacht, können aber auf andere Sport-Events aus Stadien generalisiert werden.

Die Hörversuche haben folgende Ergebnisse geliefert. Bei HV 1 konnte mit einer statistischen Signifikanz ein Unterschied der Mittelwerte bei den oben genannten Aspekte für die drei Wiedergabe-Systeme Dolby Atmos, Ambisonics und Direct Out bestätigt werden. Anhand der einfaktoriellen Varianzanalyse (ANOVA) gepaart mit den Mittelwerten der Fragen kann festgehalten werden, dass Dolby Atmos ein besseres Ergebnis bei den genannten Aspekten liefert und zurecht von der Mehrheit bevorzugt wird. Bei HV 2 konnte hingegen kein signifikanter Unterschied bei den Mittelwerten der Fragen zwischen LS und HP festgestellt werden. Beide Wiedergabe-Formate haben ihre Stärken und Schwächen. Dennoch wurde im Hörversuch die LS-Wiedergabe leicht favorisiert. Vielmehr wurden die verschiedenen Anwendungsgebiete der beiden Formate deutlich. Für die Übertragung von 3D-Audio für mehrere Personen ist nur die LS-Wiedergabe möglich. Bei der Wiedergabe über HP geht zu sehr das Gemeinschaftsgefühl verloren. Dafür eignet sich die HP-Wiedergabe für den Fall, dass einzelne Personen ein Spiel verfolgen wollen. Dies können sie dann auch unterwegs oder zu Hause, wo keine 3D-Soundanlage vorhanden ist. Bei einzelnen Personen kann es zu einem schlechten Hörerlebnis kommen, da bei dem binauralen Downmix in Dolby Atmos eine synthetische HRTF benutzt wird und diese nicht bei allen Personen ein natürliches 3D-Klangerlebnis erzeugt. Die Wiedergabe über LS erzeugt eine bessere Immersivität und die Wiedergabe über HP eine bessere OLE. Bei HV 3 wird von 73% der Proband*innen die statische Audioperspektive bevorzugt. Es wurde allgemein der auditive Perspektivenwechsel gut erkannt, allerdings auch als eher verwirrend wahrgenommen. Nur einzelne Personen fanden die mobile Audioperspektive besser und es gab allgemein größere Unterschiede bei der Bewertung der Fragen, erkennbar an der größeren Standardabweichung. Die mobile Audioperspektive kann trotzdem weiterentwickelt werden und verschiedene Ansätze verfolgen. Beispielsweise können die Übergänge der Audioszenen angepasst werden oder mehrere statische Perspektiven aus unterschiedlichen Positionen im Stadion abgemischt werden, sodass der/die Zuschauer*in einer dieser Optionen auswählen kann.

Literaturverzeichnis

Bassères, Julien & Patrick Thevenot. 2008. How to Widen the Sweet Spot in Monitoring 5.1. Audio Engineering Society.

Blumlein, Alan D. 1958. "British Patent Specification 394,325 (Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems)." *Journal of the Audio Engineering Society* 6(2):91–130. Publisher: Audio Engineering Society.

Brunnström, Kjell et al. 2013. "Qualinet White Paper on Definitions of Quality of Experience."

Dolby Atmos. 2023a. abgerufen am: 2023-11-28.

URL: <https://www.dolby.com/de/technologien/dolby-atmos/>

Dolby Atmos. 2023b. abgerufen am: 2023-11-28.

URL: https://de.wikipedia.org/wiki/Dolby_Atmos

Elen, Richard. 2001. "Ambisonics: The Surround Alternative."

Fisher, H. Geoffrey & Sanford J. Freedman. 1968. "The role of the pinna in auditory localization." *Journal of Auditory Research* 8(1):15–26. Place: US Publisher: C.W. Shilling Auditory Research Center, Inc.

Frank, Matthias & Franz Zotter. 2017. Exploring the Perceptual Sweet Area in Ambisonics. Audio Engineering Society.

Grewe, Yannik, Andreas Walther & Julian Klapp. 2019. Evaluation on the Perceptual Influence of Floor Level Loudspeakers for Immersive Audio Reproduction. Audio Engineering Society.

Im-Kopf-Lokalisation. 2019. abgerufen am: 2023-11-27.

URL: <https://de.wikipedia.org/wiki/Im-Kopf-Lokalisation>

Kamekawa, Toru, Atsushi Marui, Toshihiko Date & Masaaki Enatsu. 2011. Evaluation of Spatial Impression Comparing 2ch Stereo, 5ch Surround, and 7ch Surround with Height Channels for 3D Imagery. Audio Engineering Society.

Kim, Sungyoung, Doyuen Ko, Aparna Nagendra & Wieslaw Woszczyk. 2013. Subjective Evaluation of Multichannel Sound with Surround-Height Channels. Audio Engineering Society.

Likert-Skala | Definition, Anwendung und Auswertung. 2023. abgerufen am: 2023-11-21.

URL: <https://www.rogator.de/wissenswertes/weiteres-fachwissen/likert-skala/>

Malham, Dave. 2019. The Early Years of Ambisonics at York. Audio Engineering Society.

Olivieri, Ferdinando, Nils Peters & Deep Sen. 2019. "A technology overview and application to Next-Generation Audio, VR and 360° Video."

Oramus, Tomas & Petr Neubauer. 2019. Comparison Study of Listeners' Perception of 5.1 and Dolby Atmos. Audio Engineering Society.

Oramus, Tomas & Petr Neubauer. 2020. Comparison of Perception of Spatial Localization Between Channel and Object Based Audio. Audio Engineering Society.

Roginska, Agnieszka & Paul Geluso. 2017. *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Taylor & Francis. Google-Books-ID: IGkPEAAAQBAJ.

Schneiderwind, Christian, Annika Neidhardt & Dominik Meyer. 2021. Comparing the effect of different open headphone models on the perception of a real sound source. Audio Engineering Society.

Sky. 2023. abgerufen am: 2023-11-27.

URL: <https://www.sky.de/>

Sound Quality Assessment Material recordings for subjective tests | EBU Technology & Innovation. 2008.

Sportcast. 2023. abgerufen am: 2023-11-27.

URL: <https://www.sportcast.de/>

Stitt, Peter, Stéphanie Bertet & Maarten van Walstijn. 2014. "Off-Centre Localisation Performance of Ambisonics and HOA For Large and Small Loudspeaker Array Radii." *Acta Acustica united with Acustica* 100(5):937.

Thomas, Graham. 2017. "How Are EBU PSM Members Using 360/VR?"

Thresh, Lewis, Cal Armstrong & Gavin Kearney. 2017. A Direct Comparison of Localization Performance When Using First, Third, and Fifth Ambisonics Order for Real Loudspeaker and Virtual Loudspeaker Rendering. Audio Engineering Society.

Zotter, Franz. 2023. "Spherical harmonics — iaem - internet archiv für elektronische musik."

Anhang A

Fragebogen des Hörversuchs

Vergleich von 3D-Audio-Systemen für die Wiedergabe von bildbezogenen Mehrkanal-Audioaufnahmen aus Fußballstadien

Einleitung:

In diesem Hörversuch geht es darum, die 3D-Audio-Systeme Dolby Atmos, Ambisonics und direktes Routing miteinander zu vergleichen. Dabei gilt es, die Immersivität, Lokalisation und Sound-Qualität der Systeme zu beurteilen. Außerdem wird im zweiten Teil eine Kopfhörer-Mischung mit einer Lautsprecher-Mischung verglichen. Leitfragen sind hier, ob die Kopfhörer-Mischung ähnliche Immersivität wie die Lautsprecher-Mischung erreicht (oder besser) und welche Wiedergabe-Formate bevorzugt wird. Im letzten Teil wird eine statische Audioperspektive mit einer Mobilen verglichen. Hier wird geschaut, ob durch die mobile Perspektive eine stärkere Immersivität erzeugt wird oder ob diese verwirrend ist. Auch hier wird nach einer Präferenz gefragt.

Hörversuch 1: Audiosysteme über Lautsprecher

Frage 1: Wie einschließend ist der Sound dieser Wiedergabe?

	Sehr einschließend	Eher einschließend	Weder noch	Eher nicht einschließend	Gar nicht einschließend
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 2: Wie umfassend ist der Sound dieser Wiedergabe?

	Sehr umfassend	Eher umfassend	Weder noch	Eher nicht umfassend	Gar nicht umfassend
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 3: Wie lebendig ist der Sound dieser Wiedergabe?

	Sehr lebendig	Eher lebendig	Weder noch	Eher nicht lebendig	Gar nicht lebendig
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 4: Wie erfolgreich wird die Stadionatmosphäre vermittelt?

	Sehr erfolgreich	Eher erfolgreich	Weder noch	Eher nicht erfolgreich	Gar nicht erfolgreich
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 5: Wie gut wird auditiv die Illusion der Realität wiedergegeben?

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 6: Wie gut werden alle Frequenzbereiche wiedergegeben? (Bässe, Mitten, Höhen)

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Dolby Atmos					
Ambisonics					
Direct Out					

Falls Bereiche fehlen:

Frage 7: Wie gut können einzelne Sounds lokalisiert werden? (Fangesänge, Kommentator, ...)

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 8: Wie gut ist der Sound auf die verschiedenen Lautsprecher verteilt?

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Dolby Atmos					
Ambisonics					
Direct Out					

Frage 9: Welches der Wiedergabe-Formate bevorzugen Sie?

Dolby Atmos	Ambisonics	Direct Out

Weitere Anmerkungen:

Hörversuch 2: Vergleich von Lautsprechern und Kopfhörern

Frage 1: Wie gut gefällt Ihnen der Sound?

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Lautsprecher					
Kopfhörer					

Frage 2: Wie erfreut oder genervt sind Sie über den Sound während der Wiedergabe?

	Sehr erfreut	Eher erfreut	Weder noch	Eher genervt	Sehr genervt
Lautsprecher					
Kopfhörer					

Frage 3: Wie immersiv ist die Audio-Wiedergabe? (immersiv: umschließend, lebendig, umfassend, intensiv)

	Sehr immersiv	Eher immersiv	Weder noch	Eher nicht immersiv	Gar nicht immersiv
Lautsprecher					
Kopfhörer					

Frage 4: Wie gut wird auditiv die Illusion der Realität wiedergegeben?

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Lautsprecher					
Kopfhörer					

Frage 5: Wie gut werden alle Frequenzbereiche wiedergegeben? (Bässe, Mitten, Höhen)

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut
Lautsprecher					
Kopfhörer					

Falls Bereiche fehlen:

Frage 6: Wie groß oder mächtig wirkt die Wiedergabe auf Sie?

	Sehr groß	Eher groß	Weder noch	Eher nicht groß	Gar nicht groß
Lautsprecher					
Kopfhörer					

Frage 7: Wie stark kommt es zur Im-Kopf-Lokalisation? (Hörereignisort nicht außerhalb vom Kopf in der Umgebung, sondern Orte von Schallquellen scheinen im Kopf zu liegen) NUR BEI KOPFHÖRERN

	Sehr stark	Eher stark	Weder noch	Eher nicht stark	Gar nicht stark
Kopfhörer					

Frage 8: Wie erfolgreich wurde die Stadion-Atmosphäre vermittelt?

	Sehr erfolgreich	Eher erfolgreich	Weder noch	Eher nicht erfolgreich	Gar nicht erfolgreich
Lautsprecher					
Kopfhörer					

Frage 9: Wie gerne würden Sie zu Hause eine Sportübertragung mit 3D-Audio hören?

	Sehr gerne	Eher gerne	Weder noch	Eher nicht gerne	Gar nicht gerne
Lautsprecher					
Kopfhörer					

Frage 10: Wie gerne würden Sie zu Hause mit Freunden zusammen eine Sportübertragung per Kopfhörer hören?

	Sehr gerne	Eher gerne	Weder noch	Eher nicht gerne	Gar nicht gerne
Kopfhörer					

Frage 11: Welche Wiedergabe bevorzugen Sie?

Lautsprecher	Kopfhörer

Frage 12: Wie wichtig ist Ihnen eine gute Soundanlage zu Hause?

	Sehr wichtig	Eher wichtig	Weder noch	Eher nicht wichtig	Gar nicht wichtig
Lautsprecher					

Frage 13: Wie gerne würden Sie für ein besseres Hörerlebnis Geld ausgeben? (z.B. 5€ Eintritt für eine Sportsbar mit 3D-Sound)

	Sehr gerne	Eher gerne	Weder noch	Eher nicht gerne	Gar nicht gerne

Frage 13: Wie wahrscheinlich ist es, dass Sie sich für ein besseres Hörerlebnis ein Heimkino einrichten? (Acht oder mehr Lautsprecher 3D-Soundanlage)

	Sehr wahrscheinlich	Eher wahrscheinlich	Weder noch	Eher nicht wahrscheinlich	Gar nicht wahrscheinlich
Lautsprecher					

Weitere Anmerkungen:

Hörversuch 3: statische vs. mobile Audioperspektive

Frage 1: Wie stark merken Sie einen auditiven Perspektivwechsel?

	Sehr stark	Eher stark	Weder noch	Eher nicht stark	Gar nicht stark

Frage 2: Wie gut finden Sie den Perspektivwechsel?

	Sehr gut	Eher gut	Weder noch	Eher nicht gut	Gar nicht gut

Frage 3: Wie stark wird die Immersivität durch den Perspektivwechsel gesteigert?

	Sehr stark	Eher stark	Weder noch	Eher nicht stark	Gar nicht stark

Frage 4: Wie verwirrend finden Sie den Perspektivwechsel?

	Sehr verwirrend	Eher verwirrend	Weder noch	Eher nicht verwirrend	Gar nicht verwirrend

Frage 5: Wie nervig kann das Hörerlebnis werden, wenn die auditive Perspektive über 90 Minuten an das Kamerabild angepasst wird? (Vermutung)

	Sehr nervig	Eher nervig	Weder noch	Eher nicht nervig	Gar nicht nervig

Frage 6: Welche Audio-Perspektive bevorzugen Sie?

statisch	mobil

Weitere Anmerkungen:

--

Eigenständigkeitserklärung

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

Vergleich von 3D-Audio-Systemen für die Wiedergabe von bildbezogenen Mehrkanal-Audioaufnahmen aus Fußballstadien

selbstständig 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.

Hamburg, 07. Dezember 2023